On the equivalence between radiance and retrieval assimilation

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1175/MWR-D-10-05047.1

Publisher: American Meteorological Society

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

On the Equivalence between Radiance and Retrieval Assimilation

STEFANO MIGLIORINI

Department of Meteorology, University of Reading, Reading, United Kingdom

ABSTRACT

The need for consistent assimilation of satellite measurements for numerical weather prediction led operational meteorological centers to assimilate satellite radiances directly using variational data assimilation systems. More recently there has been a renewed interest in assimilating satellite retrievals (e.g., to avoid the use of relatively complicated radiative transfer models as observation operators for data assimilation). The aim of this paper is to provide a rigorous and comprehensive discussion of the conditions for the equivalence between radiance and retrieval assimilation. It is shown that two requirements need to be satisfied for the equivalence: (i) the radiance observation operator needs to be approximately linear in a region of the state space centered at the retrieval and with a radius of the order of the retrieval error; and (ii) any prior information used to constrain the retrieval should not underrepresent the variability of the state, so as to retain the information content of the measurements. Both these requirements can be tested in practice. When these requirements are met, retrievals can be transformed so as to represent only the portion of the state that is well constrained by the original radiance measurements and can be assimilated in a consistent and optimal way, by means of an appropriate observation operator and a unit matrix as error covariance. Finally, specific cases when retrieval assimilation can be more advantageous (e.g., when the estimate sought by the operational assimilation system depends on the first guess) are discussed.

1. Introduction

As discussed in Eyre (2007), the late 1970s saw the first attempts to assimilate temperature retrievals from satellite sounders for numerical weather prediction (NWP). These initial results had a modest impact on forecast skill, with the best performance over the ocean where other sources of data were sparse. However, the following decade saw a reduction of the impact of satellite retrievals as the improved accuracy of atmospheric models made them more sensitive to the presence of biases in the assimilated observations. These biases are due to biases in observation errors as well as to the dependence of retrieval errors on the "background" field used to constrain the satellite radiance measurements: when the background field differs from the mean atmospheric forecast field over atmospheric layers whose properties the measurements are not able to sense, the atmospheric analysis becomes biased (Migliorini et al. 2008, see their section 4b).

Since the early 1990s, the operational meteorological community has found that a way to avoid introducing these background-dependent biases is to assimilate satellite data in the form of radiances. This was made possible by the development of variational data assimilation techniques for operational NWP, which allowed the assimilation of large amounts of observational data as well as the use of nonlinear observation operators. However, the observation operator for radiance assimilation is considerably more complex than that used for the assimilation of retrievals (usually an interpolation operator) as it has to represent a solution of the radiative transfer equation as well as the characteristics of the measuring instrument (Migliorini et al. 2008, see their section 4a). Also, in the case of high-spectral-resolution sounders, the number of radiances to be assimilated can be significantly larger than the number of retrieved elements.

Over the last few years, a number of studies (e.g., Joiner and da Silva 1998; Rodgers 2000; Migliorini et al. 2008) have focused on ways to preserve the advantages of assimilating quantities in retrieval space but without its shortcomings. However, to the author's knowledge, a rigorous discussion of the conditions for equivalence between radiance and retrieval assimilation has been absent from

Corresponding author address: Stefano Migliorini, Department of Meteorology, University of Reading, Earley Gate, P.O. Box 243, Reading, RG6 6BB, United Kingdom.
E-mail: s.migliorini@reading.ac.uk

the literature and it is the purpose of this paper to provide it. A set of linearized radiance measurements for assimilation are defined in section 2. Section 3 discusses the equivalence between radiance and retrieval assimilation for the overdetermined inverse problem, while section 4 focuses on the ill-posed (or underdetermined) problem, by first discussing the case when the retrieval and the assimilation scheme share the same background information and then the more general case when they differ. Conclusions are provided in section 5.

2. Characterization of radiance measurements for assimilation

Let us denote with $\mathbf{y}^o \in \mathbb{R}^m$ a vector whose components are radiances measured by a satellite instrument over different spectral channels and with $\mathbf{x}^t \in \mathbb{R}^n$ the random vector assumed to represent the true state of a system (e.g., the atmosphere). The relationship between the radiance measurement vector and the true state can be expressed as

$$\mathbf{y}_{\text{rad}}^o = H(\mathbf{x}^t) + \boldsymbol{\epsilon}_{\text{rad}}^o, \tag{1}$$

where $H(\mathbf{x}^t)$ is the observation operator calculated in $\mathbf{x}^t$ and where $\boldsymbol{\epsilon}_{\text{rad}}^o$ is the radiance measurement error—which includes contributions due to imperfect knowledge of the observation operator and to representativeness error (e.g., Cohn 1997, see his section 2.2)—assumed Gaussian, unbiased, with covariance $\mathbf{R}_{\text{rad}} \in \mathbb{R}^{m \times m}$ and uncorrelated with $\mathbf{x}^t$. The observation operator $H(\mathbf{x}^t)$ is, in general, a nonlinear function of $\mathbf{x}^t$ and models the interaction between radiation and atmospheric constituents, the transfer of radiation through the atmosphere, the emission (and reflection) from the surface, as well as the characteristics of the instrument (e.g., its spectral response and field of view).

In the vicinity of a linearization point $\mathbf{x}_i$, the observation operator (assumed differentiable over its domain) can be approximated with its first-order Taylor expansion around $\mathbf{x}_i$. In this case, Eq. (1) can be written as

$$\mathbf{y}_{\text{rad}}^o \simeq H(\mathbf{x}_i) + \mathbf{H}^{(i)}(\mathbf{x}^t - \mathbf{x}_i) + \boldsymbol{\epsilon}_{\text{rad}}^o, \tag{2}$$

where $\mathbf{H}^{(i)} \equiv (\partial H / \partial \mathbf{x})_{\mathbf{x}=\mathbf{x}_i} \in \mathbb{R}^{m \times n}$ is the Jacobian matrix of $H(\mathbf{x})$ calculated in $\mathbf{x} = \mathbf{x}_i$. We can also define $\mathbf{y}_{\text{rad}}^{(i)}$ as

$$\mathbf{y}_{\text{rad}}^{(i)} \equiv \mathbf{y}_{\text{rad}}^o - H(\mathbf{x}_i) + \mathbf{H}^{(i)}\mathbf{x}_i \simeq \mathbf{H}^{(i)}\mathbf{x}^t + \boldsymbol{\epsilon}_{\text{rad}}^o. \tag{3}$$

It is useful to apply a whitening filter to $\mathbf{y}_{\text{rad}}^{(i)}$ and consider a new set of linearized measurements with independent errors. To this end, let us replace $\mathbf{R}_{\text{rad}}$ with its (possibly truncated) eigenvector decomposition $\mathbf{L}_p \boldsymbol{\Sigma}_p^2 \mathbf{L}_p^{\text{T}}$, where

$\mathbf{L}_p \in \mathbb{R}^{m \times p}$ is a matrix whose columns are the eigenvectors of $\mathbf{R}_{\text{rad}}$ corresponding to the $p \leq m$ nonzero (or nonsmall, as compared to machine precision) eigenvalues of $\mathbf{R}_{\text{rad}}$ lying on the diagonal of the diagonal matrix $\boldsymbol{\Sigma}_p^2 \in \mathbb{R}^{p \times p}$. The number p defines the effective rank of $\mathbf{R}_{\text{rad}}$. If we now define $\mathbf{y}_{\text{rad}}^{(i)'} \equiv \boldsymbol{\Sigma}_p^{-1}\mathbf{L}_p^{\text{T}}\mathbf{y}_{\text{rad}}^{(i)} \in \mathbb{R}^p$, from Eq. (3) we can write

$$\mathbf{y}_{\text{rad}}^{(i)'} \simeq \mathbf{H}_{\text{rad}}^{(i)'}\mathbf{x}^t + \boldsymbol{\epsilon}_{\text{rad}}', \tag{4}$$

where $\mathbf{H}_{\text{rad}}^{(i)'} \equiv \boldsymbol{\Sigma}_p^{-1}\mathbf{L}_p^{\text{T}}\mathbf{H}^{(i)} \in \mathbb{R}^{p \times n}$ and where the covariance of $\boldsymbol{\epsilon}_{\text{rad}}' \equiv \boldsymbol{\Sigma}_p^{-1}\mathbf{L}_p^{\text{T}}\boldsymbol{\epsilon}_{\text{rad}}^o$ is the unit matrix $\mathbf{I}_p \in \mathbb{R}^{p \times p}$. Similarly, it is possible to define $\mathbf{y}_{\text{rad}}^{o'} \in \mathbb{R}^p$ as $\mathbf{y}_{\text{rad}}^{o'} \equiv \boldsymbol{\Sigma}_p^{-1}\mathbf{L}_p^{\text{T}}\mathbf{y}_{\text{rad}}^o$ and $H'(\mathbf{x}^t) \in \mathbb{R}^p$ as $H'(\mathbf{x}^t) \equiv \boldsymbol{\Sigma}_p^{-1}\mathbf{L}_p^{\text{T}}H(\mathbf{x}^t)$. It follows that Eq. (1) can then be written as

$$\mathbf{y}_{\text{rad}}^{o'} = H'(\mathbf{x}^t) + \boldsymbol{\epsilon}_{\text{rad}}'. \tag{5}$$

