Cloud detection in Hyperspectral Images With Atmospheric column WAter vapor. Application to PRISMA and AVIRIS-NG images

Alexandre Alakian

Cloud detection in Hyperspectral Images With Atmospheric column WAter vapor. Application to PRISMA and AVIRIS-NG images

Alexandre Alakian, *Member, IEEE*

***Abstract*—The CHIWAWA (Cloud detection in Hyperspectral Images With Atmospheric column WAter vapor) cloud detection method is presented. It aims to detect thick and thin clouds in a single hyperspectral image covering the reflective spectral range [400-2500nm] in an unsupervised way. Its main originality is to combine the complementary information available in the ground reflectance image and the atmospheric column water vapor map (WV map). The method uses criteria based on reflectance to compute a mask of potential cloud pixels and refine it by exploiting the WV map. The use of the WV map is central to the method, as it has the greatest discriminating power. It therefore plays a role similar to that of thermal bands, which are very useful for cloud detection, but are absent in the reflective spectral range. Two specific processings are also proposed to detect snow and ice pixels and to reduce the altitude dependency of the WV map in mountain images. CHIWAWA performance is assessed on several hyperspectral images acquired with PRISMA and AVIRIS-NG sensors. Overall accuracy ranges in 97–100% for both sensors. When cloud cover is above 10%, F_1 score ranges in 94–100% for both sensors. For lower cloud covers, F_1 score ranges in 83–94% for PRISMA images and in 84–91% for AVIRIS-NG images. CHIWAWA may encounter difficulties in detecting thin clouds located over urban areas or when the contrast between the cloud and its surroundings is low in the WV map.**

***Index Terms*—atmospheric column water vapor, cloud, detection, hyperspectral, ice, reflectance, snow, unsupervised**

I. Introduction

Hyperspectral remote sensing allows the simultaneous acquisition of hundreds of narrow and contiguous spectral bands usually ranging from the visible to the short-wave infrared. The pixel spectra provided by such sensors can be exploited to extract information about the components of the studied scene (material constituents, gaseous and aerosol concentrations). Covering more than 50% of the global surface, clouds often affect hyperspectral images. Thick clouds can completely hide ground surface properties, while thin clouds allow access to the spectral properties of the surfaces below but by altering them. These cloudy areas need to be detected in order to facilitate the interpretation of the acquired images. Cloud masks therefore need to be created.

Cloud detection algorithms were developed for different sensors with various spectral ranges and spectral bands. Some methods use time series of images of the same area to calculate the cloud mask, either by comparison with a cloud-free reference image [1]–[4] or by detecting clouds as anomalies [5]. A multi-sensor approach can take advantage of the information from a sensor with a higher revisit [6]. This type of method is not easily applicable to hyperspectral images, as for each image studied there must be a cloud-free image with similar shooting conditions, and the surface must not have changed (e.g. seasonal variations).

Single-scene-based cloud detection methods mainly use spatial and spectral features to distinguish ground surfaces from clouds. Machine learning methods combined with feature analysis can be used, such as Markov random field (MRF) framework [7], [8], support vector machine (SVM) [9], neural network [10] and deep learning [11], [12]. Such methods require a large number of samples, taking into account the great variability of clouds in terms of shape, spatial extension and thickness variations within the same cloud, but also the variability of non-cloud surfaces. In the case of optically thin clouds, the surface reflectance beneath the cloud contributes significantly to the at-sensor radiance. The variability of these surfaces must also be taken into account, as well as the variability of cloud optical thickness, which has an impact on cloud transmittance and thus on ground contribution. In the end, sample diversity can be extremely high, making it difficult to detect thick and thin clouds with machine learning algorithms.

Clouds are characterized by a high apparent reflectance and a lower temperature than the underlying surface. For sensors with spectral bands in the reflective range [400-2500nm] and the thermal infrared range, cloud detection methods based on thresholding combinations of bands (single bands, band differences, band ratios or other combinations) and histograms have been developed, including ISCCP (International Satellite Cloud Climatology Project) [13], CLAVR (CLouds from the Advanced Very High Resolution Radiometer AVHRR, NOAA) [14], APOLLO (AVHRR Processing scheme Over cLouds, Land and Ocean) [15], ACCA (Cloud Cover Assessment Algorithm) [16] and Fmask [17], [18]. Fmask is the method implemented to detect clouds and associated shadows on Landsat images, with an average overall accuracy of 97%.

The temperature information enables effective discrimination, but is not accessible to sensors without thermal bands. Methods exploiting only bands in the reflective range were also developed. CDAG method [19] performs thresholding on band combinations to build a cloud mask, the thresholds used having been estimated from AVIRIS images. The Fmask

Manuscript submitted September 4, 2023.

Alexandre Alakian is with the Department of Information Processing and Systems, ONERA, The French Aerospace Lab, Université Paris Saclay, FR-91123 Palaiseau, France (e-mail:alexandre.alakian@onera.fr)

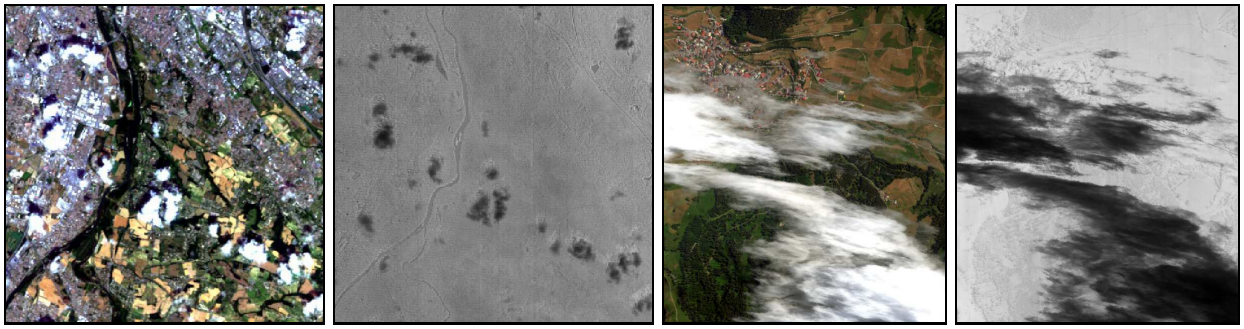

Fig. 1. RGB composition of a PRISMA image in France (left) and an AVIRIS-NG image in Switzerland (right), and corresponding WV maps. WV is lower at the location of clouds than around them: the CHIWAWA method proposes to use this information to detect thick and thin clouds. To improve visualization on the WV map, dark surfaces (mainly shadows and water) have been detected (see section III-B) and the average WV map value has been assigned to them.

method has been adapted to Sentinel2 [18]. In [20], a cloud detection method based on modified radiation transmittance is proposed and is assessed on Gaofen-2 multispectral images. Another way of detecting clouds is to retrieve cloud optical thickness from a dedicated look-up table (LUT) storing top-of-atmosphere radiances relative to a wide configuration of atmospheric scenarios including different acquisition geometries and cloud optical properties (cloud optical thickness, effective radius) [21], [22]. LUT radiances are calculated making use of a radiative transfer code.

Some specific spectral bands can be used to detect or characterize clouds, such as the 760 nm oxygen A-band [23], or the 1380 nm water vapor band which is dedicated to cirrus clouds [24]. In this paper, we will take a special look at atmospheric column water vapor, noted WV in the following, and see how it can be used to detect clouds. During atmospheric correction of a hyperspectral radiance image, WV is estimated for each image pixel using water vapor absorption bands around 940 nm or 1140 nm [25], [26]. For a given pixel, in the presence of a cloud, a variable proportion of the photons emitted by the sun are reflected by the cloud and cannot interact with the water vapor beneath the cloud (*i.e.* be absorbed), leading to an underestimation of the WV value for that pixel. As a result, the estimated WV value is generally much lower on cloud pixels than on other pixels, allowing clouds to be clearly observed on the WV map (see Fig. 1).

In this paper, the CHIWAWA method (Cloud detection in Hyperspectral Images With Atmospheric column WAter vapor) is presented. It aims to detect thick and thin clouds in a single hyperspectral image in the spectral range [400-2500nm] without supervision. The originality of the method lies in exploiting the complementarity between atmospheric water vapour, which until now has not been used to detect clouds, and ground reflectance. Ground reflectance is more suitable than apparent reflectance (or top of atmosphere reflectance), which is more commonly used, as it enables surface materials to be characterized more easily, given that shooting conditions and atmospheric effects are corrected.

Overall, CHIWAWA method is divided into two main phases.

