Matched Filter based on the Radiative Transfer Model for CO_2 estimation from PRISMA hyperspectral data

N. Acito, M. Diani, M. Alibani and G. Corsini, IEEE *Members*

***Abstract*— The rapid growth of hyperspectral satellite missions and the subsequent availability of hyperspectral images have encouraged the remote sensing community to investigate their potential in estimating the concentration of gases, as CH_4 and CO_2, which are related to the greenhouse effect. Though satellite hyperspectral sensors are not specifically designed for this purpose, they are expected to complement more specific satellite missions, such as NASA's OCO-2 and OCO-3, both in terms of enriched temporal sampling and improved spatial resolution.**

In this work, we present a new method to estimate the column-averaged dry-air mole fraction of CO_2 from hyperspectral data on a per-pixel basis. The method, which is here tailored to PRISMA images, leverages the spectral radiance samples collected in the SWIR spectral region around the CO_2 absorption band at 2000 nm. By assuming a linear model to describe the dependence of the observed radiance on the CO_2 concentration, the estimation problem is reduced to *matched filtering* and can be effectively implemented in compliance with the low computational burden required to perform a pixel by pixel analysis.

The performance of the presented method is investigated by means of a rigorous, physically based simulator that accurately reproduces the at-sensor radiance allowing one to check the validity of the assumptions and to assess the algorithm accuracy. The results show that the presented algorithm outperforms a benchmark CIBR based approach, which has been proposed in the literature to get fast per-pixel estimates of CO_2 concentration.

***Index Terms*—satellite hyperspectral data, CO2 retrieval, atmospheric parameters estimation.**

I. Introduction

NATURAL phenomena and anthropogenic activities release carbon dioxide (CO_2) and methane (CH_4) in the atmosphere. CO_2 and CH_4 are the most relevant climate-forcing agents and are closely related to the climate changes and global warming ([1]). Therefore, their characterization on both a global and local scale is of fundamental importance for monitoring the evolution of climate changes and planning the actions necessary to avoid or limit the harmful effects on the planet.

CO_2 and CH_4 absorb the radiation that propagates in the atmosphere in specific spectral windows located in the Short Wave InfraRed (SWIR) portion of the electromagnetic spectrum. The absorption features are directly related to the column-averaged dry-air mole fraction of CO_2 (XCO_2) and CH_4 (XCH_4) that can be monitored from the space by dedicated sensors.

Limiting to XCO_2 estimation, which is the focus of this work, CarbonSpec ([2]) on board of TANSAT and Orbiting Carbon Observatory-2 and -3 (OCO-2 [3] and OCO-3 [4]) are examples of satellite systems currently adopted. Both are grating spectrometers capable of measuring the radiance in the SWIR CO_2 absorption window with high spectral resolution ($\sim 0.1\ nm$) and spatial resolution of about two kilometers.

The recent availability of satellite hyperspectral imaging sensor such as PRISMA ([5], [6]) and EnMAP ([7]) operating in the Visible Near InfraRed (VNIR) and SWIR spectral range with spectral resolution of 10 nm and spatial resolution of 30 m, offers the opportunity to increase CO_2 monitoring capabilities. While they do not have the high spectral resolution of grating spectrometers and are not optimized for CO_2 mapping, they sample the CO_2 absorption window at 2000 nm with about twenty spectral channels and can potentially be used to quantify XCO_2 in the atmosphere at a finer spatial scale. This has motivated research activities aimed at analyzing the potential of the current generation of satellite hyperspectral imaging sensor for CO_2 retrieval ([8], [9], [10]) as well as this work where, with specific reference to the PRISMA sensor, a new algorithm for CO_2 estimation is proposed and a simulation-based sensitivity analysis is conducted to assess its performance.

A. Related works

The methods adopted for CO_2 retrieval from satellite spectrometers (e.g. CarbonSpec and OCO-3) mainly exploit spectroscopy and can be classified as physics-based approaches. They rely on the physical relationships that model the radiative transfer in the atmosphere (Radiative Transfer model – RTM, [11]). RTM relates the at-sensor radiance to the reflectance of the observed surface through radiative quantities that account for the parameters (e.g. XCO_2 and XCH_4) pertaining to the atmospheric constituents. Algorithms based on the RTM process the measured radiance in order to solve an inverse problem whose core is the estimation of the parameters that characterize the atmosphere at the time of the acquisition.

"This work is supported by the Italian Space Agency (ASI) under Grant DC-UOT-2019-061."

Corresponding author: Nicola Acito e-mail: nicola.acito@ unipi.it.

N. Acito, G. Corsini and M. Alibani are with the Department of Information Engineering – University of Pisa, Via Caruso 16, 56122 Pisa, Italy.

M. Diani is with Accademia Navale di Livorno – Italian Navy, Viale Italia 172,

Table 1. PRISMA sensors specifications

Swath/FOV	30 km / 2.77°
Ground Sampling Distance	30 m
Spectral Range	VNIR: 400 – 1010 nm SWIR: 920 – 2505 nm
Spectral Width (FWHM)	≤ 12 nm
Radiometric Quantization	12 bit
VNIR SNR	> 200:1
SWIR SNR	> 100:1
MTF@Nyquist Freq.	VNIR/SWIR along Track > 0.18 VNIR/SWIR across Track > 0.34
Spectral Bands	66 VNIR 173 SWIR

One of the most effective physics-based technique is the Iterative Maximum A Posteriori Differential Optical Absorption Spectroscopy (IMAP-DOAS), that was proposed in [12] to jointly estimate the atmospheric content of different absorbers such as water vapor (H_2O), CH_4 and CO_2. IMAP-DOAS is an iterative algorithm that basically has the form of a Gauss-Newton optimization problem, and combines an accurate physics model of the atmospheric absorption due to the constituents of interest, a polynomial model for the subtended reflectance and a statistical characterization for the noise and the model inaccuracy. Although it was designed for the grating spectrometer SCIAMACHY [12] that samples the incoming radiation in the spectral window of interest with high spectral resolution, IMAP-DOAS formulation has been extended to imaging hyperspectral sensors with fewer bands (i.e. coarser spectral resolution) and finer spatial resolution. In [13] a slightly modified version of the algorithm was proposed for CH_4 concentration retrieval from data acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). In a more recent work [14] IMAP-DOAS was extended for the retrieval of multiple gas species, including CH_4, CO_2, and H_2O, and applied to data acquired by the AVIRIS-Next Generation sensor (AVIRIS-NG). In [8] Cusworth et al. presented the results obtained by applying IMAP-DOAS for quantifying power plant carbon dioxide emission from data acquired by both airborne (AVIRIS-NG and Global Airborne Observatory-GAO) and spaceborne hyperspectral data (PRISMA). IMAP-DOAS performance are in general quite sensitive to the spectral reflectance of the surface observed ([13]). Furthermore, as stated by several authors ([15], [16]), the iterative form of the algorithm makes it computationally expensive thus preventing real-time mapping and application to large scenes such as those typically acquired by satellite sensors. For this reason, IMAP-DOAS is generally used for estimating CH_4 and/or CO_2 concentrations in small areas surrounding known point source emitters ([17]).

A physics-based algorithm not suitable for data acquired by grating spectrometers but quite effective when applied to data acquired by hyperspectral sensors ([9], [10]) is the Continuum Interpolated Band Ratio (CIBR). It is based on the correlation between the values of the parameter characteristic of a given atmospheric absorber and the depth of the absorption feature due to that constituent. CIBR is computationally efficient but its estimation performance is quite sensitive to various noise sources typically encountered in real data.

To break free from the complexity of the physics model and achieve higher computational efficiency, data-driven methods have been developed for application to hyperspectral data. Most of them are matched filter (MF)-based algorithms ([15]-[17], [18]-[20]) that can quantify greenhouse gas enhancement over large scenes in a computationally efficient way. Although MF-based algorithms have been applied mainly for the characterization of CH_4 point sources ([16], [18]-[20]), they also apply to the case of CO_2 emitters.

MF-based algorithms linearize the RTM so as to view the at-sensor spectral radiance as the sum of two vectors: the first represents the background ($\boldsymbol{L}_0$) while the second is the contribution due to gas enhancement that is modelled as $\alpha\boldsymbol{t}$ where $\boldsymbol{t}$ is the target spectrum and α is the gas concentration enhancement over the background. $\boldsymbol{L}_0$ is assumed as a Gaussian random vector whose mean and covariance matrix are estimated from the image pixels (or in a neighborhood of the test pixel). $\boldsymbol{t}$ is obtained as the product of the background mean and a known unit enhancement vector denoting the change in transmittance corresponding to a change in concentration of the gas of interest. So, the problem reduces to the estimation of α that is obtained by projecting the data on the subspace addressed by $\boldsymbol{t}$ (after whitening by the background covariance matrix).

In contrast to physics-based techniques, MF-based algorithms are specifically conceived to estimate the emission enhancement (with respect to the background) from point sources and are not suitable to obtain pixel-by-pixel estimation of the gas concentration.

In this work, we propose a new algorithm that estimates XCO_2 from satellite hyperspectral data on a per pixel basis. It is a physics-based approach that, as opposed to IMAP-DOAS, separates the estimation of the unknowns so as to avoid the need for iterative processing. The proposed solution recasts the XCO_2 estimation problem in a linear form that allows the use of the MF-based approach. So, it exploits the strengths of both previously summarized approaches by allowing a per-pixel estimation of XCO_2 with a low computational burden. The algorithm is derived with specific reference to the PRISMA hyperspectral sensor (whose specifications are in Table 1). However, it can be generalized to any satellite hyperspectral sensors operating in the VNIR-SWIR spectral range. To discuss the performance of the proposed technique we perform an extensive analysis on a large data set of simulated images obtained considering the specifications of the PRISMA sensor.