3. The overdetermined least squares problem

In this section we discuss the equivalence of radiance and retrieval assimilation in the case when the state of the system is well observed. This is not the typical situation that is faced for meteorological applications, but it still provides an important example that shows the equivalence of the two approaches in a mathematically consistent way.

a. Assimilation of radiances

The maximum likelihood estimate of $\mathbf{x}^t$ when radiance measurements are related to the true state as in Eq. (5) is the state that minimizes the cost function $J_o(\mathbf{x})$ defined as (e.g., Lewis et al. 2006, their section 15.3)

$$J_o(\mathbf{x}) = \frac{1}{2}[\mathbf{y}_{\text{rad}}^{o'} - H'(\mathbf{x})]^{\text{T}}[\mathbf{y}_{\text{rad}}^{o'} - H'(\mathbf{x})]. \tag{6}$$

When the number of components of the linearized measurement $\mathbf{y}_{\text{rad}}^{(i)'}$ is not less than the number of elements of the state vector $\mathbf{x}^t$, that is when $p \geq n$, and provided that $\mathbf{H}_{\text{rad}}^{(i)'}$ is full rank [i.e., the rank of $\mathbf{H}_{\text{rad}}^{(i)'}$ is equal to n], the problem of finding the minimum of $J_o(\mathbf{x})$ defined in Eq. (6) is overdetermined (e.g., Rodgers 2000, see his section 2.2). In this case, it is possible to find an estimate of $\mathbf{x}^t$ by means of Gauss–Newton iterations resulting from the minimization of the quadratic cost function $J_o^{(i)}(\mathbf{x})$ (e.g., Lewis et al. 2006, see their section 7.1), defined as

$$J_o^{(i)}(\mathbf{x}) = \frac{1}{2}[\mathbf{y}_{\text{rad}}^{(i)'} - \mathbf{H}_{\text{rad}}^{(i)'}\mathbf{x}]^{\text{T}}[\mathbf{y}_{\text{rad}}^{(i)'} - \mathbf{H}_{\text{rad}}^{(i)'}\mathbf{x}]. \tag{7}$$

The cost function $J_o^{(i)}(\mathbf{x})$ approximates $J_o(\mathbf{x})$ around a small neighborhood of $\mathbf{x}_i$. The state $\mathbf{x}_{i+1}$ is found by setting the gradient of $J_o^{(i)}(\mathbf{x})$ to zero and can be written as

$$\mathbf{x}_{i+1} = [\mathbf{H}_{\mathrm{rad}}^{(i)'\,\mathrm{T}} \mathbf{H}_{\mathrm{rad}}^{(i)'}]^{-1} \mathbf{H}_{\mathrm{rad}}^{(i)'\,\mathrm{T}} \mathbf{y}_{\mathrm{rad}}^{(i)'}. \qquad (8)$$

The iteration in Eq. (8) provides a new estimate $\mathbf{x}_{i+1}$, which can be used as a new linearization point for $H(\mathbf{x})$ and the process can be repeated until convergence (defined by some stopping criterion) is reached. Note that the progression of the Gauss–Newton iteration to a (local) minimum is ensured by the fact that the Hessian matrix $[\partial^2 J_o^{(i)}(\mathbf{x})/\partial \mathbf{x}^2]_{\mathbf{x}=\mathbf{x}_i}$ calculated in $\mathbf{x} = \mathbf{x}_i$ is the positive definite matrix $\mathbf{H}_{\mathrm{rad}}^{(i)'\,\mathrm{T}} \mathbf{H}_{\mathrm{rad}}^{(i)'} \in \mathbb{R}^{n \times n}$ (e.g., Nocedal and Wright 2006, see their section 2.1). At convergence $\mathbf{x}_{i+1} \simeq \mathbf{x}_i \equiv \hat{\mathbf{x}}_{\mathrm{ML}}$, $\mathbf{H}_{\mathrm{rad}}^{(i+1)'} \simeq \mathbf{H}_{\mathrm{rad}}^{(i)'} \equiv \hat{\mathbf{H}}_{\mathrm{rad}}'$, $\mathbf{y}_{\mathrm{rad}}^{(i+1)'} \simeq \mathbf{y}_{\mathrm{rad}}^{(i)'} \equiv \hat{\mathbf{y}}_{\mathrm{rad}}'$, and Eq. (8) becomes

$$\hat{\mathbf{x}}_{\mathrm{ML}} = (\hat{\mathbf{H}}_{\mathrm{rad}}'^{\mathrm{T}} \hat{\mathbf{H}}_{\mathrm{rad}}')^{-1} \hat{\mathbf{H}}_{\mathrm{rad}}'^{\mathrm{T}} \hat{\mathbf{y}}_{\mathrm{rad}}', \qquad (9)$$

where $\hat{\mathbf{x}}_{\mathrm{ML}}$ is the maximum likelihood estimate of $\mathbf{x}^t$, also known as the analysis. The state $\mathbf{x}^t$ may represent a number of model variables defined over a limited area or global domain at a given time. However, when the radiance measurement vector is only composed of radiances acquired over a number of spectral channels at a given time from the same field of view (e.g., from a nadir-viewing instrument at a given location and time), the state $\mathbf{x}^t$ and its estimate is usually represented as a vertical profile of a number of variables of interest. In this case, the estimate of $\mathbf{x}^t$ (e.g., $\hat{\mathbf{x}}_{\mathrm{ML}}$) is usually referred to as the retrieval.

Let us now assume that $\hat{\mathbf{x}}_{\mathrm{ML}}$ represents a satellite retrieval. An approximate expression for the retrieval error covariance $\hat{\mathbf{P}}_{\boldsymbol{\epsilon}_{\mathrm{ML}}}$ can be found by replacing $\hat{\mathbf{y}}_{\mathrm{rad}}'$ in Eq. (9) with the right-hand side of Eq. (4) calculated at convergence. We can write

$$\hat{\mathbf{x}}_{\mathrm{ML}} \simeq \mathbf{x}^t + (\hat{\mathbf{H}}_{\mathrm{rad}}'^{\mathrm{T}} \hat{\mathbf{H}}_{\mathrm{rad}}')^{-1} \hat{\mathbf{H}}_{\mathrm{rad}}'^{\mathrm{T}} \boldsymbol{\epsilon}_{\mathrm{rad}}' = \mathbf{x}^t + \boldsymbol{\epsilon}_{\mathrm{ML}}. \qquad (10)$$

It follows that the covariance of the retrieval error $\boldsymbol{\epsilon}_{\mathrm{ML}}$ can be written as $\hat{\mathbf{P}}_{\boldsymbol{\epsilon}_{\mathrm{ML}}} = (\hat{\mathbf{H}}_{\mathrm{rad}}'^{\mathrm{T}} \hat{\mathbf{H}}_{\mathrm{rad}}')^{-1} = [\partial^2 \hat{J}_o(\mathbf{x})/\partial \mathbf{x}^2]_{\mathbf{x}=\hat{\mathbf{x}}_{\mathrm{ML}}}^{-1}$, where $\hat{J}_o(\mathbf{x})$ represents $J_o^{(i)}(\mathbf{x})$ at convergence. Note that the approximation made to derive the above expression for the retrieval error covariance is justified when $H(\mathbf{x})$ can be replaced with its first-order Taylor expansion about $\hat{\mathbf{x}}_{\mathrm{ML}}$ over a region of the state space where it is likely to find $\mathbf{x}^t$, given $\hat{\mathbf{y}}_{\mathrm{rad}}'$. This region is a neighborhood around $\hat{\mathbf{x}}_{\mathrm{ML}}$ of radius comparable to the typical magnitude of the retrieval error. For a given retrieval, the validity of this approximation can be checked (e.g., Rodgers 2000, see his section 5.1). Hereafter, we only consider retrievals for which this approximation is satisfied.