- Phase 1: detection of potential cloud pixels. This phase mainly consists of detecting bright pixels while removing snow and ice pixels. The hyperspectral ground reflectance image is essentially used for this task. A processing dedicated to the detection of potential cloud pixels over green vegetation is also used.
- Phase 2: refinement of the potential cloud pixels mask using the WV map. This is achieved by combining two complementary processes: a first one is based on thresholding the WV map, a second one exploits the contrast between clouds and their environment in the WV map. A post-processing is then applied in order to remove false clouds.

The use of WV is central to phase 2: it has a high discriminating power, quite similar to the thermal bands that are often used for cloud detection. Upstream, pre-processings are applied to improve detection performance: dark pixels in hyperspectral image are identified and removed from WV map, and dependency between WV and altitude is reduced (for mountain images). The flowchart of CHIWAWA method is presented in Fig. 2.

The paper is organized as follows. The two main phases of CHIWAWA are described in sections II and III. In section IV, CHIWAWA is assessed on several hyperspectral images acquired with PRISMA [27] and AVIRIS-NG [28] sensors, and compared with other state-of-the-art methods. Finally, we give conclusion and highlight the potential directions of future research work.

II. Detection of potential cloud pixels (phase 1)

A. Overview

The goal of this phase is to identify pixels that are definitely not cloudy and thus create a mask of pixels that are potentially cloudy. First, bright pixels in the full spectral range are identified and collected in a dedicated binary mask m_{bright}. Snow and ice pixels, which are bright and can be mistaken for clouds, are also detected and removed from m_{bright}. Bright pixel detection does not retain pixels with low reflectance in the visible range, such as pixels containing green vegetation and thin clouds. A dedicated process is proposed to detect such pixels. The case of very bright pixels is also considered: these pixels are directly qualified as cloudy without going through phase 2.

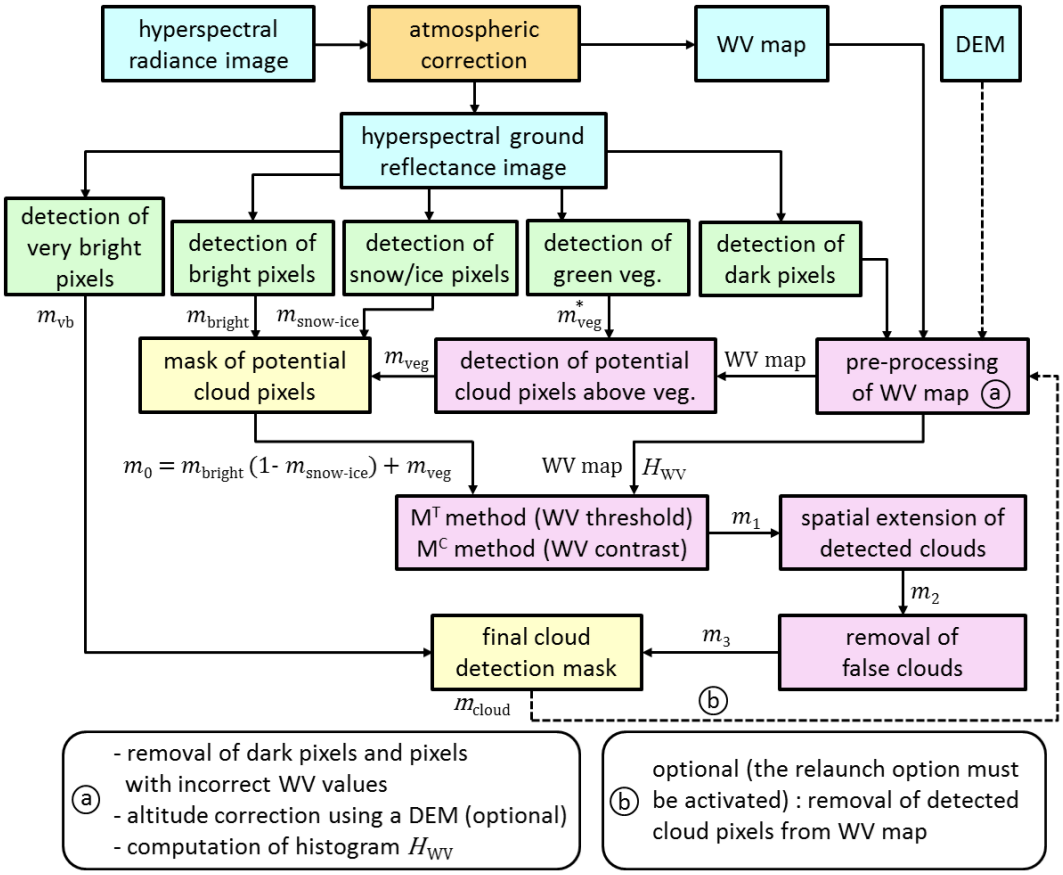

Fig. 2. Flowchart of CHIWAWA method. For the inputs, the processing requires a hyperspectral ground reflectance image in the spectral range [400-2500nm] and the associated WV map. These can be estimated from a single radiance image with atmospheric correction, or can be directly recovered as L2C products. The use of a digital elevation model (DEM) is only necessary for processing mountain images.

It is important to note that pixels excluded at this stage cannot be detected as cloudy in phase 2. On the other hand, it is necessary to eliminate as many non-cloudy pixels as possible in phase 1 in order to facilitate the processing applied during phase 2 and limit the number of false alarms. When choosing the thresholds, a compromise must therefore be made to combine these two requirements.

B. Detection of bright pixels

Clouds generally appear brighter than other surfaces in the full spectral range [400-2500nm]. Two thresholds T_{VNIR} and T_{SWIR} are applied to the spectral reflectance ρ_λ, one in the VNIR (Visible and Near-InfraRed) range [400-1000nm] and one in the SWIR (Short-Wave InfraRed) range]1000-2500nm], in order to remove non-cloudy pixels as much as possible.

$$\min_{\lambda \in I_{\text{VNIR}}} \rho_\lambda \geq T_{\text{VNIR}}, \tag{1}$$

$$\min_{\lambda \in I_{\text{SWIR}}} \rho_\lambda \geq T_{\text{SWIR}}. \tag{2}$$

Since spectral reflectance varies slowly outside gaseous absorption bands (absorption effects may persist even after atmospheric correction), estimating the minimum value of spectral reflectance in the VNIR range requires only the selection of a few low-noise bands, which must be sufficiently far apart and outside the gaseous absorption bands. The reasoning is the same for the SWIR range. The following sets of spectral bands are used: $I_{\text{VNIR}} = \{450, 550, 650, 800\}$ for VNIR and $I_{\text{SWIR}} = \{1600, 2200, 2350\}$ for SWIR.

T_{VNIR} and T_{SWIR} are sensor-dependent and may be changed depending on the type of scene. For a given sensor, T_{VNIR} and T_{SWIR} threshold values are estimated in a two-step process. Regions of interest from hyperspectral images showing various surfaces and/or clouds with different thicknesses are selected and annotated to obtain cloud maps (see section IV-A). In the first step, thresholds T_{VNIR} and T_{SWIR} are chosen in the set I = {0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.11, 0.12, 0.13, 0.15, 0.17, 0.20}: all possible combinations of thresholds (T_{VNIR}, T_{SWIR}) drawn from $I \times I$ are constructed. Each threshold combination is then applied to the annotated hyperspectral images. The resulting binary maps of bright pixels are compared with the cloud maps: the threshold combination is retained for the next step only if at least 95% of cloud pixels are detected as bright pixels for each of the images considered. In the second step, the full CHIWAWA method (phase 1 and phase 2) is applied to each combination selected in the first step. The optimal combination is the one that maximizes the average F_1 score over all the images used (see section IV-C for metrics definition). For the PRISMA sensor, around 440,000 cloud pixels out of a total of 4,000,000

pixels are used. Optimal threshold values are $T_{\text{VNIR}} = 0.07$ and $T_{\text{SWIR}} = 0.07$. For AVIRIS-NG sensor, around 800,000 cloud pixels out of a total of 4,200,000 pixels are used. Optimal threshold values are $T_{\text{VNIR}} = 0.10$ and $T_{\text{SWIR}} = 0.03$. However, in urban or mountainous areas, many surfaces appear bright and are not removed with such threshold values, especially T_{VNIR}, which can lead to poor detection performance with many false detections. For such situations, a value of 0.15 for T_{VNIR} significantly improves detection performance, even though thin clouds are less likely to be detected. The process returns the binary mask m_{bright}: the value of each pixel considered as bright is 1, and 0 elsewhere.

Note that lower clouds that are within the shadow of higher clouds may appear dark. Such cases are not specifically addressed. However, the relatively low thresholds used do not prevent the pixels associated with these dark clouds from being considered bright and therefore potentially cloudy.

