Journal of Quantitative Spectroscopy & Radiative Transfer 113 (2012) 1437–1440

Contents lists available at SciVerse ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

A generalization of optimal estimation for the retrieval of atmospheric vertical profiles

Simone Ceccherini *

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

A R T I C L E I N F O

Article history:
Received 1 February 2012
Received in revised form
19 March 2012
Accepted 21 March 2012
Available online 29 March 2012

Keywords:
Data fusion
Remote sensing
Optimal estimation
Retrieval techniques

A B S T R A C T

The optimal estimation method for the retrieval of atmospheric vertical profiles uses *a priori* information made of a profile and its covariance matrix. The underlying assumption is that the *a priori* profile has an averaging kernel matrix equal to the identity. The method is herewith generalized to the case that the *a priori* profile has a different averaging kernel matrix. The averaging kernel matrix of the *a priori* profile is properly taken into account in the cost function and a rigorous and more general solution for the optimal estimation method is derived.

1. Introduction

The optimal estimation [1] method is frequently used to perform the retrieval of vertical profiles of atmospheric parameters from spectroscopic measurements. It uses the Bayesian approach [2] in which the *a priori* information about a target vertical profile is updated with some new information provided by a measurement.

The *a priori* information about the vertical profile is represented by a Gaussian probability distribution function which is characterized by the mean vertical profile $\mathbf{x}_a$ (referred to as *a priori profile*) and by the covariance matrix (CM) $\mathbf{S}_a$.

The most satisfactory source of *a priori* information is from independent high spatial resolution measurements of ensembles of the kind of the measured profile as it may be obtained, for example, from radiosonde measurements. Such data are often available as climatologies partitioned by latitude and date (see for example [3,4]). In this case the quantities $\mathbf{x}_a$ and $\mathbf{S}_a$ correctly characterize the *a priori* information.

However, sometimes the *a priori* information is provided by another indirect measurement. This situation occurs, for example, in satellite measurements when the single measurement is too noisy to obtain a useful measurement and the profile retrieved at a nearby geolocation is used as *a priori* information for the retrieval at the next geolocation. This technique is often used for the retrieval of concentration of minor species that have weak spectral features comparable to the noise (an example can be found in [5]). Another situation where the *a priori* information is provided by another indirect measurement occurs when another instrument has measured the same vertical profile in the same geolocation or close to it. In this case the optimal estimation performs the data fusion of the two measurements [6–10].

If the *a priori* information is provided by another indirect measurement the *a priori* profile $\mathbf{x}_a$ is characterized not only by the CM $\mathbf{S}_a$, but also by the averaging kernel matrix (AKM) $\mathbf{A}_a$ [1,11,12]. In this case it is important to take into account that the elements of the vector $\mathbf{x}_a$ do not estimate the values of the true profile at the corresponding altitudes, but are related to the components of the true profile along the row vectors (the averaging kernels) of the AKM.

The classical formulas of the optimal estimation method [1] to retrieve atmospheric vertical profiles implicitly assume

* Tel.: +39 055 5226352; fax: +39 055 5226348.
E-mail address: S.Ceccherini@ifac.cnr.it

0022-4073/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jqsrt.2012.03.021

that the AKM of the *a priori* profile is equal to the identity matrix and do not take into account the possibility that it can have different values. In this paper a generalization of the formulas of the optimal estimation is derived that takes into account the AKM of the *a priori* profile. The derived formulas reduce themselves to the classical formulas when the AKM coincides with the identity matrix.

In Section 2 the basic formulas of the optimal estimation are recalled. In Section 3 the *a priori* information coming from an indirect measurement is analyzed, showing that the provided information is on the components of the true profile along the averaging kernels and describing how to calculate these components. In Section 4 the generalized formulas of the optimal estimation are derived and in Section 5 the CM and the AKM of the solution are calculated. Finally, in Section 6 the conclusions are drawn.