Let us now express $\hat{\mathbf{H}}_{\mathrm{rad}}'$ in terms of its singular vector decomposition, as $\hat{\mathbf{H}}_{\mathrm{rad}}' = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^{\mathrm{T}}$, where $\mathbf{U} \in \mathbb{R}^{p \times p}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices whose columns are the left and right singular vectors of $\hat{\mathbf{H}}_{\mathrm{rad}}'$ and where the only

nonzero elements of $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times n}$ have the same row and column indexes and are equal to the n positive singular values of $\hat{\mathbf{H}}_{\mathrm{rad}}'$. It follows that $\hat{\mathbf{P}}_{\boldsymbol{\epsilon}_{\mathrm{ML}}}$ can be expressed as $\hat{\mathbf{P}}_{\boldsymbol{\epsilon}_{\mathrm{ML}}} = \mathbf{V} \boldsymbol{\Lambda}_n^{-2} \mathbf{V}^{\mathrm{T}}$, where $\boldsymbol{\Lambda}_n^2 \in \mathbb{R}^{n \times n}$ is the diagonal matrix whose diagonal elements are the (positive) eigenvalues of $\hat{\mathbf{P}}_{\boldsymbol{\epsilon}_{\mathrm{ML}}}^{-1}$. If we multiply both sides of Eq. (10) by $\boldsymbol{\Lambda}_n \mathbf{V}^{\mathrm{T}}$, Eq. (10) can be rewritten as

$$\mathbf{y}_{\mathrm{ret}}' \equiv \boldsymbol{\Lambda}_n \mathbf{V}^{\mathrm{T}} \hat{\mathbf{x}}_{\mathrm{ML}} \simeq \mathbf{H}_{\mathrm{ret}}' \mathbf{x}^t + \boldsymbol{\epsilon}_{\mathrm{ML}}', \qquad (11)$$

where $\mathbf{H}_{\mathrm{ret}}' \equiv \boldsymbol{\Lambda}_n \mathbf{V}^{\mathrm{T}}$ and where the covariance of $\boldsymbol{\epsilon}_{\mathrm{ML}}' \equiv \boldsymbol{\Lambda}_n \mathbf{V}^{\mathrm{T}} \boldsymbol{\epsilon}_{\mathrm{ML}}$ is the identity matrix. The components of $\mathbf{y}_{\mathrm{ret}}'$ represent a new set of measurements (hence the choice of the symbol that defines the vector) that are related to the state $\mathbf{x}^t$ as shown in Eq. (11).

b. Assimilation of maximum likelihood retrievals

In section 3a it was shown that when the observation operator is a nonlinear function of $\mathbf{x}^t$ it is possible to find a maximum likelihood estimate $\hat{\mathbf{x}}_{\mathrm{ML}}$ of $\mathbf{x}^t$ by assimilating a succession (with iteration counter i) of radiance measurement vectors $\mathbf{y}_{\mathrm{rad}}^{(i)'} \in \mathbb{R}^p$ with a succession of rank-n observation operators $\mathbf{H}_{\mathrm{rad}}^{(i)'} \in \mathbb{R}^{p \times n}$, until convergence.

Let us now suppose that we want instead to determine the maximum likelihood estimate of $\mathbf{x}^t$ by assimilating observational information in the form of a linearized retrieval $\mathbf{y}_{\mathrm{ret}}' \in \mathbb{R}^n$ with its rank-n observation operator $\mathbf{H}_{\mathrm{ret}}' \in \mathbb{R}^{n \times n}$. The estimate can now be determined by finding the minimum of the quadratic cost function $J_o^{\mathrm{ret}}(\mathbf{x})$ defined as

$$J_o^{\mathrm{ret}}(\mathbf{x}) = \frac{1}{2} (\mathbf{y}_{\mathrm{ret}}' - \mathbf{H}_{\mathrm{ret}}' \mathbf{x})^{\mathrm{T}} (\mathbf{y}_{\mathrm{ret}}' - \mathbf{H}_{\mathrm{ret}}' \mathbf{x}), \qquad (12)$$

where, for simplicity, we have assumed that the representation of the state $\mathbf{x}$ in the assimilation system is the same as that used to determine $\hat{\mathbf{x}}_{\mathrm{ML}}$ [see Eq. (9)]. A brief discussion of the case when this assumption is not valid is postponed to section 4. As the number of observations (i.e., components of $\mathbf{y}_{\mathrm{ret}}'$), as well as the number of columns, rows and the rank of $\mathbf{H}_{\mathrm{ret}}'$ are all equal to n, the minimizer of $J_o^{\mathrm{ret}}(\mathbf{x})$ is given by

$$\hat{\mathbf{x}}_{\mathrm{ML}}^{\mathrm{ret}} = (\mathbf{H}_{\mathrm{ret}}')^{-1} \mathbf{y}_{\mathrm{ret}}' = (\mathbf{H}_{\mathrm{ret}}')^{-1} \boldsymbol{\Lambda}_n \mathbf{V}^{\mathrm{T}} \hat{\mathbf{x}}_{\mathrm{ML}} = \hat{\mathbf{x}}_{\mathrm{ML}}. \qquad (13)$$

Therefore, the assimilation of a succession of linearized radiances $\mathbf{y}_{\mathrm{rad}}^{(i)'}$ and the assimilation of the linearized retrieval $\mathbf{y}_{\mathrm{ret}}'$ with their respective observation operators and unit matrix error covariances produces the same result. This proves the equivalence between radiance and retrieval assimilation to solve the overdetermined least squares problem, under the assumption of moderate nonlinearity of the observation operator in the vicinity of $\hat{\mathbf{x}}_{\mathrm{ML}}$.

4. The ill-posed or underdetermined problem

Typically, remote sounding measurements of atmospheric temperature, for example, do not provide enough information to constrain all n components of the state vector, due to the finite width of the weighting functions characterizing the spectral channels of a given remote sensing instrument. In this case, it is possible to find an estimate of $\mathbf{x}^t$ by combining the information from the measurements with the *prior* or background information about the state that may be available.

a. Assimilation of radiances

The maximum a posteriori estimate of $\mathbf{x}^t$ when radiance measurements are related to the true state as in Eq. (5) is defined as the state that minimizes the cost function $J(\mathbf{x})$ defined as (e.g., Rodgers 2000, see his section 5.2)

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)$$
$$+ \frac{1}{2}[\mathbf{y}_{\mathrm{rad}}^{o'} - H'(\mathbf{x})]^{\mathrm{T}}[\mathbf{y}_{\mathrm{rad}}^{o'} - H'(\mathbf{x})]. \quad (14)$$

As explained in section 3a, it is possible to provide an estimate of the state by minimizing instead the quadratic cost function $J^{(i)}(\mathbf{x})$ defined as

$$J^{(i)}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)$$
$$+ \frac{1}{2}[\mathbf{y}_{\mathrm{rad}}^{(i)'} - \mathbf{H}_{\mathrm{rad}}^{(i)'}\mathbf{x}]^{\mathrm{T}}[\mathbf{y}_{\mathrm{rad}}^{(i)'} - \mathbf{H}_{\mathrm{rad}}^{(i)'}\mathbf{x}], \quad (15)$$

where $\mathbf{x}^b \in \mathbb{R}^n$ and $\mathbf{B} \in \mathbb{R}^{n \times n}$ are the mean and the positive-definite (i.e., nonsingular) covariance, respectively, of the prior probability density function. The cost function $J^{(i)}(\mathbf{x})$ approximates the cost function $J(\mathbf{x})$ in a small neighborhood of $\mathbf{x}_i$. Note that the Hessian matrix $[\partial^2 J^{(i)}(\mathbf{x})/\partial \mathbf{x}^2]_{\mathbf{x}=\mathbf{x}_i}$ is equal to the positive definite matrix $\mathbf{B}^{-1} + \mathbf{H}_{\mathrm{rad}}^{(i)'\mathrm{T}}\mathbf{H}_{\mathrm{rad}}^{(i)'}$. The minimum of $J^{(i)}(\mathbf{x})$ can be found by setting its gradient to zero. This procedure is repeated iteratively until convergence to the (local) minimum of $J(\mathbf{x})$ is reached. Eventually, the maximum a posteriori estimate $\hat{\mathbf{x}}_{\mathrm{MAP}} \in \mathbb{R}^n$ can be written as (e.g., Rodgers 2000, see his section 5.5)

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \mathbf{x}^b + \mathbf{K}(\hat{\mathbf{y}}_{\mathrm{rad}}' - \hat{\mathbf{H}}_{\mathrm{rad}}'\mathbf{x}^b), \quad (16)$$

with

$$\mathbf{K} \equiv \mathbf{B}\hat{\mathbf{H}}_{\mathrm{rad}}'^{\mathrm{T}}(\hat{\mathbf{H}}_{\mathrm{rad}}'\mathbf{B}\hat{\mathbf{H}}_{\mathrm{rad}}'^{\mathrm{T}} + \mathbf{I}_p)^{-1}$$
$$= (\mathbf{B}^{-1} + \hat{\mathbf{H}}_{\mathrm{rad}}'^{\mathrm{T}}\hat{\mathbf{H}}_{\mathrm{rad}}')^{-1}\hat{\mathbf{H}}_{\mathrm{rad}}'^{\mathrm{T}}, \quad (17)$$

where $\mathbf{K} \in \mathbb{R}^{n \times p}$ is often called the Kalman gain and where we assume here to consider the case when $p < n$.