C. Detection of very bright pixels

Some cloud pixels have very high reflectance levels in both the VNIR and SWIR ranges. These pixels can be directly identified as clouds by applying the criteria associated with inequalities (1) and (2) but with much higher thresholds T^{vb}_{VNIR} and T^{vb}_{SWIR}. However, inequality (2) is restricted to $\rho_{2350} \geq T^{vb}_{\text{SWIR}}$: ρ_{2350} is very useful for discriminating snow and ice (see sections II-D and IV-H). This processing is particularly useful when CHIWAWA is unable to detect clouds from the WV map, for example when the WV level is very low over the entire image.

A process similar to the one described in section II-B is used to estimate the optimal values for the thresholds used: $T^{vb}_{\text{VNIR}} = 0.40$ and $T^{vb}_{\text{SWIR}} = 0.12$ for PRISMA and AVIRIS-NG images. When studying urban areas with AVIRIS-NG, a value of 0.15 for T^{vb}_{SWIR} is preferred.

In the output, the mask of definitely cloudy pixels m_{vb} is produced, *i.e.* these cloudy pixels are kept whatever processing used afterwards in phase 2.

D. Removal of snow and ice pixels

Snow and ice surfaces are bright in VNIR range and dark in the SWIR range (see Fig. 3). However, depending on the SWIR threshold T_{SWIR} used to detect bright surfaces, snow/ice surfaces may be included in m_{bright}. We propose then to detect them in order to prevent some potential false alarms. Normalized Difference Snow Index [29] (NDSI) is often computed to detect snow, but another method more suited to hyperspectral imaging has been developed. Snow and ice have a significant absorption band around 1030 nm [30] (see Fig. 3). Detection of absorptions is based on continuum removal [31]. The spectral continuum C_λ is defined as the convex envelope of the upper part of the spectral reflectance ρ_λ, connecting the local maxima with segments. The ratio ρ_λ / C_λ highlights the local absorptions of ρ_λ. The approach is similar to those implemented in Tetracorder [32] and CHRIPS [33]: wavelengths associated to the segments ends are fixed.

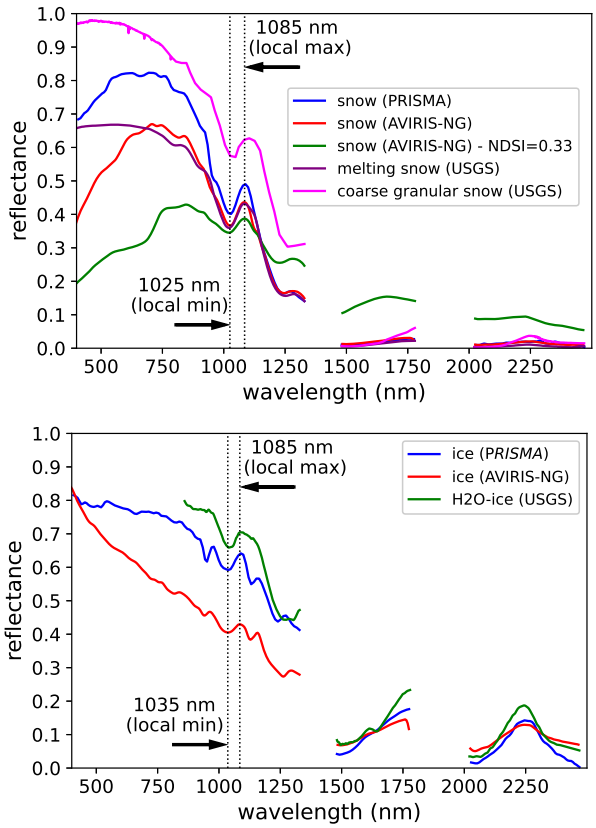

Fig. 3. Spectral reflectances of snow and ice pixels from PRISMA images, AVIRIS-NG images and the USGS spectral library [34]. A local minimum due to absorption can be seen around 1030 nm (1025 nm for snow and 1035 nm for ice) and a local maximum appears around 1085 nm. A snow reflectance spectrum with a low NDSI index (0.33) is shown: such a spectrum is difficult to detect with NDSI thresholding (NDSI is generally greater than 0.6 for snow).

For each absorption, some reference wavelengths are fixed before (λ_1) and after (λ_2) the absorption. The segment s_λ connecting ρ_{λ_1} and ρ_{λ_2} is computed:

$$s_\lambda = \rho_{\lambda_1} + (\rho_{\lambda_2} - \rho_{\lambda_1})(\lambda - \lambda_1)/(\lambda_2 - \lambda_1). \quad (3)$$

The proposed criterion for deciding whether an absorption exists is to compute the minimum value of the ratio ρ_λ / s_λ in a given spectral range $[\lambda_3, \lambda_4]$ that is included in $[\lambda_1, \lambda_2]$, and to assess if it is below a given threshold $T_1^{\text{s.i}}$:

$$\min_{\lambda_3 \leq \lambda \leq \lambda_4} \rho_\lambda / s_\lambda < T_1^{\text{s.i}}. \quad (4)$$

This criterion indicates if a local minimum or a local decrease of reflectance exists between λ_3 and λ_4. It is then defined by a 5-uplet $U = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4, T_1^{\text{s.i}}\}$. λ_3 and λ_4 must be located towards the beginning and end of the absorption band: $\lambda_3 = 1010$ nm and $\lambda_4 = 1045$ nm are chosen from a few tests and prove to be quite accurate. λ_1 must be selected before absorption, and not too far away, as reflectance increases sharply as wavelength decreases: $\lambda_1 = 980$ nm is chosen. λ_2 must lie after the absorption band: $\lambda_2 = 1085$ nm is chosen, it corresponds to a local maximum of snow and ice reflectances. In practice, for a given hyperspectral image, the spectral bands closest to the λ_i values are selected.

This criterion is highly discriminating, but does not stand alone. Four other criteria have therefore been developed to ensure accurate detection of snow and ice while minimizing confusion with other classes.

- A local maximum exists around 1085 nm: $\max_{1010\leq\lambda\leq1320}\rho_\lambda = \max_{1060\leq\lambda\leq1110}\rho_\lambda$.
- The reflectance level is sufficiently high in the [600-700nm] and [1000-1100nm] ranges: $\min_{600\leq\lambda\leq700}\rho_\lambda \geq T_2^{\text{s.i}}$ and $\min_{1000\leq\lambda\leq1100}\rho_\lambda \geq T_3^{\text{s.i}}$.
- SWIR reflectance is low. It is generally less than 0.1 for snow, but can reach 0.2 for ice. $\max_{\lambda\in\{1550,1650,2080,2300,2350\}}\rho_\lambda \leq T_4^{\text{s.i}}$.

In the last criterion, some bands in the SWIR range are not taken into account, such as those between 1660 and 1800 nm, or those close to 2250 nm. Ice reflectance is not low at these wavelengths: it increases linearly over [1500-1750nm], and a local maximum exists around 2250 nm (see Fig. 3). Criteria based on these characteristics could be used to discriminate between ice and snow. Note that the reflectance of snow and ice is very low around 2350 nm: this property is particularly useful for detecting very bright cloudy pixels without fear of confusing them with ice (see inequality $\rho_{2350} \geq T_{\text{SWIR}}^{vb}$ in section II-C).

Then, in order to estimate the optimal values for $T_1^{\text{s.i}}$, $T_2^{\text{s.i}}$, $T_3^{\text{s.i}}$ and $T_4^{\text{s.i}}$, all criteria were applied to a set of around ten thousand different snow/ice spectra and eleven million non-snow/ice spectra from eight PRISMA images and four AVIRIS-NG images varying the threshold values: $T_1^{\text{s.i}} \in [0.85, 0.98]$, $T_2^{\text{s.i}} \in [0.05, 0.40]$, $T_3^{\text{s.i}} \in [0.05, 0.40]$, $T_4^{\text{s.i}} \in [0.02, 0.30]$. The optimal combination of thresholds is the one that maximizes the F_1 score for snow/ice pixel detection on the dataset used. Before applying criteria, Gaussian filtering is applied to each reflectance spectrum to reduce noise. This does not attenuate the absorption depth, which is relatively large for snow and ice. Optimal threshold values are $\{T_1^{\text{s.i}}, T_2^{\text{s.i}}, T_3^{\text{s.i}}, T_4^{\text{s.i}}\} = \{0.94, 0.20, 0.10, 0.21\}$. The F_1 score is 98.4% for the data used (precision is 99.5% and recall is 97.4%).

The process returns a binary mask $m_{\text{snow-ice}}$ where the value of each pixel detected as snow/ice is 1, and 0 elsewhere.

E. Detection of potential cloud pixels above green vegetation

Green vegetation has a quite low reflectance for spectral bands below 650 nm. A cloud above green vegetation may increase observed reflectance but not sufficiently if cloud is thin (see Fig. 4). Detection of bright pixels described in section II-B may fail for such pixels, in particular due to the inequality (1). A dedicated green vegetation processing is applied to solve this problem. It mainly consists of detecting green vegetation pixels and retaining only those with a sufficiently low WV value.