The rest of the paper is organized as follows. In Section II, we describe the algorithm and discuss the assumptions made in its derivation. Then, in Section III, we provide a description of the simulation strategy for the generation of the image data set adopted to analyze the performance of the algorithm. Results of such an analysis are provided in Section IV. An example on real data is presented in Section V. Finally, we draw the main conclusions of the work and underline the open issues by tracing the route for the future developments.

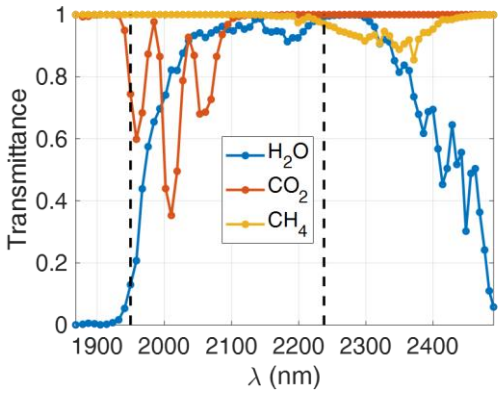


Figure 1. Spectral transmittance due to H_2O, CO_2 and CH_4 as a function of the central wavelengths of PRISMA channels in the SWIR spectral range (channels from 150 to 230).

II. Algorithm Description

This Section introduces the RTM and the main hypotheses made and provides the detailed description of the proposed algorithm. Before delving into the mathematical formulation of the problem, let us look at Figure 1 which provides an overview of the atmospheric absorption phenomena in the SWIR spectral range. Figure 1 shows, in separated plots, the contributions of H_2O, CH_4 and CO_2 to the atmospheric transmittance. These three contributions, which act together on the incoming radiance, are shown with reference to the PRISMA channels from 150 to 230 (i.e. they were convolved with the spectral response functions of the considered channels of PRISMA and sampled at the nominal central wavelengths). Figure 1 shows that CH_4 and CO_2 absorb the radiation into non-overlapping spectral windows. Conversely, H_2O interacts with both CH_4 and CO_2 making it more difficult to estimate their amount in the atmosphere. In the following, the estimation of CO_2 is approached by considering the spectral channels from 159 (1950 nm) to 194 (2237 nm) so as to minimize the influence of CH_4. The spectral range to which we refer is highlighted in Figure 1 by the two dashed vertical lines. Figure 1, also provides evidence that, as stated in Section I, the number of PRISMA samples falling within the range of interest is much lower than the hundreds of spectral samples measured by high resolution spectrometers (e.g. OCO-3 and CarbonSpec) specifically conceived for CO_2 monitoring.

A. Radiative Transfer Model (RTM)

According to the RTM in [11] and [21] and assuming the observed surface as Lambertian, the at-sensor radiance vector $\mathbf{L}\epsilon R^{N_B}$ for each pixel and for the N_B sensor channels of interest, is expressed as:

$$\mathbf{L} = \boldsymbol{A}(\boldsymbol{\Gamma}) \cdot \boldsymbol{\rho} + \mathbf{L}_p(\boldsymbol{\Gamma}) + \boldsymbol{N}$$
$$\boldsymbol{A}(\boldsymbol{\Gamma}) = \frac{\mathbf{E}cos(\vartheta)}{\pi} \cdot \boldsymbol{\tau}(\boldsymbol{\Gamma}) \quad (1)$$

where the symbol '·' indicates the element-wise product. Eq. (1) relates $\mathbf{L}$ to the surface spectral reflectance $\boldsymbol{\rho}$ and the spectral radiomentric quantities $\mathbf{E}$, $\boldsymbol{\tau}$ and $\mathbf{L}_p$ that consider the effects of the atmosphere and solar irradiance. More precisely, $\mathbf{E}$ is the solar spectral irradiance at the top of the atmosphere, ϑ is the solar zenith angle, $\boldsymbol{\tau}$ considers the direct and diffuse solar radiation in both the downward and upward path and $\mathbf{L}_p$ is the intrinsic upwelling atmospheric spectral radiance. $\boldsymbol{N}$ is the noise modelled as a zero-mean random vector.

It is important to note that, the model in eq. (1) (unlike [11] and [21]) does not include the contributions due to the background (adjacency effects) and the spectral spherical albedo because they are practically negligible in the spectral range of interest.

The radiometric quantities can be obtained by dedicated Radiative Transfer Codes (RTC) as a function of a set of parameters ($\boldsymbol{\Gamma}$ in eq. (1)) characterizing the acquisition geometry and the atmosphere. Throughout the paper we will refer to Modtran (ver. 5. [22]) which is one of the most popular commercial RTC.

$\boldsymbol{\Gamma}$ can be divided into three subsets:

- $\boldsymbol{\Gamma_g}$ that comprises all the acquisition geometry parameters such as the solar zenith (ϑ) angle, the sensor viewing angle, the terrain elevation and slope;
- $\boldsymbol{\Gamma_s}$ that includes the parameters that define the scattering effects such as the aerosol model and the visibility range (V);
- $\boldsymbol{\Gamma_a}$ that refers to the atmospheric absorbers which, in the considered case, are H_2O quantified in terms of column water vapor concentration (w) usually measured in g/cm^2 and CO_2 quantified in terms of XCO_2 (hereinafter denoted as x) generally measured in parts per million (ppm).

In practice, the parameters $\boldsymbol{\Gamma_g}$, except for the terrain elevation and slope, are known as they can be obtained from the on-board instrumentation. The terrain elevation and slope are known if a Digital Elevation Model (DEM) is available. When a DEM is not provided (as in the case of PRISMA data) the monitored surface is generally considered flat and at sea level. The scattering effects parametrized by $\boldsymbol{\Gamma_s}$ are featureless in the spectral domain (i.e. have a low pass behavior) and have a more significant impact at shorter wavelengths. Therefore, SWIR channels are less affected by scattering. This motivated the simplifying assumption made in this work that the variability of scattering effects in the considered spectral range is quite small so as to set standard values for the parameters $\boldsymbol{\Gamma_s}$. Specifically, with reference to the parameters adopted by Modtran, we consider the rural aerosol model and the standard value $\bar{V} = 23\ km$ for the visibility range.

The column water vapor concentration w generally varies within the monitored scene. A pixel-by-pixel estimation $\hat{w}$ of w can be obtained by algorithms ([22]-[25]) acting on the spectral windows where absorption effects are exclusively caused by H_2O (e.g. around 940 nm). In the rest of the paper, we will refer to the well-known Atmospheric Pre-corrected Differential Absorption (APDA) algorithm proposed in [22].

B. RTM based MF algorithm (RTM-MF) for XCO_2 estimation

Considering the previous observations, the RTM in eq. (1) is rewritten as:

$$\mathbf{L} = \boldsymbol{A}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x) \cdot \boldsymbol{\rho} + \mathbf{L}_{p,\bar{\boldsymbol{\Gamma}}}(\hat{w}, x) + \boldsymbol{N} \tag{2}$$

where $\bar{\boldsymbol{\Gamma}}$ groups the known parameters, the subscript indicates that the radiometric quantities refer to $\bar{\boldsymbol{\Gamma}}$ and their dependence on x and $\hat{w}$ is made explicit.

According to the Beer-Lambert absorption law $\boldsymbol{A}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x)$ can be expressed as:

$$\boldsymbol{A}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x) = \boldsymbol{\alpha}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}) \cdot e^{-\boldsymbol{\beta}_{\bar{\boldsymbol{\Gamma}}} x} \tag{3}$$

where $\boldsymbol{\alpha}_{\bar{\boldsymbol{\Gamma}}}(\hat{w})$ addresses the contributions of the solar radiation and the atmosphere in absence of CO_2 and $\boldsymbol{\beta}_{\bar{\boldsymbol{\Gamma}}}$ takes into account the CO_2 absorbance and the path length. Using a first-order Taylor series expansion around a given value x_0, eq. (3) can be linearized as:

$$\boldsymbol{A}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x) \cong \boldsymbol{A}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) - \boldsymbol{A}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) \cdot \boldsymbol{\beta}_{\bar{\boldsymbol{\Gamma}}} \cdot \Delta x \tag{4}$$

with $\Delta x = x - x_0$. Without loss of generality, we could consider for x_0 the current global value of XCO_2, i.e.: $x_0 = 415\ ppm$ ([26]).

In the SWIR spectral range, $\mathbf{L}_{p,\bar{\boldsymbol{\Gamma}}}(\hat{w}, x)$ is much lower than $\boldsymbol{A}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x)$ and it is neglected in most practical applications. Here we do not ignore the contribution of the upwelling radiance but we assume its variation is negligible around x_0 so as to set $\mathbf{L}_{p,\bar{\boldsymbol{\Gamma}}}(\hat{w}, x) = \mathbf{L}_{p,\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0)$.