This condition ensures that the inverse problem under consideration is underdetermined (e.g., Rodgers 2000, see his section 2.2). Let us now define $\mathbf{S} \in \mathbb{R}^{p \times n}$ as the signal-to-noise matrix, of rank $r \leq \min(p, n) = p$, given by $\mathbf{S} \equiv \hat{\mathbf{H}}_{\mathrm{rad}}'\mathbf{B}^{1/2}$, which can be expressed in terms of its singular vector decomposition, as $\mathbf{S} = \mathbf{U}_r\mathbf{\Lambda}_r\mathbf{V}_r^{\mathrm{T}}$, where $\mathbf{\Lambda}_r \in \mathbb{R}^{r \times r}$ is a diagonal matrix that contains the r positive nondimensional singular values (in this case, signal-to-noise values) λ_i of $\mathbf{S}$ on its diagonal and where $\mathbf{U}_r \in \mathbb{R}^{p \times r}$ and $\mathbf{V} \in \mathbb{R}^{n \times r}$ are matrices whose columns are the left and right singular vectors of $\mathbf{S}$ corresponding to the positive singular values of $\mathbf{S}$. It is possible to show (see the appendix) that $\mathbf{K}$ has also rank r and that can be expressed as

$$\mathbf{K} = \mathbf{B}^{1/2}\mathbf{V}_r\mathbf{\Lambda}_r(\mathbf{\Lambda}_r^2 + \mathbf{I}_r)^{-1}\mathbf{U}_r^{\mathrm{T}}. \quad (18)$$

When $H(\mathbf{x})$ can be replaced with its first-order Taylor expansion about $\hat{\mathbf{x}}_{\mathrm{MAP}}$ over a region of the state space where the posterior probability is significant, $\hat{\mathbf{y}}_{\mathrm{rad}}'$ can be approximated with the right-hand side of Eq. (4) calculated at convergence. In this case, from Eq. (16) we can write

$$\hat{\mathbf{x}}_{\mathrm{MAP}} \simeq \mathbf{x}^b + \mathbf{K}\hat{\mathbf{H}}_{\mathrm{rad}}'(\mathbf{x}^t - \mathbf{x}^b) + \mathbf{K}\boldsymbol{\epsilon}_{\mathrm{rad}}', \quad (19)$$

$$\boldsymbol{\epsilon}_{\mathrm{MAP}} \equiv \hat{\mathbf{x}}_{\mathrm{MAP}} - \mathbf{x}^t \simeq (\mathbf{I} - \mathbf{K}\hat{\mathbf{H}}_{\mathrm{rad}}')\boldsymbol{\epsilon}_b + \mathbf{K}\boldsymbol{\epsilon}_{\mathrm{rad}}',$$

where $\boldsymbol{\epsilon}_b \equiv \mathbf{x}^b - \mathbf{x}^t$ is Gaussian, unbiased, with covariance given by $\mathbf{B}$ and where $\mathbf{K}\hat{\mathbf{H}}_{\mathrm{rad}}'$ is known as the averaging kernel matrix. From Eq. (19) it is possible to determine an approximate expression for the covariance $\hat{\mathbf{P}}_{\boldsymbol{\epsilon}_{\mathrm{MAP}}}$ of the maximum a posteriori retrieval error $\boldsymbol{\epsilon}_{\mathrm{MAP}}$, given by

$$\hat{\mathbf{P}}_{\boldsymbol{\epsilon}_{\mathrm{MAP}}} \simeq (\mathbf{I} - \mathbf{K}\hat{\mathbf{H}}_{\mathrm{rad}}')\mathbf{B}(\mathbf{I} - \mathbf{K}\hat{\mathbf{H}}_{\mathrm{rad}}')^{\mathrm{T}} + \mathbf{K}\mathbf{K}^{\mathrm{T}}$$
$$= (\mathbf{B}^{-1} + \hat{\mathbf{H}}_{\mathrm{rad}}'^{\mathrm{T}}\hat{\mathbf{H}}_{\mathrm{rad}}')^{-1}, \quad (20)$$

where we have assumed that $\boldsymbol{\epsilon}_b$ and $\boldsymbol{\epsilon}_{\mathrm{rad}}'$ are mutually uncorrelated and where the rightmost expression in Eq. (20) follows from the definition of $\mathbf{K}$ given in Eq. (17). Note that it can be shown (e.g., Cohn 1997, see his section 5.1) that $\hat{\mathbf{P}}_{\boldsymbol{\epsilon}_{\mathrm{MAP}}} \simeq [\partial^2 \hat{J}_o(\mathbf{x})/\partial \mathbf{x}^2]_{\mathbf{x}=\hat{\mathbf{x}}_{\mathrm{MAP}}}^{-1}$, where $\hat{J}_o(\mathbf{x})$ represents $J_o^{(i)}(\mathbf{x})$ at convergence.

b. Assimilation of maximum a posteriori retrievals

We now want to find the analysis defined as the maximum a posteriori estimate of $\mathbf{x}^t$ over a limited area or a global domain, by assimilating a retrieval determined as shown in section 4a from radiance measurements at a given location and time and some prior information. To this end, we assume that the observation operator for the retrieval is approximately linear around a neighborhood of $\hat{\mathbf{x}}_{\mathrm{MAP}}$ of radius comparable to the estimation error. If we define $\mathbf{y}_{\mathrm{ret}} \in \mathbb{R}^n$ as

$$\mathbf{y}_{\text{ret}} \equiv \hat{\mathbf{x}}_{\text{MAP}} - \mathbf{x}^b + \mathbf{K}\hat{\mathbf{H}}'_{\text{rad}}\mathbf{x}^b, \qquad (21)$$

from Eqs. (18) and (19) we can write

$$
\begin{aligned}
\mathbf{y}_{\text{ret}} &\simeq \mathbf{K}\hat{\mathbf{H}}'_{\text{rad}}\mathbf{x}^t + \boldsymbol{\epsilon}_{\text{ret}} \\
&= \mathbf{KSB}^{-1/2}\mathbf{x}^t + \boldsymbol{\epsilon}_{\text{ret}} \\
&= \mathbf{B}^{1/2}\mathbf{V}_r\boldsymbol{\Lambda}_r^2(\boldsymbol{\Lambda}_r^2 + \mathbf{I}_r)^{-1}\mathbf{V}_r^{\text{T}}\mathbf{B}^{-1/2}\mathbf{x}^t + \boldsymbol{\epsilon}_{\text{ret}}, \qquad (22)
\end{aligned}
$$

where $\boldsymbol{\epsilon}_{\text{ret}} = \mathbf{K}\boldsymbol{\epsilon}'_{\text{rad}}$ with covariance equal to $\mathbf{KK}^{\text{T}} = \mathbf{B}^{1/2}\mathbf{V}_r\boldsymbol{\Lambda}_r^2(\boldsymbol{\Lambda}_r^2 + \mathbf{I}_r)^{-2}\mathbf{V}_r^{\text{T}}\mathbf{B}^{1/2}$. It is now possible to consider only the components of $\mathbf{B}^{-1/2}\mathbf{y}_{\text{ret}}$ along the r right singular vectors of $\mathbf{S}$ (columns of $\mathbf{V}_r \in \mathbb{R}^{n \times r}$) with positive singular values. In particular, if we define $\mathbf{y}'_{\text{ret}} \in \mathbb{R}^r$ as

$$\mathbf{y}'_{\text{ret}} \equiv \boldsymbol{\Lambda}_r^{-1}(\boldsymbol{\Lambda}_r^2 + \mathbf{I}_r)\mathbf{V}_r^{\text{T}}\mathbf{B}^{-1/2}\mathbf{y}_{\text{ret}} \qquad (23)$$

it follows that Eq. (22) can be written as

$$\mathbf{y}'_{\text{ret}} \simeq \boldsymbol{\Lambda}_r\mathbf{V}_r^{\text{T}}\mathbf{B}^{-1/2}\mathbf{x}^t + \boldsymbol{\epsilon}'_{\text{ret}} \equiv \mathbf{H}'_{\text{ret}}\mathbf{x}^t + \boldsymbol{\epsilon}'_{\text{ret}}, \qquad (24)$$

where $\mathbf{H}'_{\text{ret}} \equiv \boldsymbol{\Lambda}_r\mathbf{V}_r^{\text{T}}\mathbf{B}^{-1/2} \in \mathbb{R}^{r \times n}$ and where the covariance of $\boldsymbol{\epsilon}'_{\text{ret}} \equiv \boldsymbol{\Lambda}_r^{-1}(\boldsymbol{\Lambda}_r^2 + \mathbf{I}_r)\mathbf{V}_r^{\text{T}}\mathbf{B}^{-1/2}\boldsymbol{\epsilon}_{\text{ret}}$ is equal to the identity matrix $\mathbf{I}_r \in \mathbb{R}^{r \times r}$.