As a first step, pixels of green vegetation (sparse and dense) are searched for in the whole image by computing dedicated CHRIPS criteria [33]. These criteria evaluate geometric properties associated with vegetation spectral reflectances: local maximum at 1660 nm, parabolic shape around 1660 nm, local maximum around 2210 nm, NDVI index (Normalized Difference Vegetation Index) above a given threshold. These criteria remain almost true for reflectances that mix green vegetation and clouds (see Fig. 4). They are however less and less true as the cloud thickness increases (notably local maximum around 2210 nm), but for such cases, the reflectance in the visible domain is singularly increased and the inequality (1) is true: such pixels are already embedded in m_{bright}. The constraints imposed by the CHRIPS criteria for detecting vegetation have been relaxed compared with those recommended in the original method, to take account of spectral distortion due to the presence of clouds: T_{g1} = 0.15, T_{g2} = 0.25, T_{g3} = 0.55 (NDVI thresholds), the criterion $\max_{1640\leq\lambda\leq1670}\rho_\lambda = \max_{1520\leq\lambda\leq1770}\rho_\lambda$ has become $\max_{1640\leq\lambda\leq1670}\rho_\lambda \geq 0.9 \max_{1520\leq\lambda\leq1770}\rho_\lambda$, the criterion $a < -8\rho^*$ has become $a < -6\rho^*$ and the criterion $\rho_\lambda^* / \rho_{1300} < 1.1$ has become $\rho_\lambda^* / \rho_{1300} < 1.3$ (see [33] for more details on the criteria). These modifications have been obtained by gradually varying the CHRIPS parameters on pixels from AVIRIS-NG images including green vegetation and thin clouds. Pixels identified as dense green vegetation and sparse green vegetation (*i.e.* those checking NDVI $\geq$ 0.25) are gathered in the binary mask m_{veg}^*: each pixel considered as green vegetation is 1, and 0 elsewhere.

In a second step, only the green vegetation pixels that may contain clouds should be kept. These pixels therefore have a lower WV value than non-cloudy pixels. Two constraints must be checked.

- The WV values must be sufficiently low: WV $< \mu_0 - 1.8\ \sigma_0$, where μ_0 and σ_0 are respectively the mean and the standard deviation of the histogram of WV values not detected as potentially cloudy, i.e. WV values of pixels checking m_{bright} = 0. Note that dark pixels are removed from the WV map (see processing in section III-B1) and are not taken into account when calculating the histogram.
- Reflectance in the VNIR range is not too low: $\rho_\lambda \geq 0.03$ in visible range and $\rho_{800} \geq 0.16$. This makes it possible to exclude dark vegetation, whose low reflectance is often noisy and for which the WV estimate is unreliable.

The binary detection mask D_{p} is obtained: the value of each pixel considered as potential cloud above vegetation is 1, and 0 elsewhere. Finally, the mask of potential clouds above green vegetation m_{veg} is the combination of both binary masks: $m_{\text{veg}} = m_{\text{veg}}^* \times D_{\text{p}}$.

Note that here, the WV map is used to detect potential cloud pixels (phase 1), but it is also used in phase 2 to possibly remove them: the processing used differs. In the example of an image covered mainly with vegetation but without clouds, it is likely that there are vegetation pixels that would be identified by this method as potential cloud pixels (those with the lowest WV values). It is during phase 2 that it will be decided whether these pixels are definitely cloudy or not.

F. Mask of potential cloud pixels

The mask of potential cloud pixels m_0 is the combination of two masks: $m_0 = m_{\text{bright}^*} + m_{\text{veg}}$, where

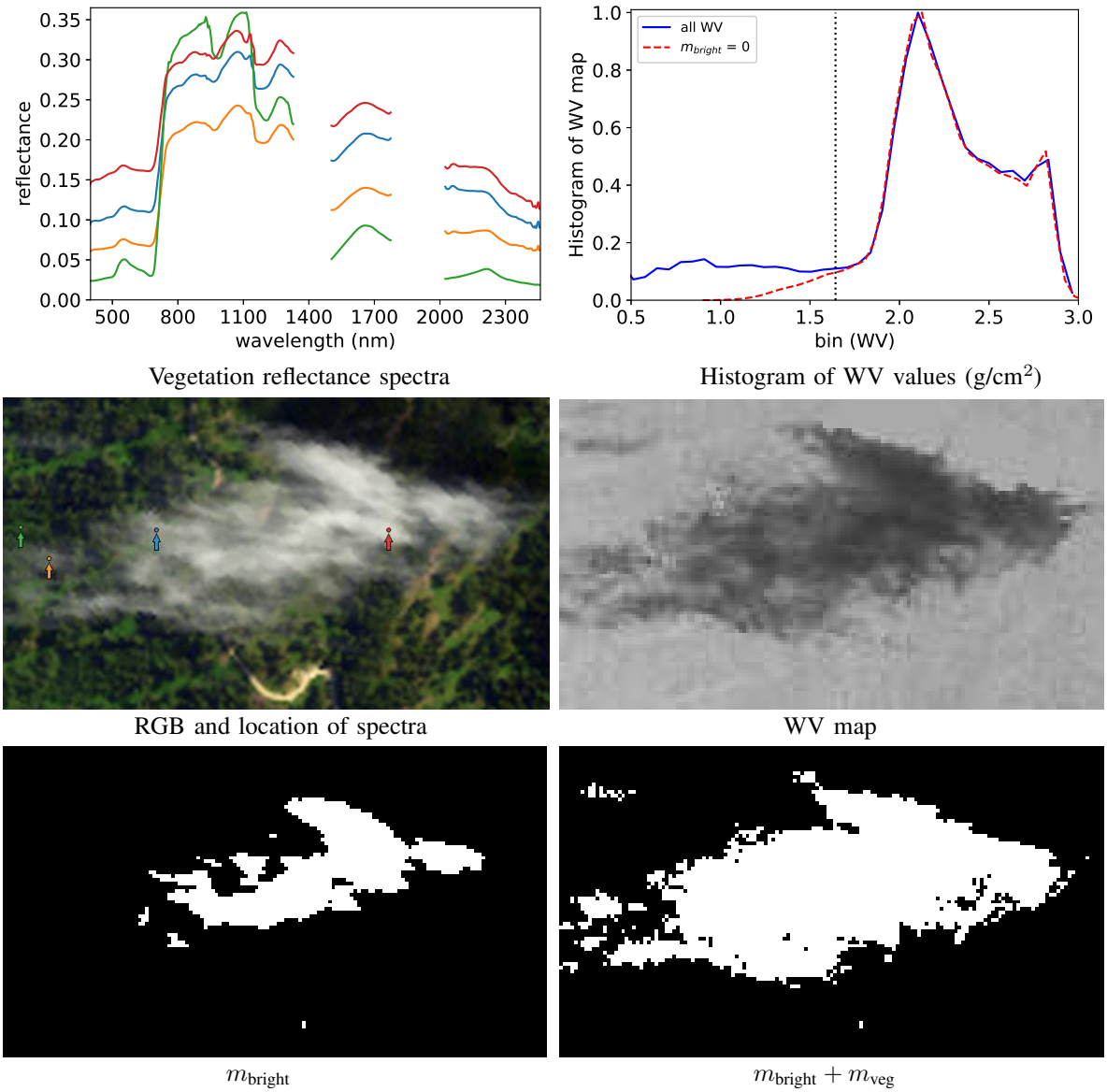

Fig. 4. Spectral reflectances of four pixels including green vegetation and more or less dense clouds (no cloud for green spectrum) in an AVIRIS-NG image. The presence of clouds does not distort the spectral variations too much: the spectral properties of green vegetation such as the presence of red-edge around 700 nm, local maximum around 1660 nm (see CHRIPS criteria [33]) appear whatever the cloud thickness. However, the presence of a cloud tends to increase the average reflectance level everywhere (except perhaps around 800-1200 nm). In the case of a thin cloud (orange pixel), the increase in reflectance in the visible range due to the cloud is insufficient for the minimum reflectance in the VNIR to exceed T_{VNIR} (see inequality (1)). Histograms of WV values are shown for all pixels (blue) and for pixels checking $m_{\text{bright}} = 0$ (dotted red). Estimated threshold $\mu - 1.8\sigma$ is also represented (dotted black): pixels detected as green vegetation and with a WV value below this threshold are considered as potential cloud pixels. The use of m_{veg} makes it possible to better identify all potential cloud pixels, even in thin areas (see $m_{\text{bright}} + m_{\text{veg}}$).