Consequently, and assuming for the moment that $\boldsymbol{\rho}$ is known, eq. (2) is rewritten as:

$$\begin{gathered} \mathbf{L} = \boldsymbol{c}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) + \boldsymbol{t}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) \cdot \Delta x + \mathbf{L}_{p,\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) + \boldsymbol{N} \\ \textit{with} \\ \boldsymbol{c}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) = \boldsymbol{A}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) \cdot \boldsymbol{\rho} \\ \boldsymbol{t}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) = -\boldsymbol{c}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) \cdot \boldsymbol{\beta}_{\bar{\boldsymbol{\Gamma}}} \end{gathered} \tag{5}$$

that reformulates the approached problem as a linear estimation problem whose Minimum Mean Square Error (MMSE) solution is:

$$\begin{gathered} \widehat{\Delta x} = \frac{\boldsymbol{t}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0)^T \boldsymbol{Y}(\hat{w}, x_0)}{\|\boldsymbol{t}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0)\|_2^{\,2}} \\ \textit{with} \\ \boldsymbol{Y}(\hat{w}, x_0) = \mathbf{L} - \boldsymbol{c}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) - \mathbf{L}_{p,\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0) \end{gathered} \tag{6}$$

where ^{T} is the vector transpose operator and $\| \ \|_2$ is the l_2-norm. Eq. (6) can be interpreted as the application of the MF $\frac{\boldsymbol{t}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0)}{\|\boldsymbol{t}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0)\|_2^{\,2}}$ to the observation vector $\boldsymbol{Y}(\hat{w}, x_0)$.

Of course, the estimate $\hat{x}$ of XCO_2 is obtained as:

$$\hat{x} = x_0 + \widehat{\Delta x} \tag{7}$$

It is worth noting that, eq. (6) is applied on a pixel-by-pixel basis. However, when we can assume that the parameter x is constant in an MxM pixels neighborhood Ω of the pixel under test, a more robust estimator can be obtained as:

$$\widehat{\Delta x} = \frac{\sum_{k \in \Omega} \boldsymbol{t}_{\bar{\boldsymbol{\Gamma}}}^{(k)}\left(\hat{w}^{(k)}, x_0\right)^T \boldsymbol{Y}^{(k)}\left(\hat{w}^{(k)}, x_0\right)}{\sum_{k \in \Omega} \left\|\boldsymbol{t}_{\bar{\boldsymbol{\Gamma}}}^{(k)}(\hat{w}^{(k)}, x_0)\right\|_2^{\,2}} \tag{8}$$

where $\boldsymbol{Y}^{(k)}$, $\boldsymbol{t}_{\bar{\boldsymbol{\Gamma}}}^{(k)}$ and $\hat{w}^{(k)}$ are the observation vector, the matched filter and the estimate of w for the k-th pixel of Ω, respectively.

Up to now, to derive the MF-based XCO_2 estimation algorithm we have assumed that $\boldsymbol{\rho}$ is known. Of course, such an assumption is not realistic and a possible way to deal with this problem is to estimate the surface reflectance from the data themselves. For this purpose, we exploit the idea proposed in two recently published papers ([21], [27]) where machine learning methods are used to directly estimate the spectral reflectance from the at-sensor radiance addressing the variability due to one or more parameters of the RTM (and the noise) without estimating them.

Following the mentioned approach, we define a parametric transformation that allows us to obtain the estimate $\hat{\boldsymbol{\rho}}$ of the spectral reflectance in each pixel from the observation $\tilde{\boldsymbol{\rho}}(x)$ which is obtained by inverting the model in eq. (2) with a fixed value of XCO_2 (e.g. x_0). To be more precise, let us start by defining the observation vector $\tilde{\boldsymbol{\rho}}(x)$ as:

$$\tilde{\boldsymbol{\rho}}(x) = \frac{\mathbf{L} - \mathbf{L}_{p,\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0)}{\boldsymbol{A}_{\bar{\boldsymbol{\Gamma}}}(\hat{w}, x_0)} \tag{9}$$

It can be seen as a rough estimate of the spectral reflectance which contains the effects of noise and does not take into account the actual value of XCO_2 (this is made explicit in eq. (9) through the dependence on x).

$\hat{\boldsymbol{\rho}}$ is obtained by applying the parametric estimation function g to $\tilde{\boldsymbol{\rho}}(x)$, i.e.:

$$\hat{\boldsymbol{\rho}} = g(\tilde{\boldsymbol{\rho}}(x); \boldsymbol{\Psi}) \tag{10}$$

with $\boldsymbol{\Psi}$ denoting the parameters of the estimation function. The response of the estimator to the predictors $\tilde{\boldsymbol{\rho}}(x)$ depends on the specific value of the parameters $\boldsymbol{\Psi}$ that must be determined through the learning process. Learning is performed on N_t training samples, made up of the predictors $\tilde{\boldsymbol{\rho}}_{\boldsymbol{t}}(x)$ and the corresponding reflectance vector $\boldsymbol{\rho}_t$ $(t = 1, \dots, N_t)$. They are obtained by simulation using the RTM in eq. (2), Modtran for the generation of the radiometric quantities and by exploiting

Table 2. List of the parameters related to the acquisition geometry and the scattering effects considered in the simulation

Acquisition Geometry	
Parameter	**Values** (lower bound : step size : upper bound)
Solar Zenith Angle (deg)	0 : 10 : 40
Sensor Viewing Angle (deg)	0 : 5 : 35
Terrain elevation (m)	0

Scattering	
Parameter	**Values**
Aerosol model (IHAZE in Modtran)	1 (Rural), 3 (Navy Maritime), 5 (Urban)
Visibility (km)	10:5:60 (lower bound : step size : upper bound)

the availability of a database of the spectral reflectances of different surfaces/materials. The learning process consists in minimizing with respect to $\boldsymbol{\Psi}$ the loss function in eq. (11) over the training samples.

$$\underset{t=1,\dots,N_t}{loss}\left[g\left(\tilde{\boldsymbol{\rho}}_{\boldsymbol{t}}(x);\boldsymbol{\Psi}\right),\boldsymbol{\rho}_t\right] = \sum_{t=1}^{N_t}\left\|\boldsymbol{\rho}_t - g\left(\tilde{\boldsymbol{\rho}}_{\boldsymbol{t}}(x);\boldsymbol{\Psi}\right)\right\|_2^2 + \phi\|\boldsymbol{\Psi}\|_F^2 \quad (11)$$

The loss function is composed of a data fidelity term that computes the square estimation error and a regularization term based on the Frobenious norm ($\|\ \ \|_F$) of $\boldsymbol{\Psi}$ which is introduced to avoid overfitting and increase robustness to the noise. According to the results in [21] and [27], we consider a multilinear parametric estimator:

$$\hat{\boldsymbol{\rho}} = \boldsymbol{\Psi}\begin{bmatrix}\tilde{\boldsymbol{\rho}}(x)\\ 1\end{bmatrix} \quad (12)$$

which gives the loss function in eq. (12) the form of a Ridge regression whose closed-form solution is:

$$\boldsymbol{\Psi} = \left[\sum_{t=1}^{N_t}\begin{bmatrix}\tilde{\boldsymbol{\rho}}_{\boldsymbol{t}}(x)\\ 1\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\rho}}_{\boldsymbol{t}}(x)\\ 1\end{bmatrix}^T + \phi\cdot\mathbf{I}\right]^{-1}\cdot\sum_{t=1}^{N_t}\begin{bmatrix}\tilde{\boldsymbol{\rho}}_{\boldsymbol{t}}(x)\\ 1\end{bmatrix}\boldsymbol{\rho}_{\boldsymbol{t}}{}^T \quad (13)$$

with $\mathbf{I}$ denoting the (N_B+1)x(N_B+1) identity matrix.
Details on training set generation are provided in Section III, here we would like to conclude with a summary of the RTM based MF algorithm (RTM-MF, hereinafter) proposed in this Section. First, for each image pixel, RTM-MF derives the estimation of $\hat{w}$ by applying the APDA algorithm to the channels falling in the H_2O absorption window around 940 nm. Then, it estimates the spectral reflectance ($\hat{\boldsymbol{\rho}}$) by applying eq. (12) to the observation vector $\tilde{\boldsymbol{\rho}}(x)$**.** The latter is obtained from the at-sensor radiance in the selected CO_2 absorption bands (Figure 1) by means of eq. (9). Finally, the MF based estimator in eq. (6) (or that in eq. (8)) is applied to each image pixel (or to its surrounding) in order to obtain the estimate $\hat{x}$ of XCO_2.

III. Simulation strategy

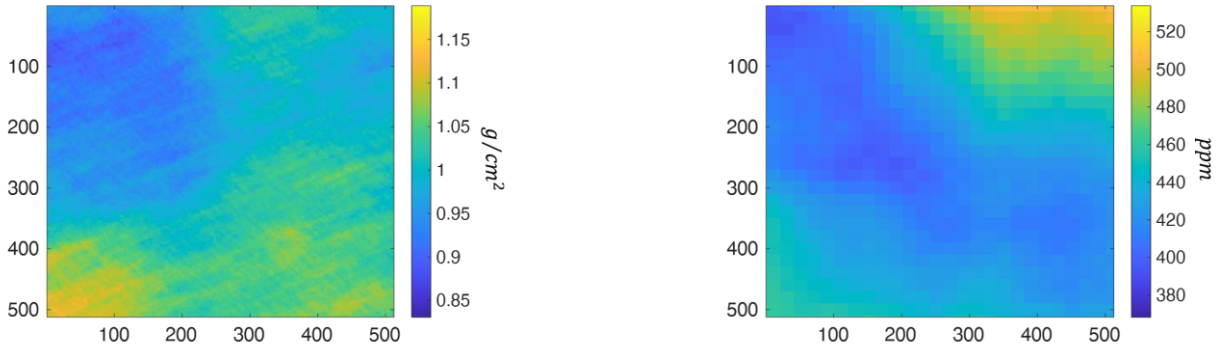
This Section describes the simulation strategy for the generation of the PRISMA images (PRISMA-like images, hereinafter) adopted for the performance analysis and presents the procedure followed to build the training set used in the learning process of the reflectance estimator introduced in Section II.