Note that from Eqs. (16), (18), and (21) we can write

$$\mathbf{y}_{\text{ret}} = \mathbf{K}\hat{\mathbf{y}}'_{\text{rad}} = \mathbf{B}^{1/2}\mathbf{V}_r\boldsymbol{\Lambda}_r(\boldsymbol{\Lambda}_r^2 + \mathbf{I}_r)^{-1}\mathbf{U}_r^{\text{T}}\hat{\mathbf{y}}'_{\text{rad}}, \quad (25)$$

so that from Eqs. (23) and (25) we can write

$$\mathbf{y}'_{\text{ret}} = \mathbf{U}_r^{\text{T}}\hat{\mathbf{y}}'_{\text{rad}}. \qquad (26)$$

Also, from the definitions of $\mathbf{H}'_{\text{ret}}$ and $\mathbf{S}$ we can write

$$\mathbf{H}'_{\text{ret}} = \mathbf{U}_r^{\text{T}}\mathbf{SB}^{-1/2} = \mathbf{U}_r^{\text{T}}\hat{\mathbf{H}}'_{\text{rad}}. \qquad (27)$$

This shows that $\mathbf{y}'_{\text{ret}}$ can be calculated either as a scaled projection of the nondimensional vector $\mathbf{B}^{-1/2}(\hat{\mathbf{x}}_{\text{MAP}} - \mathbf{x}^b + \mathbf{K}\hat{\mathbf{H}}'_{\text{rad}}\mathbf{x}^b)$ in the direction of the columns of $\mathbf{V}_r$ [see Eqs. (21) and (23)], or as the projection of the nondimensional vector $\hat{\mathbf{y}}'_{\text{rad}}$, in the direction of the columns of $\mathbf{U}_r$ [see Eq. (26)]. Similarly, $\mathbf{H}'_{\text{ret}}$ can be calculated according to either Eq. (24) or Eq. (27). As the covariance of $\mathbf{x}^t$ is given by $\mathbf{B}$, from Eq. (24) it follows that the covariance of $\mathbf{y}'_{\text{ret}}$ results equal to $\boldsymbol{\Lambda}_r^2 + \mathbf{I}_r$. If we now denote with λ_i the ith diagonal element of $\boldsymbol{\Lambda}_r$, it is possible to show (Rodgers 2000, see his section 2.5) that $(1/2)\ln(1 + \lambda_i^2)$ is the contribution of the ith component of $\mathbf{y}'_{\text{ret}}$ to the total information content of the measurements. This means that the components of $\mathbf{y}'_{\text{ret}}$ that provide most information are those that are characterized by having signal-to-noise values λ_i greater than about 1.

The transformed retrieval $\mathbf{y}'_{\text{ret}}$ can be assimilated by means of its observation operator $\mathbf{H}'_{\text{ret}}$. In particular, it is possible to find the analysis by minimizing the cost function $J^{\text{ret}}(\mathbf{x})$ defined by replacing $\mathbf{y}_{\text{rad}}^{(i)'}$ with $\mathbf{y}'_{\text{ret}}$ and $\mathbf{H}_{\text{rad}}^{(i)'}$ with $\mathbf{H}'_{\text{ret}}$ in Eq. (15). For this to be meaningful, it is necessary that the representation of the state $\mathbf{x}$ considered in order to perform the minimization of $J^{\text{ret}}(\mathbf{x})$ is the same as that used to minimize $J(\mathbf{x})$. If this is not the case (e.g., when the retrieval is produced outside the data assimilation system), each variable in $\mathbf{x}$ that contributes to the signal needs to be interpolated onto the grid that allows the state to be consistently multiplied (on the left) by $\mathbf{H}'_{\text{ret}}$, in a way to minimize information loss (Migliorini et al. 2008, see their section 7). Hereafter, it is understood that $\mathbf{H}'_{\text{ret}}$ may include a suitable interpolation operator that makes $\mathbf{H}'_{\text{ret}}\mathbf{x}$ meaningful. Two different cases are discussed.

1) ASSIMILATION OF RADIANCES AND RETRIEVALS WITH THE SAME PRIOR INFORMATION

Let us consider first the case when the prior information used for data assimilation is the same as that used to determine the retrieval, for the common components of the state. In this case, the analysis $\hat{\mathbf{x}}_{\text{MAP}}^{\text{ret}}$ can be found by minimizing the cost function $J^{\text{ret}}(\mathbf{x})$ defined as

$$
\begin{aligned}
J^{\text{ret}}(\mathbf{x}) = &\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^{\text{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) \\
&+ \frac{1}{2}(\mathbf{y}'_{\text{ret}} - \mathbf{H}'_{\text{ret}}\mathbf{x})^{\text{T}}(\mathbf{y}'_{\text{ret}} - \mathbf{H}'_{\text{ret}}\mathbf{x}) \quad (28)
\end{aligned}
$$

and $\hat{\mathbf{x}}_{\text{MAP}}^{\text{ret}}$ can be expressed as

$$\hat{\mathbf{x}}_{\text{MAP}}^{\text{ret}} = \mathbf{x}^b + \mathbf{K}_{\text{ret}}(\mathbf{y}'_{\text{ret}} - \mathbf{H}'_{\text{ret}}\mathbf{x}^b), \qquad (29)$$

where $\mathbf{K}_{\text{ret}} \equiv \mathbf{BH}_{\text{ret}}'^{\text{T}}(\mathbf{H}'_{\text{ret}}\mathbf{BH}_{\text{ret}}'^{\text{T}} + \mathbf{I}_r)^{-1}$ can be written as [see Eqs. (17), (18) and (27)]

$$
\begin{aligned}
\mathbf{K}_{\text{ret}} &= \mathbf{B}\hat{\mathbf{H}}_{\text{rad}}'^{\text{T}}\mathbf{U}_r(\mathbf{U}_r^{\text{T}}\hat{\mathbf{H}}'_{\text{rad}}\mathbf{B}\hat{\mathbf{H}}_{\text{rad}}'^{\text{T}}\mathbf{U}_r + \mathbf{I}_r)^{-1} \\
&= \mathbf{B}^{1/2}\mathbf{S}^{\text{T}}\mathbf{U}_r(\mathbf{U}_r^{\text{T}}\mathbf{SS}^{\text{T}}\mathbf{U}_r + \mathbf{I}_r)^{-1} \\
&= \mathbf{B}^{1/2}\mathbf{V}_r\boldsymbol{\Lambda}_r(\boldsymbol{\Lambda}_r^2 + \mathbf{I}_r)^{-1} \\
&= \mathbf{KU}_r. \qquad (30)
\end{aligned}
$$

From Eqs. (16), (23), (27), (29), and (30) it follows that the analysis $\hat{\mathbf{x}}_{\text{MAP}}^{\text{ret}}$ can be written as

$$
\begin{aligned}
\hat{\mathbf{x}}_{\text{MAP}}^{\text{ret}} &= \mathbf{x}^b + \mathbf{KU}_r(\mathbf{U}_r^{\text{T}}\hat{\mathbf{y}}'_{\text{rad}} - \mathbf{U}_r^{\text{T}}\hat{\mathbf{H}}'_{\text{rad}}\mathbf{x}^b) \\
&= \mathbf{x}^b + \mathbf{KU}_r\mathbf{U}_r^{\text{T}}(\hat{\mathbf{y}}'_{\text{rad}} - \hat{\mathbf{H}}'_{\text{rad}}\mathbf{x}^b) \\
&= \mathbf{x}^b + \mathbf{K}(\hat{\mathbf{y}}'_{\text{rad}} - \hat{\mathbf{H}}'_{\text{rad}}\mathbf{x}^b) \\
&= \hat{\mathbf{x}}_{\text{MAP}}, \qquad (31)
\end{aligned}
$$

where we have used the equivalence $\mathbf{K} = \mathbf{KU}_r\mathbf{U}_r^{\mathrm{T}}$ that follows from Eq. (18). This proves the equivalence between assimilating radiances and retrievals in the case when the prior information used first to determine and then to assimilate the retrieval are the same.