$m_{\text{bright*}} = m_{\text{bright}} \times (1 - m_{\text{snow-ice}})$ is the mask of bright pixels for which snow and ice pixels have been removed, and m_{veg} is the mask of potential cloud pixels above green vegetation. In order to avoid pixels with m_0 values of 2, m_0 can be re-written as $m_0 = m_{\text{bright*}} + m_{\text{veg}} \times (1_M - m_{\text{bright*}})$, where 1_M is a mask full of 1. In the following, we will consider that a pixel belongs to m_0 if its value in m_0 is 1.

III. Refinement of the Mask of Potential Cloud Pixels Using the WV Map (Phase 2)

A. Overview

The goal is to refine the mask of potential cloud pixels using the WV map. Several stages follow: pre-processing of the WV map, application of two cloud detection methods M^{T} and M^{C} and spatial extension of clouds.

First, three pre-processings of the WV map are presented: identification and removal of dark pixels, reduction of dependency between WV map and altitude, which is useful when altitude varies greatly in a single image, and computation of the WV map histogram. Two cloud detection methods, M^{T} and M^{C}, are then proposed. The M^{T} method is based on thresholding the WV map: cloudy pixels have WV values below a given value, which is assessed by analyzing the histogram of the WV map. Unfortunately, the automatical search for thresholds is a very hard task as histogram of WV values may change a lot depending on the distribution

of clouds. Using this method, a number of clouds may be missed, especially when clouds of very different thicknesses or altitudes are present on the same scene. M^C method is less sensitive to this kind of problem: it exploits the WV contrast between the cloudy pixels and their environment. The two methods offer redundant but also complementary results, so their results are merged after the post-processings (see section III-E).

Afterwards, a processing allowing to identify cloud pixels that could have been missed by M^T and M^C methods is proposed: it makes it possible to spatially extend the clouds already detected. At the end, a processing is applied to eliminate false clouds.

B. Pre-processing of WV map

1) Removal of dark pixels

WV map is pre-processed in order to remove pixels that may cause false cloud detection. As signal-to-noise ratio is low around 940 nm and 1140 nm for dark surfaces such as water or shadows, WV may be underestimated and may lead to local minima on WV map. Such surfaces are obviously not clouds, they are detected with the following criterion:

$$\max_{\lambda \geq 900} \rho_\lambda \leq T_{dark} \tag{5}$$

where T_{dark} is a threshold. From a few tests performed by applying the full CHIWAWA method to a few PRISMA and AVIRIS-NG images, it appeared that choosing a T_{dark} value between 0.06 and 0.08 is a good compromise. A value below 0.06 may miss dark pixels. With a value higher than 0.08, cloud pixels with low reflectance may be missed, such as thin clouds over water. In the following, T_{dark} = 0.07 is used.

Once detected, the dark pixels are located on the WV map. Their WV value is set to -1. These pixels will later be used in a special way, for example to detect clouds over water. Moreover, pixels that may have a non-valid WV estimation (WV values too close to 0 or missing values) are identified: WV value of these pixels is set to -2. These pixels will not be taken into account in the following processings. To improve the visualization of WV maps in this article, dark pixels are detected and the average value of the WV map (excluding dark pixels) is assigned to them (see Fig. 5).

2) Reduction of dependency between WV map and altitude

WV value highly depends on altitude: it decreases as altitude increases. In areas with relief, WV variations are therefore strongly impacted by altitude. The highest altitude areas could be mistakenly considered as clouds. Then, to improve cloud detection performance in such areas, we propose to reduce altitude dependency of WV values. First, this dependency is modelled with a regression line:

$$w = a\, z + b, \tag{6}$$

where w is the WV value, z is altitude, a and b are coefficients that are estimated with least squares. Next, a corrected value of w, denoted w^c, is obtained by:

$$w^c = \frac{w}{a\, z + b} w_0, \tag{7}$$

where w_0 is the mean WV value of pixels not belonging to m_0. w_0 allows w^c to be homogeneous to w.

A digital elevation model (DEM) is used to give an estimated value of ground altitude of each pixel. DEM is resampled to be superposable to the WV map. In this study, DEM were downloaded from Earth Explorer (https://earthexplorer.usgs.gov/).

For the regression, all pixels in the image are used except those with unreliable WV values (-1 and -2) and those that may be potentially cloudy, *i.e.* pixels verifying m_{bright} = 1, as their WV values may not be correlated with the ground elevation provided by the DEM (when these are truly cloud pixels). After regression, a new WV map is then computed by replacing each w value by its corresponding w^c value. Pixels with WV values of -1 and -2 in the original WV map are reported in the corrected WV map.

In practice, this processing will only be used on mountain images. It will be illustrated in section IV-G.

RGB initial WV map WV map corrected for visualization

Fig. 5. AVIRIS-NG image including land, water and clouds. The WV estimate is erroneous above dark surfaces such as water (middle image). To improve visualization, dark pixels are detected and the mean WV value of image (excluding dark pixels) is assigned to them (right image): clouds become easier to observe.

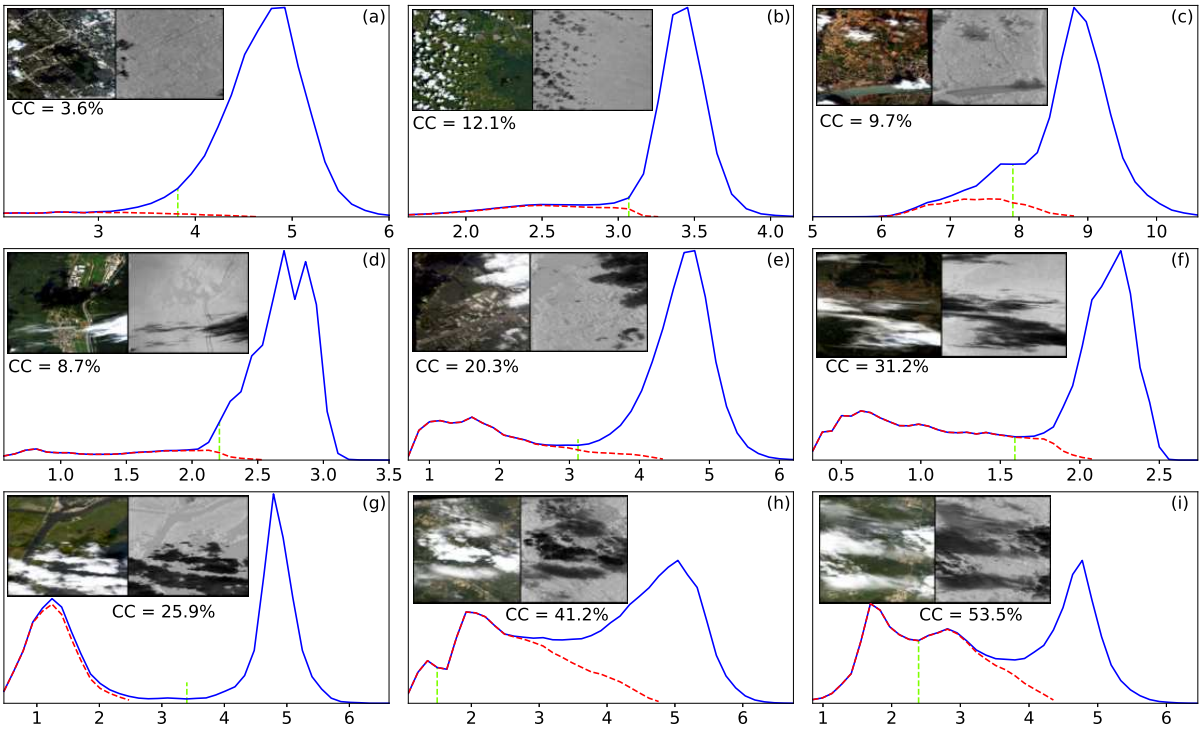

Fig. 6. Histograms of WV map (g/cm^2 for the abscissa, arbitrary units on the ordinate) for different AVIRIS-NG and PRISMA images containing clouds with different cloud covers (CC). For each image, the RGB composition, the WV map, the histogram of the WV map (blue), the histogram of the WV map for cloud pixels only (dotted red) and the threshold estimated with the M_T method (dotted green) are shown. The histograms have highly variable shapes, and the WV values associated with clouds are also very variably distributed. This explains why automatic threshold estimation (M_T method) is a relatively complicated task.

3) Computation of histogram H_{WV}

Histogram of WV map, denoted H_{WV}, is computed with N_b bins of equal size (N_b = 41), gathered in a vector w^b (histogram abscissa). H_{WV} is normalized by its maximal value, *i.e.* max H_{WV} = 1. Let us note k_i the number of the first bin and k_e the number of the final bin. At this step, k_i = 1 and k_e = N_b.