A. Simulation of PRISMA-like images

Note that, our simulator generates synthetic images covering the whole spectral range of the PRISMA sensor. However, only a small portion of the SWIR and NIR spectral regions is actually used. In fact, the RTM-MF procedure leverages the spectral channels falling between 1950 nm and 2237 nm, while to estimate *w* we exploit a small spectral region around 940 nm.

The simulation of each PRISMA-like image starts from a real reflectance image that is picked up from an AVIRIS-NG image repository made available by NASA Jet Propulsion Lab ([28]). AVIRIS-NG sensor covers the VNIR-SWIR spectral range from 380 nm to 2510 nm (with 5 nm spectral sampling) that completely overlaps the portion of the spectrum measured by PRISMA. In this work we use ten AVIRIS-NG reflectance images acquired during the India Campaign held in 2016. From each image we extracted non-overlapped tiles of 512x512 pixels (100 in total), discarding cloud covered areas and corrupted data. A detailed list of the characteristics (included the file identifiers and the number of tiles extracted) of the ten AVIRIS-NG images is given in Appendix I where the false color (FC) representation of one of the images is also shown along with the extracted tiles. The reflectance spectra in each image were low-pass filtered in order to remove residual noise and interpolated so as to achieve a 1 nm spectral sampling.

For a given reflectance image the atmospheric effects are simulated to obtain the at-sensor radiance. This task is performed by applying the RTM in [21], with the radiometric



(a) *(b)*

Figure 2. Examples of simulated w (a) and x (b) maps.

quantities obtained by specific Look-Up-Tables (LUTs) generated by Modtran (with 1 nm spectral sampling) for various combinations of the significant parameters. In all the simulations the "Mid-Latitude Summer" model of Modtran is adopted. Table 2 lists the parameters related to the acquisition geometry and the scattering effects and the corresponding values considered in the simulations.

As to the absorption effects both w and x are generated in terms of bidimensional (2D) maps with the same spatial extent as the image so as to consider their spatial variability within the scene. Specifically, with reference to w, first we generate a 2D map whose pixels are realizations of independent random variables uniformly distributed in the range $[0,1]$. Then, we introduce a spatial correlation by applying a 2D linear (zero phase) space invariant filter. The amplitude response of such a filter is defined according to the average power spectrum of the maps obtained by applying APDA to dozens of real PRISMA images. Finally, the generated spatially correlated map is rescaled in the range $[w_{LB}, w_{UB}]$ with w_{LB} and w_{UB} in turn randomly selected in the range $[0.1,5]$ $^{g}/_{cm^2}$.

A similar procedure is adopted to generate the maps of XCO_2. In this case the spatial correlation is obtained by applying the same linear filter of the previous case followed by a 2D moving window average filter so as simulate the higher spatial correlation expected in the x maps. The obtained correlated maps are rescaled in the range $[x_{LB}, x_{UB}]$, with x_{LB} and x_{UB} randomly selected in the range $[360,550]$ ppm.

Figure 2 (a) and (b) show an example of the generated maps for the parameters w and x, respectively.

Once the high spectral resolution radiance image is obtained, the at-sensor radiance measured by PRISMA is derived by applying the Spectral Response Functions (SRFs) of each channel and adding noise. The SRF of each channel is modelled as a Gaussian kernel centered at the nominal central wavelength and having width in agreement with the nominal Full Width Half Maximum (FWHM) of that channel. Both the mentioned parameters are known as they are included in the metadata provided with the PRISMA images.

Noise is generated according to the model in [29] as a spectrally uncorrelated Gaussian random vector with zero mean, and variance per band linearly dependent on the noise-free radiance in order to consider both the photon noise and the thermal noise contributions. Specifically, for the b-th channel of a given pixel, noise is obtained as a realization of a zero mean gaussian variable with variance:

$$\sigma_b^2 = \gamma_{ph,b} L_b + \gamma_{th,b} \quad (14)$$

In eq. (14), L_b is the noise-free radiance in the b-th channels, while $\gamma_{ph,b}$ and $\gamma_{th,b}$ are the parameters related to photon and thermal noise, respectively. To make noise simulation realistic, for both $\gamma_{ph,b}$ and $\gamma_{th,b}$ we use the average values estimated from dozens of real PRISMA image by the algorithm proposed in [29].

The simulation strategy has been used to simulate $N_{IM} = 840$ PRISMA-like images by selecting randomly the reflectance images, the combinations of the parameters in Table 2 and by generating the maps of w and x according to the previously described procedure. As an example, Figure 3 shows the FC representation, which is obtained by combining the channels corresponding to red, green and blue, of one of the generated images.

B. Simulation of the training set for the reflectance estimator

Regarding the learning of the reflectance estimator, the training set is obtained by a procedure similar to that previously

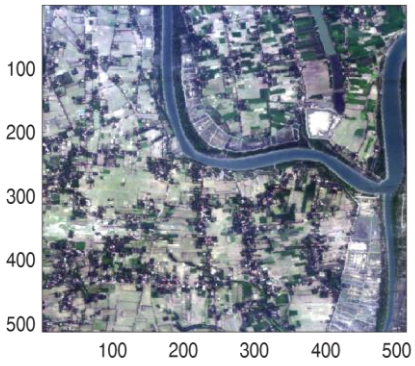


Figure 3. FC representation of one of the simulated PRISMA-like image.

described, except for the generation of the surface reflectances. The t-th reflectance spectrum $\boldsymbol{\rho}_t$ is generated by exploiting a library $\mathbf{D}$ which contains more than 1000 measured spectral reflectances of several man-made materials, vegetation, water, soils and minerals. $\mathbf{D}$ is obtained by merging the three available spectral libraries ASTER ([30]), USGS (version released in 2007, [31]) and ANGERS ([32]).

Each $\boldsymbol{\rho}_t$ is obtained by combining N randomly selected elements of $\mathbf{D}$ ($\mathbf{d}_i, i = 1, \dots, N$) and N randomly generated mixing coefficients a_i, $i = 1, \dots, N$.

$$\boldsymbol{\rho}_t = \sum_{i=1}^{N} a_i \mathbf{d}_i \tag{15}$$

The a_is are drawn according to the Dirichlet distribution so as to meet the sum-to-one and the non-negative constraints. N is also generated as a random integer in the range [1,5]. This strategy increases the spectral variability of the training set with the aim of encompassing the wide range of the spectral reflctances encountered in real applications.

For each $\boldsymbol{\rho}_t$ the corresponding radiance is obtained according to the above-described simulation strategy, with w and x randomly drawn in the range $[0.1,5]\ {}^{g}/_{cm^2}$ and $[330,550]\ ppm$, respectively. Random noise is added to the simulated radiance according to the model in eq. (14).

Of course, the obtained radiance is processed according to eq. (9) to obtain the t −th training pair $\{\boldsymbol{\rho}_t, \tilde{\boldsymbol{\rho}}_t(x)\}$.

By means of the described procedure we simulated $N_t = 5 \cdot 10^4$ training samples. It is worth noting that, the simulation of the training set (in particular, that of the reflectance data) is made different from that of the PRISMA-like images in order to have training and test sets as independent as possible.

IV. Results on PRISMA-like data

This Section is dedicated to the analysis of the performance of the proposed RTM-MF algorithm. This analysis is carried out on the test set consisting of $N_{IM} = 840$ PRISMA-like images generated by the simulation strategy described in Section III.

For completeness RTM-MF is compared with the CIBR algorithm that provides pixel-by-pixel estimates of XCO_2 by exploiting the correlation between XCO_2 values and the depth of the absorption feature in a specific measurement channel. More precisely, for a given measurement channel m within the CO_2 absorption window, CIBR computes the depth of the absorption feature relative to a local linear continuum as:

$$\tilde{d} = \frac{L_m}{w_1 L_{r1} + w_2 L_{r2}} \tag{16}$$

$r1$ and $r2$ are two channels at the left and right sides of the absorption window where the effect of CO_2 is negligible. The two weighting coefficients w_1 and w_2 are non-negative and sum to unity so as to make the denominator the linearly interpolated continuum at the location m of the absorption window.

According to the RTM in eq. (2), neglecting the upwelling radiance and assuming as constant the reflectance values in the three channels m, $r1$ and $r2$, the theoretical values of the absorption window depth as a function of x is expressed as:

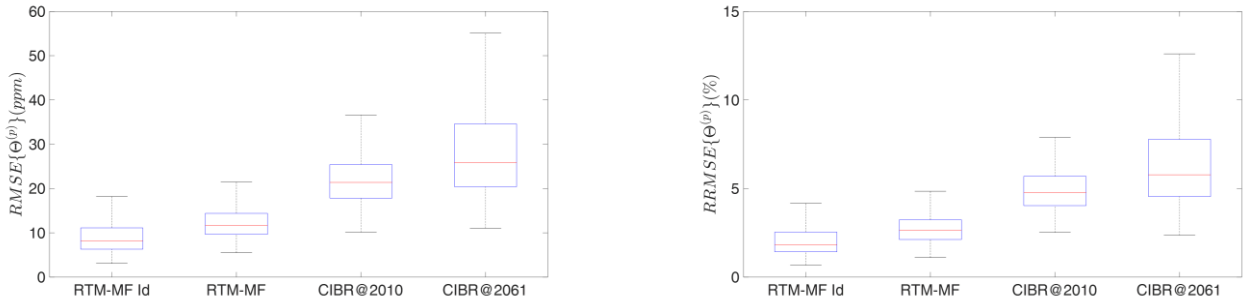
$$d(x) = \frac{A_{\bar{\Gamma},m}(\hat{w}, x)}{w_1 A_{\bar{\Gamma},r1}(\hat{w}) + w_2 A_{\bar{\Gamma},r1}(\hat{w})} \tag{17}$$

$d(x)$ can be computed by Modtran for different values of x so as to estimate the XCO_2 value in each pixel as:

$$\hat{x} = arg \min_x \left| \tilde{d} - d(x) \right| \tag{18}$$

In the tested implementation of the CIBR algorithm we consider $r1 = 163$ and $r2 = 178$ corresponding to the PRISMA channels at 1985 nm and 2110 nm, respectively. As to the measurement channel we tested two different choices: 1) $m = 166$ (2010 nm) as suggested in [10] and 2) $m = 172$ (2061 nm) as indicated in [9]. In the following we will refer to the two implementations as CIBR@2010 and CIBR@2061, respectively.