2) ASSIMILATION OF RADIANCES AND RETRIEVALS WITH DIFFERENT PRIOR INFORMATION

Let us now assume that we want to assimilate a succession of radiance measurements $\mathbf{y}_{\mathrm{rad}}^{(i)\prime}$ and estimate the analysis by finding the minimum of a succession of cost functions similar to that in Eq. (15), but using a prior density function with different mean $\mathbf{x}^{b*} \in \mathbb{R}^n$ and covariance $\mathbf{B}^* \in \mathbb{R}^{n \times n}$. This case is of practical importance, as the retrieval to be assimilated may have been estimated by an external data provider, who may have used prior constraints that differ from those used for assimilation. In this case, the resulting analysis $\hat{\mathbf{x}}_{\mathrm{MAP}}^*$ can be written from Eq. (16) as

$$\hat{\mathbf{x}}_{\mathrm{MAP}}^* = \mathbf{x}^{b*} + \mathbf{K}^*(\hat{\mathbf{y}}_{\mathrm{rad}}^{*\prime} - \hat{\mathbf{H}}_{\mathrm{rad}}^{*\prime}\mathbf{x}^{b*}), \qquad (32)$$

where $\hat{\mathbf{y}}_{\mathrm{rad}}^{*\prime}$ and $\hat{\mathbf{H}}_{\mathrm{rad}}^{*\prime}$ differ from $\hat{\mathbf{y}}_{\mathrm{rad}}^{\prime}$ and $\hat{\mathbf{H}}_{\mathrm{rad}}^{\prime}$, respectively, for the different value of the retrieval used as linearization point of $H(\mathbf{x})$. Note that, in general, the rank of $\hat{\mathbf{H}}_{\mathrm{rad}}^{*\prime}$ is $s \neq r$. From Eq. (4) we can write

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{MAP}}^* &\simeq \mathbf{x}^{b*} + \mathbf{K}^*\hat{\mathbf{H}}_{\mathrm{rad}}^{*\prime}(\mathbf{x}^t - \mathbf{x}^{b*}) + \mathbf{K}^*\boldsymbol{\epsilon}_{\mathrm{rad}}^{\prime} \\ &= \mathbf{x}^{b*} + \mathbf{K}^*\mathbf{S}^*\mathbf{B}^{*-1/2}(\mathbf{x}^t - \mathbf{x}^{b*}) + \mathbf{K}^*\boldsymbol{\epsilon}_{\mathrm{rad}}^{\prime},\end{aligned}$$
$$(33)$$

with $\mathbf{S}^* \equiv \hat{\mathbf{H}}_{\mathrm{rad}}^{*\prime}\mathbf{B}^{*1/2} = \mathbf{U}_s^*\boldsymbol{\Lambda}_s^*\mathbf{V}_s^{*\mathrm{T}}$, where from Eqs. (17), (A2), (A3), and (A5) we can write

$$\begin{aligned}\mathbf{K}^* &\equiv \mathbf{B}^*\hat{\mathbf{H}}_{\mathrm{rad}}^{*\prime\mathrm{T}}(\hat{\mathbf{H}}_{\mathrm{rad}}^{*\prime}\mathbf{B}^*\hat{\mathbf{H}}_{\mathrm{rad}}^{*\prime\mathrm{T}} + \mathbf{I}_p)^{-1} \\ &= \mathbf{B}^{*1/2}\mathbf{S}^{*\mathrm{T}}(\mathbf{S}^*\mathbf{S}^{*\mathrm{T}} + \mathbf{I}_p)^{-1} \\ &= \mathbf{B}^{*1/2}\mathbf{V}_s^*\boldsymbol{\Lambda}_s^*(\boldsymbol{\Lambda}_s^{*2} + \mathbf{I}_s)^{-1}\mathbf{U}_s^{*\mathrm{T}} \\ &= \mathbf{B}^{*1/2}\mathbf{S}^{*\mathrm{T}}\mathbf{U}_s^*(\mathbf{U}_s^{*\mathrm{T}}\mathbf{S}^*\mathbf{S}^{*\mathrm{T}}\mathbf{U}_s^* + \mathbf{I}_s)^{-1}\mathbf{U}_s^{*\mathrm{T}}, \qquad (34)\end{aligned}$$

where $\boldsymbol{\Lambda}_s^* \in \mathbb{R}^{s \times s}$ is a diagonal matrix that contains the s positive nondimensional singular values of $\mathbf{S}^*$ on its diagonal and $\mathbf{U}_s^* \in \mathbb{R}^{p \times s}$ and $\mathbf{V}_s^* \in \mathbb{R}^{n \times s}$ are the matrices whose columns are the left and right singular vectors of $\mathbf{S}^*$ corresponding to the positive singular values of $\mathbf{S}^*$.

Consider now the retrieval $\mathbf{y}_{\mathrm{ret}}^{\prime}$ defined in Eq. (23) and estimated by using prior information $\mathbf{x}^b$ and $\mathbf{B}$. We want to assimilate $\mathbf{y}_{\mathrm{ret}}^{\prime}$ with its observation operator $\mathbf{H}_{\mathrm{ret}}^{\prime}$ by finding the state $\hat{\mathbf{x}}_{\mathrm{MAP}}^{\mathrm{ret}*}$ that minimizes $J^{\mathrm{ret}}(\mathbf{x})$ [see

Eq. (28)], in the case when the prior information used to constrain $\mathbf{y}_{\mathrm{ret}}^{\prime}$ is $\mathbf{x}^{b*}$ and $\mathbf{B}^*$. To prove the equivalence between radiance and retrieval assimilation in the case when the prior information used first to determine and then to assimilate the retrieval are different, we need to show that $\hat{\mathbf{x}}_{\mathrm{MAP}}^{\mathrm{ret}*} \simeq \hat{\mathbf{x}}_{\mathrm{MAP}}^*$. From Eq. (29) it follows that $\hat{\mathbf{x}}_{\mathrm{MAP}}^{\mathrm{ret}*}$ can be written as

$$\hat{\mathbf{x}}_{\mathrm{MAP}}^{\mathrm{ret}*} = \mathbf{x}^{b*} + \mathbf{K}_{\mathrm{ret}}^*(\mathbf{y}_{\mathrm{ret}}^{\prime} - \mathbf{H}_{\mathrm{ret}}^{\prime}\mathbf{x}^{b*}), \qquad (35)$$

where, from Eqs. (17) and (27), $\mathbf{K}_{\mathrm{ret}}^{\star} \in \mathbb{R}^{n \times r}$ can be expressed as

$$\begin{aligned}\mathbf{K}_{\mathrm{ret}}^{\star} &\equiv \mathbf{B}^*\mathbf{H}_{\mathrm{ret}}^{\prime\mathrm{T}}(\mathbf{H}_{\mathrm{ret}}^{\prime}\mathbf{B}^*\mathbf{H}_{\mathrm{ret}}^{\prime\mathrm{T}} + \mathbf{I}_r)^{-1} \\ &= \mathbf{B}^*\hat{\mathbf{H}}_{\mathrm{rad}}^{\prime\mathrm{T}}\mathbf{U}_r(\mathbf{U}_r^{\mathrm{T}}\hat{\mathbf{H}}_{\mathrm{rad}}^{\prime}\mathbf{B}^*\hat{\mathbf{H}}_{\mathrm{rad}}^{\prime\mathrm{T}}\mathbf{U}_r + \mathbf{I}_r)^{-1} \\ &= \mathbf{B}^{*1/2}\mathbf{S}^{\star\mathrm{T}}\mathbf{U}_r(\mathbf{U}_r^{\mathrm{T}}\mathbf{S}^{\star}\mathbf{S}^{\star\mathrm{T}}\mathbf{U}_r + \mathbf{I}_r)^{-1}, \qquad (36)\end{aligned}$$

with $\mathbf{S}^{\star} \equiv \hat{\mathbf{H}}_{\mathrm{rad}}^{\prime}\mathbf{B}^{*1/2} \in \mathbb{R}^{p \times n}$. In analogy with Eq. (30), let us now find the conditions when it is possible to write $\mathbf{K}_{\mathrm{ret}}^{\star} = \mathbf{K}^*\mathbf{U}_s^*$. A comparison between Eqs. (34) and (36) shows that $\mathbf{K}_{\mathrm{ret}}^{\star} = \mathbf{K}^*\mathbf{U}_s^*$ when $s = r$ and $\mathbf{U}_r^{\mathrm{T}}\mathbf{S}^{\star} = \mathbf{U}_r^{*\mathrm{T}}\mathbf{S}^*$. Therefore, by assuming $\mathbf{U}_r^{\mathrm{T}}\mathbf{S}^{\star} = \mathbf{U}_r^{*\mathrm{T}}\mathbf{S}^*$, from Eqs. (4), (26), (27), (35), and (36) we can write