The WV range of histogram may be contaminated by very low or very high values that are not representative of the image. High WV values cannot be associated to clouds and then can be suppressed without being afraid of decreasing cloud detection. In practice, k_e is the bin at which H_{WV} exceeds 0.05 for the first time (starting from the end of the histogram). The same process is led for low WV values but the threshold is lower as most low WV values can be associated with clouds. k_i is the bin at which H_{WV} exceeds 0.01 for the first time (starting from the beginning of the histogram).
Finally, only the part of the histogram between bins k_i and k_e is kept. For the retained part of histogram, minimal and maximal values of vector w^b are then $w^b[k_i]$ and $w^b[k_e]$, and will be respectively noted w^b_{min} and w^b_{max} in the following. The range R_{WV} of a WV map is then defined as the difference between w^b_{min} and w^b_{max}: $R_{WV} = w^b_{\text{max}} - w^b_{\text{min}}$.

C. M^T method: WV histogram thresholding

Histogram H_{WV} is composed of one mode or more. In this study, a mode is defined in a particular way: it is a part of H_{WV} located between two successive local minima, and by construction contains a local maximum of H_{WV}. Two successive modes do not overlap. Thereafter, we will assign a number to each mode according to its position in the histogram: the first mode is the leftmost mode (with the lowest WV values), the second mode is the mode following the first mode, etc. For illustration, histograms with different number of modes are shown in Fig. 6. The number of modes depends on cloud distribution and image context. Several cases can be distinguished:

- only one mode: it may happen when there is little or no cloud cover (see Figs. 6a and 6c), or when the WV values corresponding to cloud pixels are very spread out and form a long left-hand tail on the histogram (see Figs. 6b and 6d).
- two modes: cloud pixels are generally located in the first mode, i.e. corresponding to the lowest WV value. Meanwhile, if the proportion of cloud is large, it may happen that cloud pixels are included in the beginning of the second mode (see Figs. 6e, 6f, 6g).
- three modes or more. This can happen when the proportion of clouds is high or when their thickness or altitude varies greatly (see Figs. 6h, 6i). First mode corresponds to clouds. Last mode does not correspond to clouds. Other modes are generally associated to clouds.

The M$^{\text{T}}$ method consists in thresholding the histogram: a pixel is cloudy if it belongs to m_0 and if its WV value is below a given threshold T_{WV}. Such process needs to be carried out

with caution. If T_{WV} is too low, some clouds may not be detected. If T_{WV} is too high, false detections may rise. An empirical approach was opted. Based on tests on different images, three main cases are considered, allowing an empirical estimate of T_{WV}.

- Case 1: this case occurs especially when many clouds are present: the histogram has at least two modes and the first mode on the left is significant (Figs. 6e, 6f, 6g, 6h, 6i). If the first mode with a significant amplitude on the left is sufficiently close to the left edge of the histogram, i.e. the abscissa position p_{max} of the mode maximum checks: $p_{\text{max}} \leq w^b_{\text{min}} + 0.4\, R_{WV}$, then the threshold T_{WV} is set at the end of the mode, *i.e.* it is positioned at the first local minimum of the histogram to the right of the mode.
- Case 2: this case is considered if case 1 has not been fulfilled. It deals with a histogram with a long tail on the left (histogram in general monomodal), this tail being associated with the presence of clouds. (Figs. 6a, 6b, 6c) This case is validated when the mode furthest to the right of the histogram contains at least 70% of the pixels in the image. Let p_{max} be the WV value associated with the maximum of this mode. The threshold T_{WV} is then the WV value v to the left of p_{max} that checks $H_{\text{WV}}(v) = 0.15\, H_{\text{WV}}(p_{\text{max}})$.
- Case 3: if the two previous cases are not fulfilled, the threshold T_{WV} is set in order to retain only the 15% of the lowest values of the WV map.

The output of the process is the binary mask m_1^{T}. When clouds of different altitudes and thicknesses are present (see Figs. 6h and 6i), threshold estimation is generally underestimated. The contrast-based M^{C} method is a best way to handle such scenes.

D. M^C method: WV contrast analysis

In this approach, a pixel is considered as cloudy if it belongs to m_0 and if its WV value is significantly below the mean WV value of its environment.

Let us consider a pixel p belonging to the mask of potential cloud pixels (*i.e.* $m_0 = 1$) with a WV value w_0. In the WV map, all pixels in a square window of length L centered on p are retained. Among these, pixels with wrong WV estimation are removed, *i.e* pixels for which WV value is -2 (see section III-B). Pixels belonging to m_0 are also removed because WV contrast needs to be computed between pixel p and pixels that are not cloudy for sure.

Among remaining pixels, let us consider S_i the set of pixels for which WV value is positive (i.e "normal" pixels), and S_d the set of pixels for which WV value is -1, *i.e.* dark pixels. Numbers of elements of these sets are respectively noted N_i and N_d. The median WV value of the set S_i is noted w_i. The WV contrast is then defined as the difference $w_i - w_0$. In the following, four thresholds will be used: a minimal number of "normal" pixels N_i^*, a threshold T_w on minimal WV constrast, a minimal number of dark pixels N_d^* and a minimal ratio F_d characterizing the proportion of dark pixels. Three cases can occur.

- Case 1: $N_i \geq N_i^*$. This is the most common case: the number of pixels remaining in square window is sufficient to make a decision. If the WV contrast $w_i - w_0$ is higher than T_w, p is cloudy, otherwise it is not.
- Case 2: $N_i < N_i^*$, $N_d > N_d^*$ and $N_d/(N_d + N_i) > F_d$. The number of "normal" pixels in the square window is not sufficient to compute a reliable WV constrast. This is mainly due to the high presence of dark pixels such as water or shadows. It generally happens over the sea. According to our experiments, if an area checks the criteria of a potential cloud and it is surrounded by water or shadows, it is very likely to be a cloud. Then, decision is that p is cloudy. For the user's eventual needs, a dedicated mask could be created for these pixels to indicate why they were detected as cloudy. Wrong detection may happen with specific cases though, such as an emerging carbonate sand bank, but are quite rare. Such cases could be solved with dedicated processes such as detection of carbonates.
- Case 3: if neither case 1 nor case 2 are validated, p is not cloudy.

If a cloud is wide enough so its minimal spatial size is above L, pixels that are close to the center of the cloud have neighbours that are still inside the cloud. Such pixels will be considered as not cloudy by applying defined criteria (we are in case 3). The processing described in section III-F aims to solve this problem.

From tests, the following threshold values are used: N_i^* = 50, N_d^* = 100 and F_d = 0.8. L depends on spatial resolution: the selected value is 41 for PRISMA images and 101 for AVIRIS-NG images. The choice of T_w is crucial because it is mainly this parameter that will determine the detection performance. It was chosen to use a fraction of the WV range R_{WV} (see. section III-B.3): T_w = 0.1 R_{WV}. A value of 0.1 allows the best tradeoff between false negative detections and false positive detections.

If the value of T_w is very low, typically lower than 0.1, the contrast measurement is unreliable and many false alarms may occur. To solve this problem, another threshold T_m is computed: T_m = 0.06 V_m, where V_m is the average value of the WV map. In the end, the threshold used is T_w^* = max(T_w, T_m).

The binary mask of detected cloud pixels is m_1^{C}. Fig. 7 illustrates the WV contrast process for cases 1 and 2.

E. Fusion of cloud detection masks

The independent application of M^{T} and M^{C} methods produces two detection masks m_1^{T} and m_1^{C}. In practice, the M^{C} method performs much better than the M^{T} method, and can identify all the clouds present on its own. There are, however, cases where it performs less well, notably when clouds contrast poorly with their surroundings in the WV map (see P_9 image in Figs. 13 and 15). In order to benefit from the complementarities of both methods, a more complete detection mask m_1 is computed by combining both detection masks: $m_1 = m_1^{\text{T}} + m_1^{\text{C}}$, that can be-rewritten as $m_1 = m_1^{\text{T}} + m_1^{\text{C}}(1 - m_1^{\text{T}})$ so that no value exceeds 1.