It is worth noting that, as in the case of RTM-MF, when we can assume x constant in a neighborhood Ω of the pixel under test, a more robust CIBR estimator can be obtained as:



(a) *(b)*

Figure 4. Distribution in terms of boxplot of $RMSE\{\Theta^{(p)}\}$ expressed in ppm (a) and $RRMSE\{\Theta^{(p)}\}$ in percent (b) for the algorithms: RTM-MF Id, RTM-MF, CIBR@2010 and CIBR@2061.

$$\tilde{d} = \frac{\sum_{k\in\Omega} L_m^{(k)}}{w_1 \sum_{k\in\Omega} L_{r1}^{(k)} + w_2 \sum_{k\in\Omega} L_{r2}^{(k)}} \quad (19)$$

where $L_m^{(k)}$, $L_{r1}^{(k)}$ and $L_{r2}^{(k)}$ are radiance values in the k-th pixel of Ω.

In all the experiments presented in this Section, for RTM-MF and the two CIBR algorithms we consider the robust versions with Ω defined as a 3x3 pixels neighborhood. In applying the RTM-MF algorithm the pivot point x_0 of the first order Taylor approximation of $e^{-\boldsymbol{\beta}\bar{\Gamma}x}$ (eqs. (3) and (4)) is set to 415 *ppm* according to the current global mean value of XCO_2 ([26]). Furthermore, to evaluate the impact of the estimation $\hat{\boldsymbol{\rho}}$ (eq. 10) on the performance of the RTM-MF we consider an ideal version of the algorithm (named "RTM-MF Id") in which the true value of $\boldsymbol{\rho}$ is used instead of $\hat{\boldsymbol{\rho}}$.

In all the discussed results we discarded the pixels with reflectance lower than 0.03 because the variations of the corresponding radiance values with respect to x are less than the sensitivity of any algorithm considered.

In order to test the performance of the various XCO_2 estimation methods we analyze the random variable $\Theta = \hat{x} - x$ which corresponds to the estimation error.

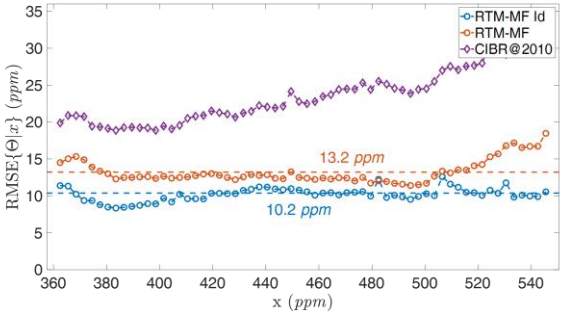
The first analysis is focused on the root mean square error (RMSE) and the root relative mean square error (RRMSE) of Θ obtained on each test image. For the p −th image RMSE and RRMSE are defined as:

$$\mathrm{RMSE}\{\Theta^{(p)}\} = \sqrt{\frac{1}{Np}\sum_{k=1}^{Np}\left[\hat{x}_k^{(p)} - x_k^{(p)}\right]^2}$$
$$\mathrm{RRMSE}\{\Theta^{(p)}\} = \sqrt{\frac{1}{Np}\sum_{k=1}^{Np}\left[\frac{\hat{x}_k^{(p)} - x_k^{(p)}}{x_k^{(p)}}\right]^2} \quad (20)$$

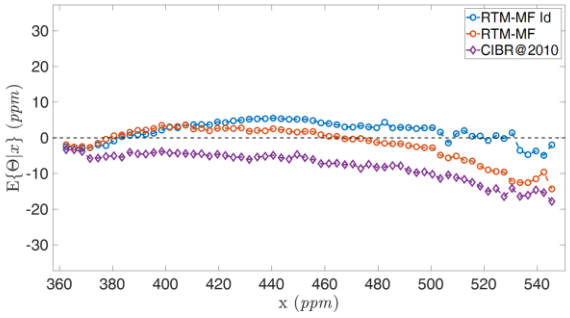
where $x_k^{(p)}$ and $\hat{x}_k^{(p)}$ are the true and the estimated values of XCO_2 in the k −th pixel of the image, respectively, and Np is the number of image pixels considered.

Figure 4 (a)-(b) show the distribution in terms of boxplot of RMSE (Figure 4 (a)) and RRMSE (Figure 4 (b)) all over the considered images for the four algorithms RTM-MF Id, RTM-MF, CIBR@2010 and CIBR@2061. Each box plot provides the following information about the distribution of the index to which it refers: the red line is the median, the edges of the box are the 25th and the 75th percentiles, respectively, and the whiskers are the minimum and the maximum values.

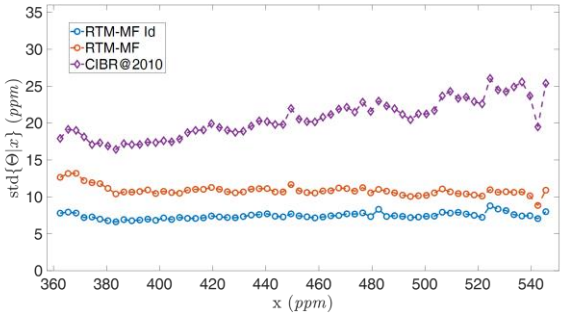
Figure 4 (a) shows that over the whole dataset, aside for the ideal case, the proposed RTM-MF achieves the best performance in terms of RMSE. On 75% of the images the values of RMSE obtained by RTM-MF are lower than 14.3 *ppm* with x ranging from 360 to 550 *ppm*. Conversely, CIBR@2010 provides RMSE values higher than 17 *ppm* on 75% of the analyzed images. The median RMSE attained by RTM-MF is of about 11 *ppm* while that of CIBR@2010 is approximately



(a)



(b)



(c)

Figure 5. (a) RMSE of the estimation error conditioned to the values of x ($\Theta|x$) estimated on the entire dataset for RTM-MF Id, RTM-MF and CIBR@2010. (b) Mean value of $\Theta|x$ and (c) standard deviation of $\Theta|x$ estimated on the entire dataset for RTM-MF Id, RTM-MF and CIBR@2010.

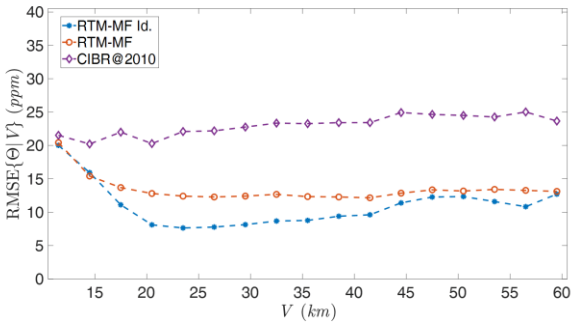


Figure 6. RMSE of the estimation error conditioned to the values of V ($\Theta|V$) estimated on the entire dataset for RTM-MF Id, RTM-MF and CIBR@2010.

21 *ppm*.

By comparing the distributions of the RMSE in the ideal case (RTM-MF Id) and in the more realistic case (RTM-MF), we can see that, as expected, RTM-MF Id performs better than RTM-MF. In the ideal case, the RMSE is below 11 *ppm* on the 75% of the images and its median value is 8.17 *ppm*. In general, by using the estimated value of $\boldsymbol{\rho}$, RTM-MF has a performance loss of about 3.5 *ppm* on average.

Figure 4 (a) also shows that CIBR@2010 outperforms CIBR@2061. This can be explained by considering that the strength of the CO_2 absorption is greater at 2010 nm than 2061 nm. It is worth noting that, the absorption feature at 2061 nm is less sensitive to water vapor than that at 2010 nm. Therefore, CIBR@2061 should be preferred to CIBR@2010 when water vapor cannot be estimated and compensated for. Since, in our experiments the estimate (via APDA) of the water vapor is available, in the rest of the article we only discuss the performance of CIBR@2010.

Figure 4 (b) confirms all the conclusions drawn. In particular, it shows that RTM-MF attains RRMSE values below 3% on most of the analyzed image with a median of about 2.6%. Conversely, RRMSE for CIBR@2010 is greater than 4% for most images. Figure 4 (b) also confirms the degradation of the performance of RTM-MF with respect to the ideal case. To give a numerical example, the median RRMSE for ideal case drops to 1.83%.