2. Optimal estimation

Consistently with the notations adopted by Rodgers [1], we represent the observations (radiances) 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

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

where $\boldsymbol{\varepsilon}$ is the vector containing the experimental errors of the observations, characterized by a CM $\mathbf{S_y}$ given by the mean value of the product of $\boldsymbol{\varepsilon}$ times its transposed.

In the optimal estimation method we suppose to have an *a priori* information about the profile $\mathbf{x}$ that is described by an *a priori* profile $\mathbf{x}_a$ characterized by a CM $\mathbf{S}_a$ and the optimal estimation solution is obtained minimizing the cost function:

$$J(\mathbf{x})=(\mathbf{y}-\mathbf{F}(\mathbf{x}))^T\mathbf{S}_y^{-1}(\mathbf{y}-\mathbf{F}(\mathbf{x}))+(\mathbf{x}-\mathbf{x}_a)^T\mathbf{S}_a^{-1}(\mathbf{x}-\mathbf{x}_a). \tag{2}$$

The minimization of the cost function corresponds to maximize the *a posteriori* probability distribution function $P(\mathbf{x}|\mathbf{y})$ (conditional probability to obtain $\mathbf{x}$ given $\mathbf{y}$) that, for the Bayes theorem, is proportional to the product of the likelihood function $P(\mathbf{y}|\mathbf{x})$ times the *a priori* probability distribution function $P(\mathbf{x})$ [1,2]. If $P(\mathbf{y}|\mathbf{x})$ and $P(\mathbf{x})$ are Gaussian distributions, then also $P(\mathbf{x}|\mathbf{y})$ is a Gaussian distribution and the cost function is given by minus twice the exponent of $P(\mathbf{x}|\mathbf{y})$. Therefore, the terms that appear in the cost function correspond to minus twice the exponents of $P(\mathbf{y}|\mathbf{x})$ and $P(\mathbf{x})$.

Generally the minimum of $J(\mathbf{x})$ is estimated with the Gauss-Newton method that provides the following iterative formula:

$$\mathbf{x}_{i+1}=\mathbf{x}_i+(\mathbf{K}_i^T\mathbf{S}_y^{-1}\mathbf{K}_i+\mathbf{S}_a^{-1})^{-1}[\mathbf{K}_i^T\mathbf{S}_y^{-1}(\mathbf{y}-\mathbf{F}(\mathbf{x}_i))+\mathbf{S}_a^{-1}(\mathbf{x}_a-\mathbf{x}_i)], \tag{3}$$

where $\mathbf{K}_i$ is the Jacobian matrix of $\mathbf{F}(\mathbf{x})$ calculated in $\mathbf{x}_i$.

The solution $\hat{\mathbf{x}}$ obtained at convergence of the iterative process is characterized by the CM $\mathbf{S}$ and by the AKM $\mathbf{A}$ given by [1]

$$\mathbf{S}=(\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K}+\mathbf{S}_a^{-1})^{-1} \tag{4}$$

$$\mathbf{A}=(\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K}+\mathbf{S}_a^{-1})^{-1}\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K}, \tag{5}$$

where $\mathbf{K}$ is the Jacobian matrix of $\mathbf{F}(\mathbf{x})$ calculated in $\hat{\mathbf{x}}$.

3. *A priori* information coming from another indirect measurement

Now we suppose to have an *a priori* information $\mathbf{x}_a$ about the true profile $\mathbf{x}$ that is the result of the retrieval of another indirect measurement, therefore, the $\mathbf{x}_a$ profile is characterized by a CM $\mathbf{S}_a$ and by an AKM $\mathbf{A}_a$. We can perform an expansion at the first order of the relationship between the retrieved profile $\mathbf{x}_a$ and the true profile $\mathbf{x}$ obtaining [1,11,13]:

$$\mathbf{x}_a=\hat{\mathbf{x}}_0+\mathbf{A}_a(\mathbf{x}-\mathbf{x}_0)+\boldsymbol{\varepsilon}_a, \tag{6}$$