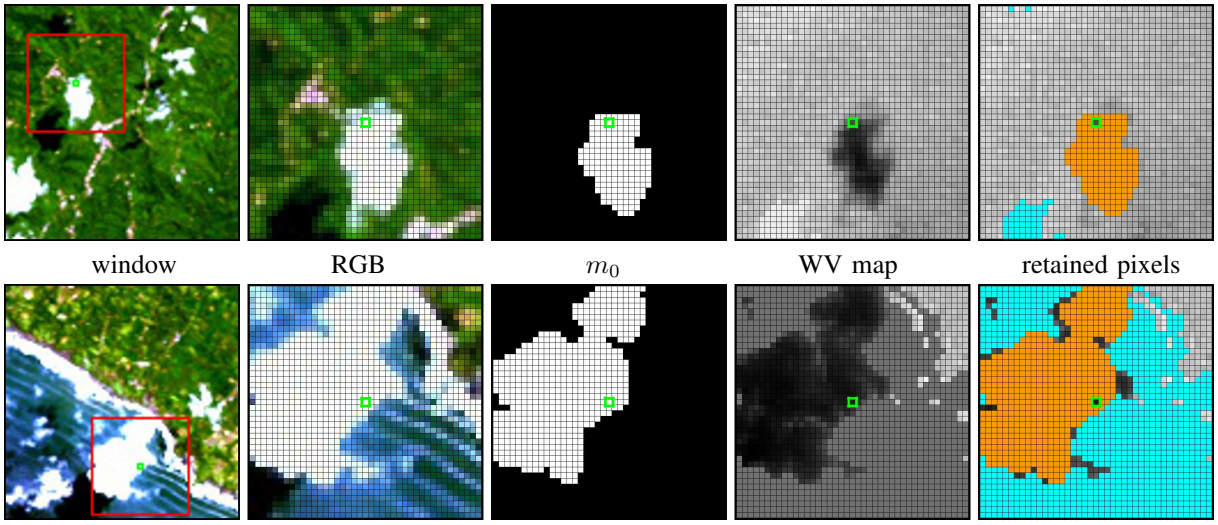

Fig. 7. Steps in the M^C method for a given pixel p in a cloud over land (top, common case) and in a cloud over water (bottom) for a PRISMA image. Pixels from m_0 (orange pixels) and dark pixels (cyan pixels) are removed from the window on WV map. WV value of the studied pixel p (green) is compared with the remaining pixels in the window. To improve visualization on the WV map, the average WV map value has been assigned to dark surfaces (shadows and water).

F. Spatial extension of clouds

Some cloudy pixels may have not been detected with M^T and M^C methods. These are generally pixels located at cloud boundaries or within extended clouds. The following processing is designed to solve this problem. The set of cloudy pixels candidates S is defined as follows: a pixel p is a candidate if it belongs to m_0, does not belong to m_1, and has a coherent WV value (not -1 or -2).

Each pixel of S is successively scanned. Let us consider a pixel p from S with a WV value w. p is considered as cloudy if it has at least one neighbouring pixel p_i that belongs to m_1 and has a WV value w_i close enough to w, *i.e.* $|w - w_i| \leq T_{se}$, where the threshold T_{se} is defined from the range of WV values R_{WV} (see. section III-B.3): $T_{se} = 0.05\, R_{WV}$. m_1 is updated each time a new pixel p is validated as cloudy. The process is recursively launched until the number of new pixels in m_1 does not change from previous launch.

Some "holes" may still appear inside some extended clouds after this processing. The pixels inside these holes are most likely cloudy as long as these pixels belong to m_0. These pixels are considered cloudy. False alarms may occur but are very rare in practice. The output binary mask of this full process is noted m_2.

G. Post-processing: removal of false clouds

Two post-processings are applied in order to remove false detected clouds from detection mask m_2.

The first post-processing runs through each cloud and compares its average WV level with that of its environment. It is similar to M^C method. The main difference is that it is region-based whereas M^C is pixel-based. The binary mask m_2 is labelled into separate regions: each cluster of pixels forming a cloud are thus associated with a single region. For each region R, the median WV value w_1 is computed inside the cloud, the median WV value w_2 among the neighboring pixels of R that do not belong to a cloud is computed. These neighboring pixels are obtained by the difference between two morphological dilations of R. Let us consider a binary mask M_R in which all pixels in R are 1 and pixels are 0 elsewhere, and note $f(R, r)$ a spatial dilation of M_R with a square of width r. Then all neighboring pixels of R are pixels for which values of $f(R, r_2)$ - $f(R, r_1)$ are 1 ($r_2 > r_1$). If the difference $w_2 - w_1$ is above a given threshold S_{fc}, the region R is validated as a cloud. S_{fc} is computed from the range of WV values R_{WV} : $S_{fc} = 0.1 R_{WV}$, which is the same value as T_w (see M^C method).

We do not use r_1 = 0 because pixels too close to a cloud are likely to be undetected cloudy pixels whose WV value is close to the WV value inside the cloud. From tests, the following values provide satisfactory results: r_1 = 15, r_2 = 25 for PRISMA and AVIRIS-NG images. The full post-processing is illustrated in Fig. 8. Its output binary mask is m_3.

The second post-processing removes small or narrow regions detected as clouds. It is particularly useful to eliminate small urban structures or roads that could be mistakenly identified as clouds. As clouds may have some minimal spatial extension, a post-process can be applied to remove detected clouds with too few pixels. The clouds are detected by labelling the mask m_2 into separate regions. The number of pixels in each region can be then thresholded in order to remove too small "clouds". Another method consists in applying an erosion of a given radius r_3 to each region R. If no pixels from the R region remain after erosion, the cloudy pixels associated with R are deleted. Otherwise, they are retained. This approach also makes it easy to remove long, thin structures, such as roads, which might otherwise have been detected as clouds. The cloudy pixel output mask is denoted m_{cloud}.

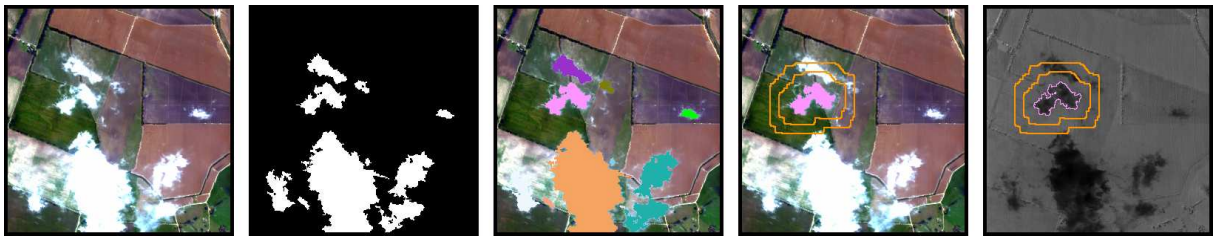

Fig. 8. Detection of false clouds. Cloud detection mask m_2 is labelled into separate regions. For each region, a crown of pixels is located in the WV map: median values are compared between pixels inside cloud region and pixels inside the crown that do not belong to m_2.

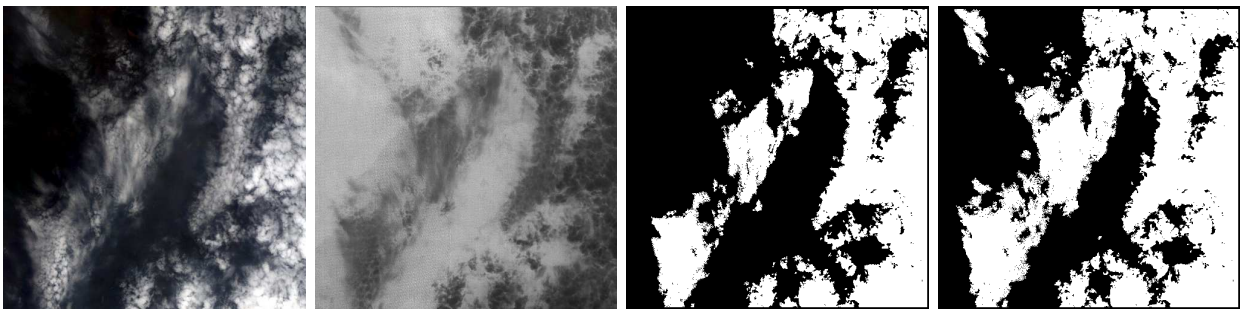

Fig. 9. Application of the CHIWAWA method once and twice to a region of interest in a PRISMA image containing clouds with various thicknesses. Thicker clouds are detected in the first run, and thinner clouds in the second.

H. Relaunch of cloudy pixels validation

In the case of images containing clouds of very different thicknesses or altitudes, the WV values of cloud pixels can be very variable. This has the effect that clouds with the highest WV values may not be detected by the methodology described above. This happens especially on the peripheral areas of the clouds, which are often less dense than the interior areas. This kind of problem can be solved by applying the cloud pixel detection twice.

For the first launch, the full processing chain described in Fig. 2 is applied normally. The cloud detection mask m_F^1 is obtained (noted m_{cloud} in section III-G and in Fig. 2). For the second launch, the WV map is modified: values of the detected cloud pixels in m_F^1 are set to -2 so that these pixels are not used anymore. The processing chain is then applied with this new WV map and produces the detection mask m_F^2. The final detection mask m_F is the fusion of m_F^1 and m_F^2: $m_F = m_F^1 + m_F^2$. The interest of this processing is illustred in Fig. 9. Relaunching CHIWAWA can increase the false detection rate, particularly in urban areas, and remains optional.