In deriving RTM-MF we used the first order Taylor approximation of $e^{-\boldsymbol{\beta}_{\bar{\Gamma}}x}$ in eq. (4) around the pivot point x_0. To test the validity of this hypothesis we analyze the estimation error Θ conditioned to XCO_2 ($\Theta|x$). In particular, Figure 5 a) shows the RMSE of $\Theta|x$ (RMSE$\{\Theta|x\}$) for all the range of x values considered in our simulation. The values of RMSE$\{\Theta|x\}$ achieved by RTM-MF are compared with those obtained by CIBR@2010 and those achieved in the ideal case. As to RTM-MF, Figure 5 a) shows that, in the ideal case (RTM-MF Id), RMSE$\{\Theta|x\}$ is quite stable as all index values are very close to 10.2 *ppm*. This leads to conclude that the error due to the first order Taylor approximation of the term $e^{-\boldsymbol{\beta}_{\bar{\Gamma}}x}$ in eq. (4) does not significantly affects the performance of the algorithm. As to the realistic case (RTM-MF), we can see that, for most of the x values, RMSE$\{\Theta|x\}$ is concentrated around 13.2 *ppm*. It deviates from 13.2 *ppm* when approaching the limits of the considered range, where, probably, the error introduced by the estimation of $\boldsymbol{\rho}$ has the greatest impact. However, for all x considered, RTM-MF provides values of RMSE$\{\Theta|x\}$ well below those achieved by CIBR@2010.

In addition to the RMSE analysis, Figure 5 b) and c) show the results concerning the bias and the standard deviation of the estimators. In particular, Figure 5 b) displays the values of the bias for each algorithm, which represent the average value of Θ conditioned to XCO_2 ($E\{\Theta|x\}$). On the other hand, Figure 5 c) shows the standard deviation of $\Theta|x$ ($std\{\Theta|x\}$) for the considered algorithms. The bias exhibited by all the algorithms is not constant and varies with the specific value of x. CIBR@2010 has negative bias for each value of x with values ranging from -3 ppm to -17.8 ppm, while the biases for RTM-MF Id and RTM-MF have both negative and positive values. Regardless of x, the absolute values of the bias obtained by CIBR@2010 are higher than those obtained by RTM-MF and RTM-MF Id. Notice however that the worse performance attained by CIBR@2010 in terms of RMSE is not solely due to the estimation bias. In fact, Figure 5 c) shows that CIBR@2010 exhibits a significantly higher standard deviation of $\Theta|x$ compared to both RTM-MF and RTM-MF Id. As expected, Figure 5 c) also shows that $std\{\Theta|x\}$ of RTM-MF Id is lower than that of RTM-MF.

The other hypothesis made in the algorithm derivation concerns the variability with respect to V of the radiometric quantities in the RTM. Specifically, it is assumed negligible and the standard value $\bar{V} = 23\ km$ is used. In order to evaluate the impact of this hypothesis on the performance of RTM-MF, we analyze the RMSE of Θ conditioned to V (RMSE$\{\Theta|V\}$) on the whole dataset. Results of such an analysis are summarized in Figure 6 where, for completeness, RMSE$\{\Theta|V\}$ obtained by CIBR@2010 and in the ideal case are also reported.

Figure 6 shows that for V greater than 20 km, RMSE$\{\Theta|V\}$ for RTM-MF is quite constant (approximately 13.3 *ppm*), while it increases for lower values of V and rises up to 20 *ppm* with $V = 10\ km$. This leads us to conclude that the hypothesis made can be considered valid for almost all the visibility conditions that occur in real applications. It is less reasonable for very low visibilities. In those cases, the assumption made has a negative effect on the performance of the RTM-MF algorithm. It is worth noting that, Figure 6 confirms the better results achieved by RTM-MF compared to CIBR@2010 for almost all the values of V considered. When very low V values occur, RTM-MF and CIBR@2010 have very similar performance.

Finally, we evaluate the performance of the algorithms with respect to the average value $\bar{\rho}$ of the reflectance in the spectral range of interest. Specifically, we analyze RMSE of Θ conditioned to $\bar{\rho}$ (RMSE$\{\Theta|\bar{\rho}\}$). Results of such analysis are reported in Figure 7. For all the algorithms RMSE$\{\Theta|\bar{\rho}\}$ decreases as $\bar{\rho}$ increases. It is an expected behavior as the impact of the noise tends to decrease as $\bar{\rho}$ increases. Figure 7 shows that RTM-MF outperforms CIBR@2010 for all the values of $\bar{\rho}$ occurred in the analyzed dataset. RMSE values for the two algorithms converge approximately to the same value

for high $\bar{\rho}$, while they differ significantly for lower values of $\bar{\rho}$ (<0.4) thus showing the higher noise robustness of RTM-MF.

Performance of the algorithms degrade significantly with $\bar{\rho} < 0.1$. Specifically, with $\bar{\rho} < 0.1$ RTM-MF attains RMSE$\{\theta|\bar{\rho}\}$ higher than 11 *ppm*, the value of the index rises up to 24 ppm for $\bar{\rho}$ in the order of 0.03 *ppm*.

We would like to conclude this section by providing information about the computational time of the proposed algorithm. By using a MATLAB implementation of the RTM-MF algorithm on a personal computer equipped with Intel Core i9 CPU at 2.5 GHz and with 64 GB of RAM, we found an average computing time of about $6\,s$ for a full image of 512x512 pixels. The MATLAB implementation of the CIBR algorithm takes about 4 *s* on average.

V. Example on real PRISMA data

To give an example on real data we considered a portion of the PRISMA image acquired on a subregion of the Po Valley-Italy on 15th July 2022. The false color representation of the considered image obtained by combining the channels corresponding to red, green and blue is shown in Figure 8. Such image has been chosen due to the availability of OCO-2 data acquired on the same day and in the same region. The longitude and latitude coordinates of OCO-2 data samples are highlighted in Figure 8 (red plots).

Figure 9 (a) and (b) show per pixel estimates $\hat{x}$ of XCO_2 obtained on the portion of the PRISMA image by RTM-MF and CIBR@2010 respectively. Both the algorithms are applied to PRISMA L1 Hyperspectral COregistered (HCO) data, i.e. the radiance data that are spatially and spectrally resampled in order to correct smile and keystone effects and spatial/spectral misalignments.

Figure 9 (a) and (b) show that RTM-MF and CIBR@2010 provide different results. Specifically, the estimated values obtained by RTM-MF are in average higher than those provide by CIBR@2010.

Figure 10shows the scatterplot of the XCO_2 values $\bar{x}$ for each sample of OCO-2 data and the corresponding estimates $\hat{x}$ obtained by RTM-MF (blue points) and CIBR@2010 (red points). To consider the different spatial resolutions of OCO-2 and PRISMA, the estimates $\hat{x}$ shown in Figure 10 are averaged over a 5x5 pixels neighborhood. Comparing XCO_2 estimates obtained by the two considered algorithms with those provided by OCO-2, it is evident that CIBR@2010 underestimates XCO_2 by providing an average value for the error $\hat{x} - \bar{x}$ of about -25 *ppm* and RMSE of 26.8 *ppm*. Results improve by adopting RTM-MF that provides an average error of -1.9 *ppm* and RMSE of 7.6 *ppm*.

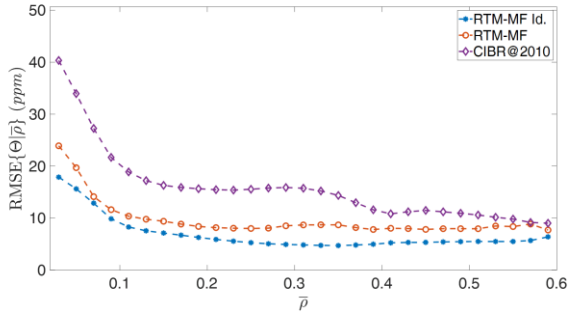


Figure 7. RMSE of the estimation error conditioned to the values of the spectral reflectance average value ($\Theta|\bar{\rho}$) estimated on the entire dataset for RTM-MF Id, RTM-MF and CIBR@2010.

We would like to remark that results presented in this Section are just an example of the application of the proposed algorithm and do not have the claim to be an experimental proof of its effectiveness. There are some specific aspects related to real data that are not considered in this work and deserve further analysis. For instance, we have not yet investigated the impact of a) spectral calibration errors which generally affect real hyperspectral data, b) instrumental fixed pattern noise (e.g. striping) and c) not correctly compensated for bad samples that are experienced in PRISMA images as discussed in [33].

Such analysis along with the possible strategies to mitigate the effects of those noise sources not modelled by eq. (2) are the main goal of the ongoing research activity.