where the vector $\boldsymbol{\varepsilon}_a$ (whose CM is $\mathbf{S}_a$) contains the errors on the retrieved profile $\mathbf{x}_a$, $\mathbf{x}_0$ is a linearization point, that is a profile close enough to the true profile $\mathbf{x}$ in such a way that the linear approximation of Eq. (6) is appropriate, and $\hat{\mathbf{x}}_0$ is the retrieved profile when $\mathbf{x}_0$ is the true profile and the observations are not affected by errors. Eq. (6) is equivalent to Eq. (3.9) in Section 3.1.4 of [1], but is obtained using a generic linearization point that does not necessarily coincide with the *a priori* profile.

From Eq. (6) we can see that if we know $\mathbf{x}_0$ and $\hat{\mathbf{x}}_0$ the profile $\mathbf{x}_a$ allows to determine the product $\mathbf{A}_a\mathbf{x}$, which is the components of $\mathbf{x}$ along the averaging kernels. Indeed, we can calculate the vector:

$$\boldsymbol{\alpha}=\mathbf{x}_a-\hat{\mathbf{x}}_0+\mathbf{A}_a\mathbf{x}_0. \tag{7}$$

Using Eq. (6), we can see that $\boldsymbol{\alpha}$ corresponds to:

$$\boldsymbol{\alpha}=\mathbf{A}_a\mathbf{x}+\boldsymbol{\varepsilon}_a \tag{8}$$

and provides the values of the components of $\mathbf{x}$ along the rows of $\mathbf{A}_a$ with CM $\mathbf{S}_a$.

In the light of this consideration a procedure to calculate $\mathbf{x}_0$ and $\hat{\mathbf{x}}_0$ is herewith described.

A possible choice for the linearization point $\mathbf{x}_0$ is the best estimation that we have of the true profile, that is $\mathbf{x}_a$, the retrieved profile itself. Then $\hat{\mathbf{x}}_0$ can be obtained with a simulated retrieval from error-free observations computed with the forward model applied to $\mathbf{x}_a$.

We can obtain an analytical expression of $\hat{\mathbf{x}}_0$ in the following way. Generally the retrieved profile has been obtained minimizing a cost function $J_a(\mathbf{x})$ that is the sum of the chi-square function plus a constraint function:

$$J_a(\mathbf{x})=(\mathbf{y}_a-\mathbf{F}_a(\mathbf{x}))^T\mathbf{S}_{ya}^{-1}(\mathbf{y}_a-\mathbf{F}_a(\mathbf{x}))+(\mathbf{x}-\mathbf{x}'_a)^T\mathbf{R}_a(\mathbf{x}-\mathbf{x}'_a), \tag{9}$$

where $\mathbf{y}_a$ is the observations vector with CM $\mathbf{S}_{ya}$, $\mathbf{F}_a(\mathbf{x})$ is the forward model, $\mathbf{R}_a$ is a constraint matrix and $\mathbf{x}'_a$ is an *a priori* profile (an apex is used to distinguish this *a priori* from the *a priori* $\mathbf{x}_a$ that we are presently characterizing). All these quantities are related to the retrieval that produced $\mathbf{x}_a$. The constraint function $(\mathbf{x}-\mathbf{x}'_a)^T\mathbf{R}_a(\mathbf{x}-\mathbf{x}'_a)$ is a term that increases when the value of $\mathbf{x}$ (in the case of optimal estimation) or its derivatives (in the case of Tikhonov regularization [14–18]) are different from those of the *a priori* estimation $\mathbf{x}'_a$.

Since the observations computed with the forward model assuming the $\mathbf{x}_a$ profile as the true profile are given by $\mathbf{F}_a(\mathbf{x}_a)$, $\hat{\mathbf{x}}_0$ corresponds to the minimum of the following