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{MAP}}^{\mathrm{ret}*} &= \mathbf{x}^{b*} + \mathbf{K}^*\mathbf{U}_r^*\mathbf{U}_r^{\mathrm{T}}(\hat{\mathbf{y}}_{\mathrm{rad}}^{\prime} - \hat{\mathbf{H}}_{\mathrm{rad}}^{\prime}\mathbf{x}^{b*}) \\ &\simeq \mathbf{x}^{b*} + \mathbf{K}^*\mathbf{U}_r^*\mathbf{U}_r^{\mathrm{T}}\mathbf{S}^{\star}\mathbf{B}^{*-1/2}(\mathbf{x}^t - \mathbf{x}^{b*}) \\ &\quad + \mathbf{K}^*\mathbf{U}_r^*\mathbf{U}_r^{\mathrm{T}}\boldsymbol{\epsilon}_{\mathrm{rad}}^{\prime} \\ &= \mathbf{x}^{b*} + \mathbf{K}^*\mathbf{S}^*\mathbf{B}^{*-1/2}(\mathbf{x}^t - \mathbf{x}^{b*}) + \mathbf{K}^*\mathbf{U}_r^*\mathbf{U}_r^{\mathrm{T}}\boldsymbol{\epsilon}_{\mathrm{rad}}^{\prime},\end{aligned}$$
$$(37)$$

where $\mathbf{K}^*\mathbf{U}_r^*\mathbf{U}_r^{*\mathrm{T}} = \mathbf{K}^*$. From Eqs. (33) and (37) it follows that the condition $\mathbf{U}_r^{\mathrm{T}}\mathbf{S}^{\star} = \mathbf{U}_r^{*\mathrm{T}}\mathbf{S}^*$ implies that $\hat{\mathbf{x}}_{\mathrm{MAP}}^{\mathrm{ret}*} \simeq \hat{\mathbf{x}}_{\mathrm{MAP}}^*$ within retrieval noise (note that the covariance of $\mathbf{K}^*\boldsymbol{\epsilon}_{\mathrm{rad}}^{\prime}$ and $\mathbf{K}^*\mathbf{U}_r^*\mathbf{U}_r^{\mathrm{T}}\boldsymbol{\epsilon}_{\mathrm{rad}}^{\prime}$ are both equal to $\mathbf{K}^*\mathbf{K}^{*\mathrm{T}}$). Now, by noting that $\mathbf{S}^{\star}$ can in general also be written as $\mathbf{S}^{\star} = \mathbf{SB}^{-1/2}\mathbf{B}^{*1/2}$, it follows that $\hat{\mathbf{x}}_{\mathrm{MAP}}^{\mathrm{ret}*} \simeq \hat{\mathbf{x}}_{\mathrm{MAP}}^*$ holds when $\mathbf{U}_r^{*\mathrm{T}}\mathbf{S}^*\mathbf{B}^{*-1/2} = \mathbf{U}_r^{\mathrm{T}}\mathbf{SB}^{-1/2}$, that is, when $\mathbf{H}_{\mathrm{ret}}^{\prime} \equiv \boldsymbol{\Lambda}_r\mathbf{V}_r^{\mathrm{T}}\mathbf{B}^{-1/2} = \boldsymbol{\Lambda}_r^*\mathbf{V}_r^{*\mathrm{T}}\mathbf{B}^{*-1/2}$. This means that $\hat{\mathbf{x}}_{\mathrm{MAP}}^{\mathrm{ret}*} \simeq \hat{\mathbf{x}}_{\mathrm{MAP}}^*$ holds when $\mathbf{H}_{\mathrm{ret}}^{\prime}\mathbf{BH}_{\mathrm{ret}}^{\prime\mathrm{T}}$ (the covariance of $\mathbf{H}_{\mathrm{ret}}^{\prime}\mathbf{x}^t$ in the case when the covariance of $\mathbf{x}^t$ is $\mathbf{B}$) is equal to $\mathbf{H}_{\mathrm{ret}}^{\prime}\mathbf{B}^*\mathbf{H}_{\mathrm{ret}}^{\prime\mathrm{T}}$ (the covariance of $\mathbf{H}_{\mathrm{ret}}^{\prime}\mathbf{x}^t$ in the case when the covariance of $\mathbf{x}^t$ is $\mathbf{B}^*$; i.e., when $\boldsymbol{\Lambda}_r^* = \boldsymbol{\Lambda}_r$). In other words, the equivalence is satisfied when the difference between $\hat{\mathbf{x}}_{\mathrm{MAP}}^*$ and $\hat{\mathbf{x}}_{\mathrm{MAP}}$—arising from the use of a different prior constraint—preserves the information content of the measurements, defined in terms of the diagonal elements of $\boldsymbol{\Lambda}_r$ as shown in section 4b. In this respect, note that $\boldsymbol{\Lambda}_r^* = \boldsymbol{\Lambda}_r$ does not necessarily implies that $\mathbf{B}^* = \mathbf{B}$—in which case the equivalence between $\hat{\mathbf{x}}_{\mathrm{MAP}}^{\mathrm{ret}*}$ and $\hat{\mathbf{x}}_{\mathrm{MAP}}^*$ would

be trivially ensured by Eq. (31)—because the covariance of the components of the state $\mathbf{x}^t$, which lie in the null space of $\mathbf{H}'_{\text{ret}}$ in the case when the covariance of $\mathbf{x}^t$ is $\mathbf{B}^*$, do not alter the information content that the same measurements have in the case when the covariance of $\mathbf{x}^t$ is $\mathbf{B}$.

The above discussion proves the equivalence between assimilating radiances and retrievals in the case when the prior information used first to determine and then to assimilate the retrieval differ, provided that (i) the observation operator is approximately linear over a region of the state space centered in $\hat{\mathbf{x}}_{\text{MAP}}$ where it is likely to find $\mathbf{x}^t$ given $\hat{\mathbf{y}}'_{\text{rad}}$ and that (ii) both the prior constraint used to determine the retrieval and that used for radiance assimilation are chosen in a way not to lose the information content of the measurements.

5. Conclusions

In this paper a rigorous proof of the equivalence between assimilating radiance measurements and assimilating appropriately transformed retrievals are determined offline (i.e., outside a given data assimilation system) from the same set of radiance measurements. The only requirements for the equivalence are the following: (i) the radiance observation operator needs to be approximately linear in a region of the state space centered at the retrieval and with a radius of the order of the retrieval error and (ii) any prior information used to constrain the retrieval should not underrepresent the variability of the state, so as to retain the information content of the measurements. Requirement (i) can be tested once the retrieval has been determined, while it is possible to compare an ensemble of measurements linearized about the retrieval (and adjusted to account for observation error) with the chosen prior information in observation space to ensure the validity of requirement (ii). If the requirements are met, it is possible to improve the efficiency of the assimilation of remote sounding measurements by performing the nonlinear (due to the nonlinear radiance observation operator) least squares estimation before assimilation.

Estimated retrievals can then be transformed so as to represent only the portion of the state that is well constrained by the original radiance measurements and assimilated in a consistent and optimal way, by means of an appropriate observation operator and a unit matrix as error covariance. Note that the number of elements needed to fully describe a remote sensing measurement is now $r + r \times n$, where r is the rank of the signal-to-noise matrix $\mathbf{S}$ (as well as of $\hat{\mathbf{H}}_{\text{rad}}$) and n is the dimension of the state vector used to perform the retrieval, usually representing vertical profiles and surface values of a set of geophysical fields. Note that $r \leq \min(m, n)$, where m is the dimension of the measurement vector (i.e., the number of spectral channels in a remote sounding instrument). It is also important to note that the observation operator used for assimilation of transformed retrievals (as well as that used for radiance assimilation) must include all components of the state that are relevant for the purpose of representing such measurements.