VI. Conclusion

In this work we have proposed a new algorithm for XCO_2 retrieval from PRISMA hyperspectral data. The proposed algorithm is based on the RTM and, exploiting the estimates of the column water vapor and the spectral reflectance, reformulates XCO_2 estimation problem in a linear form. This form makes it possible to tackle the problem by means of a computationally efficient algorithm based on the matched filter. The algorithm performance has been evaluated on a large simulated dataset of PRISMA-Like images by testing the various hypotheses made in the derivation of the method. The analysis has included the comparison with the CIBR algorithm that is commonly adopted for estimating XCO_2 from

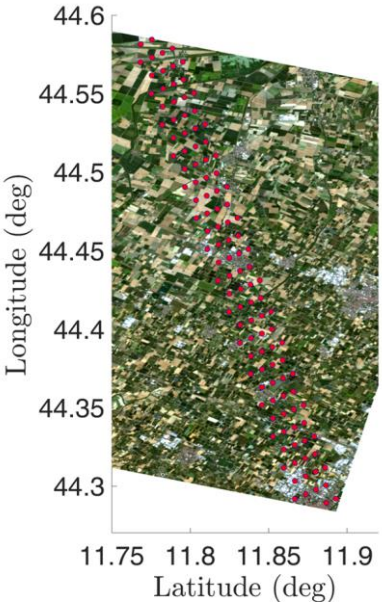
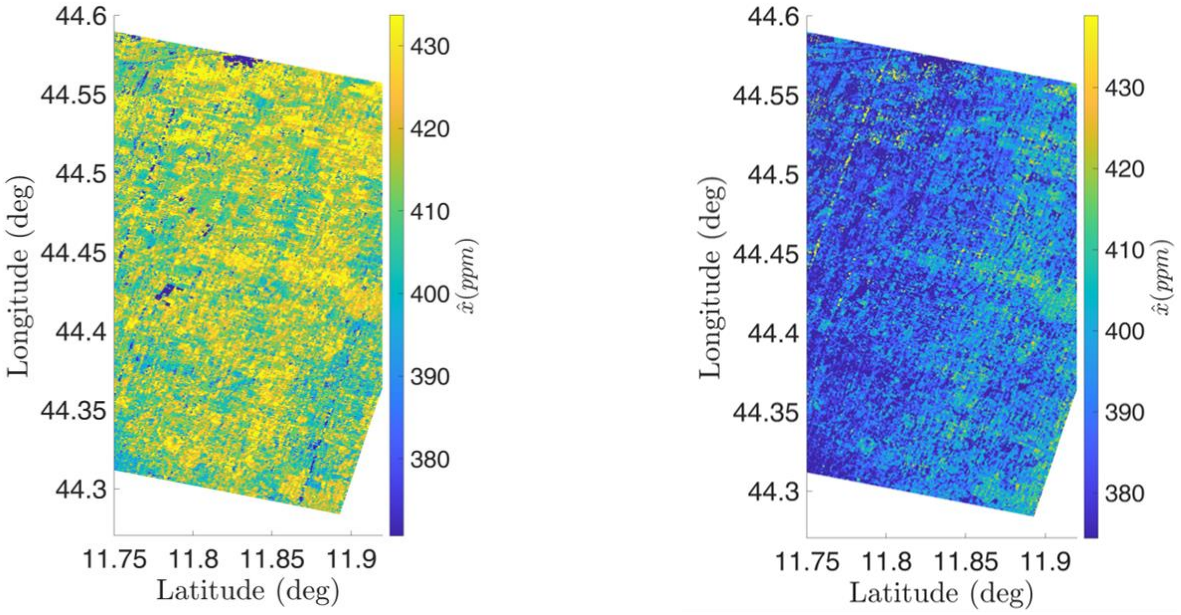


Figure 8. False color representation of the portion of the PRISMA image acquired on 15th July 2022 over the Po-Valley, Italy. The red circles indicate the coordinates of the OCO-2 data samples.



(a) *(b)*

Figure 9. Estimates of XCO_2 obtained on the considered PRISMA image by RTM-MF (a) and CIBR@2010 (b).

hyperspectral data. Results on simulated data have shown that the proposed solution provides better estimation accuracy than CIBR with a gain in terms of RMSE of about 10 *ppm*. Such a conclusion has been confirmed by an example on real PRISMA data where the estimates provided by RTM-MF and those obtained by CIBR have been compared with OCO-2 data. It worth nothing that, the conducted analysis has confirmed the limitations of hyperspectral imaging sensors in the estimation of XCO_2. The obtained errors are above those (in the order of few ppm) attained by processing data from very high spectral resolution sensors specifically designed for XCO_2 estimation (e.g. OCO-2 and OCO-3). However, the PRISMA mission is the precursor of a series of recently launched or forthcoming spaceborne spectroscopy missions with potential for CO2 mapping. The availability of estimation algorithms specifically conceived for hyperspectral imaging sensors is a pre-requite for the exploitation of the high spatial and temporal resolution observations that are or will be made available from the spaceborne missions mentioned above.

We would like to emphasize again that the analysis presented in this work does not include the effects of the spectral calibration errors that persist after smile and keystone noise correction, nor it does consider the impact of noise sources like instrumental fixed pattern noise and bad samples, which are frequently encountered in real hyperspectral images. The analysis of those effects, together with the possible solutions for their mitigation will be one of the objectives of our research activity in the near future.

APPENDIX

In this Section we provide details about the simulated data set. Particularly, in Table 3 we list the AVIRIS-NG images adopted in our simulation. For each image we provide the following information: the file identifier, the sensing date, the nominal GSD, the number of lines and samples, the number of tiles extracted having size 512x512 pixels. The total number of image tiles is 100.

In Figure 11, for one of the considered AVIRIS-NG images (file identifier: ang20160101t072436) we show the FC representations of the entire image and the extracted tiles.

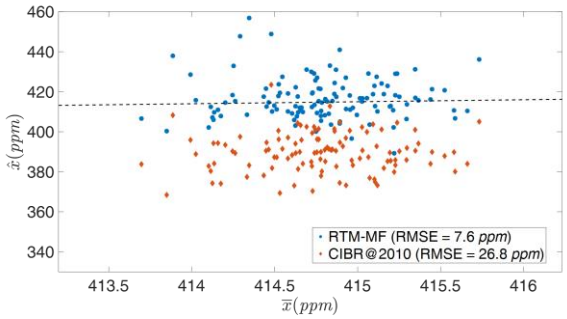


Figure 10. Scatterplot of the XCO_2 values $\bar{x}$ for each sample of OCO-2 data and the corresponding estimates $\hat{x}$ obtained by RTM-MF (blue points) and CIBR@2010 (red points).

ACKNOWLEDGMENT

This project is carried out using PRISMA products, © of the Italian Space Agency (ASI), delivered under an ASI License to use.

REFERENCES

[1]. H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama, "IPCC, 2022: Summary for Policymakers. In: Climate Change 2022: Impacts, Adaptation and Vulnerability," Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 3–33, doi:10.1017/9781009325844.001.

[2]. D. Yang, Y. Liu, Z. Cai, X. Chen, L. Yao, D. Lu, "First Global Carbon Dioxide Maps Produced from TanSat Measurements". *Advances in Atmospheric Sciences*, Vol. 35, pp. 621-623, June 2018.

[3]. https://ocov2.jpl.nasa.gov/, last accessed 11 April 2023.

[4]. A. Eldering, T. E. Taylor, C. W. O'Dell, R. Pavlick, "The OCO-3 mission: Measurement objectives and expected performance based on 1 year of simulated data," Atmos. Meas. Tech., Vol 12, No. 4, pp. 2341-2370, April 2019.

[5]. R. Guarini, R. Loizzo, C. Facchinetti, F. Longo, B. Ponticelli, M. Faraci, M. Dami, M. Cosi, L. Amoruso, V. De Pasquale, N. Taggio, F. Santoro, P. Colandrea, E. Miotti, W. Di Nicolantonio, "Prisma Hyperspectral Mission Products," Proceedings of the IEEE *International Geoscience and Remote Sensing Symposium*, 2018, IGARSS '18, July 2018, pp. 179-182.

[6]. R. Loizzo, M. Daraio, R, Guarini, F. Longo, R. Lorusso, L. Dini, E. Lopinto, "Prisma Mission Status and Perspective," Proceedings of the *IEEE International Geoscience and Remote Sensing Symposium*, 2019, IGARSS '19, July 2019, pp. 4503-4506.

[7]. S. Chabrillat, K. Segl, S. Goerster, M. Bell, L. Guanter, A. Schickling, T. Storch, H.-P- Honold, S. Fisher, "EnMAP pre-launch and start phase: Mission update," *IEEE International Geoscience and Remote Sensing Symposium*, 2022, IGARSS '22, July 2022, pp. 5000-5003.

[8]. D. H. Cusworth, R. M. Duren, A. K. Thorpe, M. L. Eastwood, R. O. Green, P. E. Dennison, C. Frankenberg, J. W. Heckler, G. P. Asner, C. E. Miller, "Quantifying global power plant carbon dioxide emissions with imaging spectroscopy," *AGU Advances Earth and Space Science*, Vol. 2, No. 2, pp. 1-8, June 2021.

[9]. V. Romaniello, C. Spinetti, M. Silvestri, M. F. Buongiorno, "A methodology fo CO2 retrieval applied to hyperspectral PRISMA data," *Remote Sensing*, Vol. 13, No. 22, pp. 1-13, 2021.

[10]. P. E. Dennison, "Fire detection in imaging spectrometer data using atmospheric carbon dioxide absorption," *Intern. Journ. of Remote Sensing*, Vol. 27, No. 14, pp. 3049-3055, 2006.

[11]. E. J. Ientilucci, S. Adler-Golden, "Atmospheric compensation of hyperspectral data: An overview and review of In-Scene and Physics-Based approaches," *IEEE Geosci. and Remote Sens. Magazine*, Vol. 7, No.2, pp. 31-50, June 2019.

[12]. C. Frankenberg, U. Platt, T. Wagner, "Iterative maximum a posterior (IMAP)-DOAS for retrieval of strongly absorbing trace gases: Model studies for CH4 and CO2 retrieval from near infrared spectra of SCIAMACHY onboard ENVISAT," *Atmospheric Chemistry and Physics*, Vol. 5, No. 1, pp. 9–22, Jan. 2005.

[13]. A. K. Thorpe, C. Frankenberg, D. A. Roberts, "Retrieval techniques for airborne imaging of methane concentration using spatial and moderate spectral resolution: application to AVIRIS," *Atmospheric Chemistry and Physics*, Vol. 7, No. 1, pp. 491-506, 2014.

[14]. A. K. Thorpe, C. Frankenberg, D. Thompson, R. M. Duren, A. d. Aubrey, B. D. Bue, R. O. Green, K. Gerilowski, T. Krigs, J. Borchardt, E. a. Kort, C. Sweeney, S. Conley, D. A. Roberts, P. E. Dennison, "Airborne DOAS retrieval of methane, carbon dioxide, and water vapor concentrations at high spatial resolution: application to AVIRIS-NG," *Atmospheric Chemistry and Physics*, Vol. 10, pp. 3833-3850, 2017.

[15]. M. D. Foote, P. E. Dennison, A. K. Thorpe, D. R. Thompson, S. Jongaramrungruang, C. Frankenberg, S. C. Joshi, "Fast and accurate retrieval of methane concentration from imaging spectrometer data using sparsity prior," *IEEE Trans. Geosci. Remote Sens.*, Vol. 58, No. 9, pp. 6480-6492, Sept. 2020.

[16]. L. Guanter, I. Irakulis-Loitxate, J. Gorroio, E. Sanchez-Garcia, D. H. Cusworth, D. J. Varon, S. Cogliatti, R. Colombo, "Mapping methane point emission with PRISMA spaceborne imaging spectrometer," *Remote Sens. of Environm.*, Vol. 256, 112671, pp. 1-14, 2021.

[17]. M. D. Foote, P. E. Dennison, P. R. sullivan, K. B. O'Neill, A. K. Thorpe, D. R. Thompson, D. H. Cusworth, R. Duren, S. G. Joshi, "Impact of scene-specific enhancement spectra on matched filter greenhouse gas retrievals from imaging spectroscopy," *Remote Sensi. of Environm.*, Vol. 264, 112574, pp. 1-12, 2021.

[18]. D. R. Thompson *et al.*, "Real-time remote detection and measurement for airborne imaging spectroscopy: A case study with methane," *Atmos. Meas. Techn.*, vol. 8, no. 10, pp. 4383–4397, Oct. 2015

[19]. D.R. Thompson, A.K Thorpe, C. Frankenberg, R. O. Green, R. Duren, L. Guanter, A. Hollstein, E. Middleton, L. Ong, S. Ungar, "Space-based remote imaging spectroscopy of the Aliso canyon CH4 superemitter," Geophys. Res. Lett., Vol. 43, pp. 6571–6578, 2016.

[20]. R. M. Duren, A. K. Thorpe, K. T. Foster, T. Rafiq, F. M. Hopkins, V. Yadav, B. D. Bue, D. R. Thompson, S. Conley, N. K. Colombi, C. Frankenberg, I. B. McCubbin, K. L. Eastwood, M. Falk, J. D. Herner, B. E. Croes, R. O. Green, C. E. Miller, "California's methane super-emitters," Nature, Vol. 575, pp. 180–184, 2019.

[21]. N. Acito, M. Diani, G. Corsini, "Learning-Based Approach for Atmospheric Compensation of VNIR Hyperspectral Data," IEEE Trans. on Geosci. and Remote Sens., Vol. 59, No. 5, pp. 4218-4232, 2021.

[22]. A. Berk, G P. Anderson, P. K. Acharya, L. S. Bernstein, L. Muratov, J. Lee, M. J. Fox, S. M. Adler-Golden, J. H. Chetwynd Jr., M. L. Hoke, R. B. Lockwood, J. A. Gardner, T. W. Cooley, P. E. Lewis, "Modtran5: a reformulated atmospheric band model with auxiliary species and practical multiple scattering," Proceedings of SPIE, Vol. 5425, 2004

[23]. D Schlaepfer, C. C. Borel, J. Keller and K. I. Itten, "Atmospheric Precorrected Differential Absorption Technique to Retrieve Columnar Water Vapor," Remote Sensing of Environment, Vol. 65, No. 3, pp: 353-366, Sept. 1998.

[24]. N. Acito, M. Diani, "Atmospheric Column Water Vapor Retrieval from Hyperspectral VNIR Data Based on Low-Rank Subspace Projection," IEEE Trans. on Geosci. and Remote Sens., Vol. 56, No. 7, pp. 3924-3940, 2018.

[25]. N. Acito, M. Diani, G. Corsini, "CWV-Net: A Deep Neural Network for Atmospheric Column Water Vapor Retrieval from Hyperspectral VNIR Data," IEEE Trans. on Geosci. and Remote Sens., Vol. 58, No. 11, pp. 8163-8175, 2020.

[26]. WMO Greenhouse gas bullettin, World Meteorological Organization, No. 18, pp. 1-10, October 2022. [Online]. Available: https://library.wmo.int/doc_num.php?explnum_id=11352.

[27]. N. Acito, "Improved Learning-Based Approach for Atmospheric Compensation of VNIR-SWIR Hyperspectral Data," IEEE Trans. on Geosci. and Remote Sens., Vol. 60, pp. 1-15, 2022.

[28]. https://avirisng.jpl.nasa.gov/index.html, last accessed 14 March 2023.

[29]. N. Acito, M. Diani, G. Corsini, "Signal dependent noise modelling and model parameters estimation in hyperspectral images," IEEE Trans. on Geosci. and Remote Sens., Vol. 49, No. 8, August 2011, pp. 2957-2971.

[30]. A.M. Baldridge, S.J. Hook, C.I. Grove and G. Rivera, "The ASTER Spectral Library Version 2.0," Remote Sensing of Environment, Vol 113, pp. 711-715, 2009.

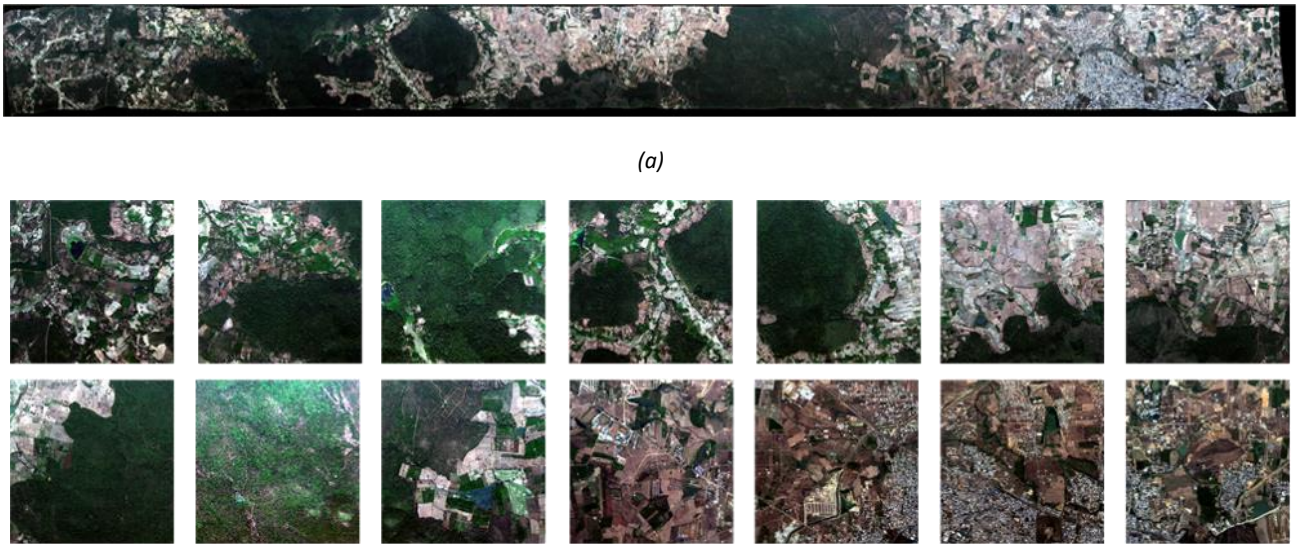
[31]. R.N. Clark, G.A. Swayze , R.A. Wise, K.E. Live, T.M. Hoefen, R.F. Kokaly, and S.J. Sutley, USGS Digital Spectral Library splib06a: U.S. Geological Survey Data Series 231, 2007.

[32]. Available on line at http://opticleaf.ipgp.fr/index.php?page.

[33]. N. Acito, M. Diani, M. Alibani, G. Corsini, "Automatic detection and correction of defective pixels in PRISMA hyperspectral data," IEEE Trans. on Geoscience and Remote Sensing, Vol. 60, pp. 1-15, Art no. 5538315, 2022.

Table 3. List of AVIRIS-NG images adopted for the simulated dataset. For each image the table gives: the name of the file corresponding to the image (File Identifier), the sensing date, the nominal GSD (in m), the number of lines and samples and the number of extracted tiles (512x512 pixels).

File Identifier	***Sensing date (day/month/year)***	***GSD (m)***	***Lines***	***Samples***	***No of extracted tiles***
ang20160101t072436	01/01/2016	5	7891	693	14
ang20160102t061817	02/01/2016	5	7836	694	14
ang20160102t064351	02/01/2016	5	7761	682	14
ang20160223t060914	23/02/2016	5	3591	711	4
ang20160223t062639	23/02/2016	5	3657	718	5
ang20160224t053740	24/02/2016	5	2670	671	4
ang20160224t055506	24/02/2016	5	4710	676	8
ang20160305t062710	05/03/2016	5	9177	756	12
ang20160305t063928	05/03/2016	5	9167	747	13
ang20160305t065158	05/03/2016	5	9086	768	12
Total no of extracted tiles					**100**



(a)

(b)

Figure 11. (a) FC representation of the AVIRIS-NG image whose file identifier is ang20160101t072436. (b) FC representations of the tiles (512x512 pixels) extracted from the image.