Assimilation of transformed retrievals may be particularly advantageous for remote sounding instruments with a very high number of channels or when efficient radiative transfer models used for operational assimilation of radiance measurements are not able to model the spectral regions (e.g., visible or ultraviolet) observed by the instrument. An estimation of the retrieval before assimilation would also be beneficial when the nonlinearity of the observation operator is such that $J(\mathbf{x})$ admits multiple minima (i.e., when the posterior probability density function is multimodal) over its domain, in the assumption that the retrieval could be estimated by using more sophisticated techniques than those currently used for operational assimilation of radiances. In this case, for example, the usual (e.g., quasi-Newton) minimization algorithm normally used to determine the retrieval could be applied successively to explore each local minimum of $J(\mathbf{x})$ and estimate the retrieval, which attains the global minimum of $J(\mathbf{x})$. When the number of local minima is large, the global minimum of $J(\mathbf{x})$ could instead be found by using a Monte Carlo method (e.g., Tarantola 2005, see his section 2.4). In the case when $H(\mathbf{x})$ was approximately linear about the retrieval, which results in the global minimum of $J(\mathbf{x})$, it would be possible to assimilate the linearized and transformed retrieval in the operational data assimilation system in the place of a succession of linearized radiances. In this way, the accuracy of the analysis would improve with respect to the case when radiance data are assimilated directly with a first guess (e.g., the background) leading to a local but not global minimum of $J(\mathbf{x})$.

APPENDIX

Signal-to-Noise Representation of the Kalman Gain

As shown in Eq. (17), the Kalman gain $\mathbf{K} \in \mathbb{R}^{n \times p}$ is defined as

$$\mathbf{K} = \mathbf{B}\hat{\mathbf{H}}'^{\text{T}}_{\text{rad}}(\hat{\mathbf{H}}'_{\text{rad}}\mathbf{B}\hat{\mathbf{H}}'^{\text{T}}_{\text{rad}} + \mathbf{I}_p)^{-1}, \qquad (A1)$$

where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is a symmetric positive definite (i.e., nonsingular) matrix with rank $(\mathbf{B}) = n$ and where

$\hat{\mathbf{H}}'_{\text{rad}} \in \mathbb{R}^{p \times n}$ and $\mathbf{I}_p \in \mathbb{R}^{p \times p}$ is the identity matrix with rank $(\mathbf{I}_p) = p$. Let us now express the signal-to-noise matrix $\mathbf{S}$ as $\mathbf{S} \equiv \hat{\mathbf{H}}'_{\text{rad}} \mathbf{B}^{1/2} \in \mathbb{R}^{p \times n}$—where $\mathbf{B}^{1/2}$ is the symmetric square root of $\mathbf{B}$, with rank$(\mathbf{B}^{1/2}) = n$—in terms of its singular vector decomposition, as $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\text{T}}$, where $\mathbf{U} \in \mathbb{R}^{p \times p}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices whose columns are the left and right singular vectors of $\mathbf{S}$ and where the only nonzero elements of $\mathbf{\Lambda} \in \mathbb{R}^{p \times n}$ have the same row and column indexes and are equal to the $r \leq \min(n, p)$ positive nondimensional singular values λ_i of $\mathbf{S}$. It follows that Eq. (A1) can be written as

$$\begin{aligned} \mathbf{K} &= \mathbf{B}^{1/2}\mathbf{S}^{\text{T}}(\mathbf{S}\mathbf{S}^{\text{T}} + \mathbf{I}_p)^{-1} \\ &= \mathbf{B}^{1/2}\mathbf{V}\mathbf{\Lambda}^{\text{T}}\mathbf{U}^{\text{T}}(\mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}^{\text{T}}\mathbf{U}^{\text{T}} + \mathbf{I}_p)^{-1} \\ &= \mathbf{B}^{1/2}\mathbf{V}\mathbf{\Lambda}^{\text{T}}(\mathbf{\Lambda}\mathbf{\Lambda}^{\text{T}} + \mathbf{I}_p)^{-1}\mathbf{U}^{\text{T}}, \end{aligned} \quad (A2)$$

where we have used the property of the inverse of the product between square nonsingular matrices and that $\mathbf{U}^{-1} = \mathbf{U}^{\text{T}}$. If we now define $\mathbf{\Lambda}_r \in \mathbb{R}^{r \times r}$ as the diagonal matrix that contains the r positive nondimensional singular values (in this case, signal-to-noise values) λ_i of $\mathbf{S}$ on its diagonal and $\mathbf{U}_r \in \mathbb{R}^{p \times r}$ and $\mathbf{V}_r \in \mathbb{R}^{n \times r}$ as the matrices whose columns are the left and right singular vectors of $\mathbf{S}$ corresponding to the positive singular values of $\mathbf{S}$, from Eq. (A2) the Kalman gain can also be written as

$$\mathbf{K} = \mathbf{B}^{1/2}\mathbf{V}_r\mathbf{\Lambda}_r(\mathbf{\Lambda}_r^2 + \mathbf{I}_r)^{-1}\mathbf{U}_r^{\text{T}} \equiv \mathbf{B}^{1/2}\mathbf{C}, \quad (A3)$$

where $\mathbf{C} \equiv \mathbf{V}_r\mathbf{\Lambda}_r(\mathbf{\Lambda}_r^2 + \mathbf{I}_r)^{-1}\mathbf{U}_r^{\text{T}}$. Let us now recall that the rank of a matrix $\mathbf{C} \in \mathbb{R}^{n \times p}$ is the dimension of the range of $\mathbf{C}$ defined as the subspace of $\mathbb{R}^p$ generated by the columns of $\mathbf{C}$ (e.g., Meyer 2000, see his section 4.2). Given that the r left singular vectors corresponding to the r positive singular values of $\mathbf{C}$ define a basis for the range of $\mathbf{C}$ (e.g., Golub and van Loan 1996, see their section 2.5.3), it follows that rank $(\mathbf{C}) = r$. It is also possible to prove that the rank is invariant under multiplication by

a nonsingular matrix (e.g., Meyer 2000, see his section 4.5). We can then write

$$\text{rank}(\mathbf{K}) = \text{rank}(\mathbf{B}^{1/2}\mathbf{C}) = \text{rank}(\mathbf{C}) = r \leq \min(n, p). \quad (A4)$$

Note that this also implies that the Kalman gain has the same rank as the signal-to-noise matrix $\mathbf{S}$ and as the observation operator $\hat{\mathbf{H}}'_{\text{rad}}$. Finally, if we express $\mathbf{S}$ as $\mathbf{S} = \mathbf{U}_r\mathbf{\Lambda}_r\mathbf{V}_r^{\text{T}}$, from Eq. (A3) we can also write

$$\mathbf{K} = \mathbf{B}^{1/2}\mathbf{S}^{\text{T}}\mathbf{U}_r(\mathbf{U}_r^{\text{T}}\mathbf{S}\mathbf{S}^{\text{T}}\mathbf{U}_r + \mathbf{I}_r)^{-1}\mathbf{U}_r^{\text{T}}. \quad (A5)$$

REFERENCES

Cohn, S., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan,* **75,** 257–288.

Eyre, J., 2007: Progress achieved on assimilation of satellite data in numerical weather prediction over the last 30 years. *Proc. ECMWF Seminar on "Recent Developments in the Use of Satellite Observations in Numerical Weather Prediction,"* Reading, United Kingdom, ECMWF, 1–27.

Golub, G., and C. van Loan, 1996: *Matrix Computations.* 3rd ed. The John Hopkins University Press, 694 pp.

Joiner, J., and A. da Silva, 1998: Efficient methods to assimilate remotely sensed data based on information content. *Quart. J. Roy. Meteor. Soc.,* **124,** 1669–1694.

Lewis, J., S. Lakshmivarahan, and S. Dhall, 2006: *Dynamic Data Assimilation: A Least Squares Approach.* Cambridge University Press, 654 pp.

Meyer, C., 2000: *Matrix Analysis and Applied Linear Algebra.* Society for Industrial and Applied Mathematics, 718 pp.

Migliorini, S., C. Piccolo, and C. Rodgers, 2008: Use of the information content in satellite measurements for an efficient interface to data assimilation. *Mon. Wea. Rev.,* **136,** 2633–2650.

Nocedal, J., and S. Wright, 2006: *Numerical Optimization.* 2nd ed. Springer, 664 pp.

Rodgers, C., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice.* World Scientific Publishing Co., 240 pp.

Tarantola, A., 2005: *Inverse Problem Theory and Methods for Model Parameter Estimation.* Society for Industrial and Applied Mathematics, 342 pp.