IV. Experimental tests: performance and analysis

A. Hyperspectral datasets

CHIWAWA is assessed on eighteen PRISMA images (see Table II) and nine AVIRIS-NG images (see Table III). PRISMA [27] and AVIRIS-NG [28] data, *i.e.* ground reflectance images and associated WV maps, were retrieved directly from the corresponding websites (L2C or L2D products). For each AVIRIS-NG image, regions of interest (ROIs) were selected and processed separately to facilitate memory management (some images exceed 10 GB in size). Main characterics of these images are shown in Table I. All images are located on a world map in Fig. 10.

To build ground truth, each image was manually annoted by the author as accurately as possible using RGB images, SWIR spectral bands (including the cirrus band around 1380 nm) and WV maps. However, uncertainties may remain, particularly concerning cloud boundaries and the location of very thin clouds.

Only images P_{13}, P_{15} and some regions of interest in A_9 are mountainous. These images are specifically processed in sections IV-G and IV-H. All other images are analyzed in sections IV-E and IV-F.

TABLE I
MAIN CHARACTERICS OF HYPERSPECTRAL IMAGES USED: SENSOR USED FOR ACQUISITION, NUMBER N_I OF IMAGES, NUMBER N_S OF SPECTRAL BANDS, MEAN VALUE OF FULL WIDTH AT HALF MAXIMUM (FHWM) IN FULL SPECTRAL RANGE AND SPATIAL RESOLUTION.

Sensor	N_I	N_S	FWHM	Spatial resolution
PRISMA	18	239	10 nm	30 m
AVIRIS-NG	9	425	5 nm	3.4 – 8.4 m

B. Cloud detection methods assessed

The performance of the CHIWAWA method is compared with three other methods: Fmask, CDAG and COTLUT

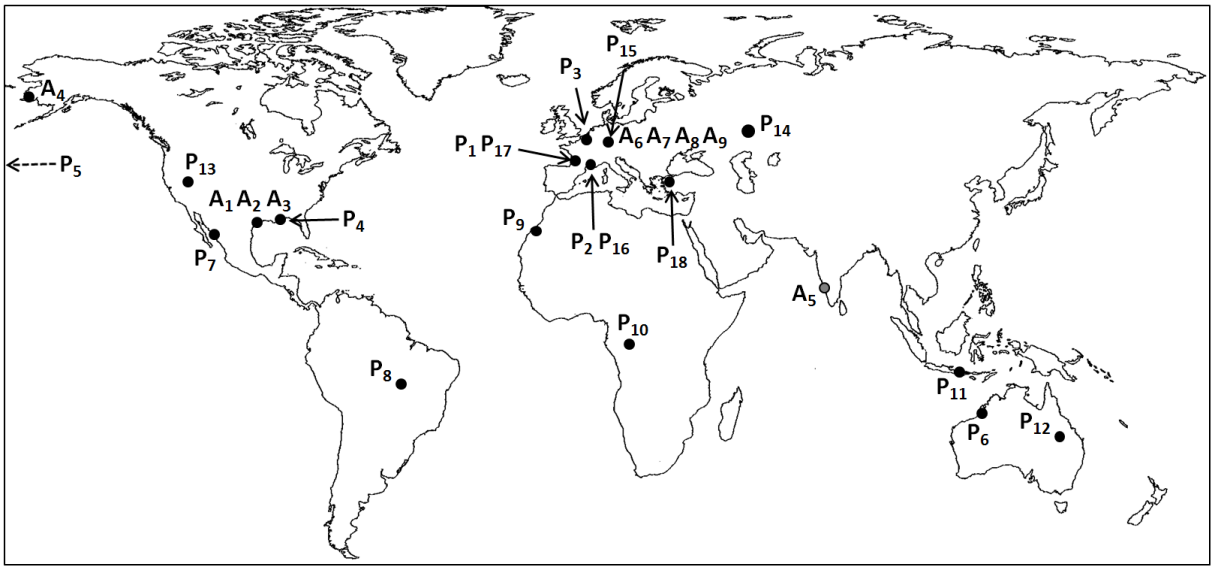

Fig. 10. Location of PRISMA images (P_1–P_{18}) and AVIRIS-NG images (A_1–A_9).

TABLE II

IDENTIFICATION (ID.) AND REFERENCES OF PRISMA IMAGES USED IN THIS PAPER. ALL IMAGES HAVE 1000 ROWS AND 1000 COLUMNS. P_{13}, P_{14} AND P_{15} ARE MOUNTAINOUS AND/OR CONTAIN SNOW/ICE. P_{16}, P_{17} AND P_{18} DO NOT CONTAIN ANY CLOUD BUT ARE USEFUL TO ASSESS ALGORITHM PERFORMANCE IN TERMS OF FALSE DETECTIONS.

Id.	Site	Reference
P_1	France (Toulouse)	PRS_L2C_STD_20200710104733_20200710104738
P_2	France (Marseille)	PRS_L2C_STD_20200831104351_20200831104356
P_3	Belgium	PRS_L2C_STD_20200710104532_20200710104536
P_4	USA (Louisiane)	PRS_L2C_STD_20220911165323_20220911165328
P_5	USA (Hawaii)	PRS_L2C_STD_20201002212222_20201002212227
P_6	Australie (NW)	PRS_L2C_STD_20200127021532_20200127021536
P_7	Mexico	PRS_L2C_STD_20190805175721_20190805175725
P_8	Brasil	PRS_L2C_STD_20191004134858_20191004134902
P_9	Western Sahara	PRS_L2C_STD_20191003114530_20191003114535
P_{10}	Gabon	PRS_L2C_STD_20220913093006_20220913093010
P_{11}	Indonesia (Bali)	PRS_L2C_STD_20200825024901_20200825024905
P_{12}	Australia	PRS_L2C_STD_20220914001035_20220914001039
P_{13}	USA (Cuprite)	PRS_L2C_STD_20200413183824_20200413183829
P_{14}	Kazakhstan	PRS_L2C_STD_20230211074155_20230211074159
P_{15}	Switzerland	PRS_L2C_STD_20210617104005_20210617104009
P_{16}	France (Fos-sur-Mer)	PRS_L2C_STD_20200408104431_20200408104435
P_{17}	France (Bordeaux)	PRS_L2C_STD_20210404110407_20210404110412
P_{18}	Turkey	PRS_L2C_STD_20200513091135_20200513091139

TABLE III

SPECIFIC CHARACTERISTICS OF AVIRIS-NG IMAGES USED IN THIS PAPER: IDENTIFICATION (ID.), SITE, IMAGE FILENAME, SPATIAL RESOLUTION, SIZE (NUMBER OF ROWS $\times$ NUMBER OF COLUMNS), NUMBER OF SELECTED REGIONS OF INTEREST N_{ROI}. EACH CHOSEN ROI HAS AROUND 1000 ROWS AND THE SAME NUMBER OF COLUMNS AS THE COMPLETE IMAGE. SOME ROIS IN A_9 ARE MOUNTAINOUS AND ARE PROCESSED USING ALTITUDE CORRECTION.

Id.	Site	Reference	spatial res.	size	N_{ROI}
A_1	USA (Texas site 1)	ang20191004t221515	8.3 m	2692 $\times$ 658	3
A_2	USA (Texas site 2)	ang20191004t201341	8.3 m	8431 $\times$ 702	9
A_3	USA (Texas site 3)	ang20191003t175539	8.4 m	6241 $\times$ 674	6
A_4	USA (Alaska)	ang20170720t225802	5.2 m	12149 $\times$ 695	7
A_5	India (South-East)	ang20180318t062047	4.0 m	5052 $\times$ 686	6
A_6	France (East)	ang20180626t122850	4.6 m	4168 $\times$ 656	4
A_7	Switzerland (site 1)	ang20180630t111646	4.1 m	10208 $\times$ 738	8
A_8	Switzerland (site 2)	ang20180627t123936	4.0 m	11112 $\times$ 717	9
A_9	Switzerland (site 3)	ang20180701t092524	3.4 m	11815 $\times$ 702	10

(PRISMA only).

Like the CHIWAWA method, Fmask [17], [18] first uses rules based on cloud physical properties to separate potential cloud pixels and clear-sky pixels. A cloud mask is then obtained by combining the potential cloud pixels with a cloud probability layer computed from a brightness probability, a variability probability, a Haze Optimized Transformation [35] and a cirrus test. The Fmask version dedicated to Sentinel-2 data [18] is used in this paper.

The CDAG method [19] performs thresholding on spectral bands, band differences and band ratios to build a cloud mask. The thresholds used were estimated from the analysis of various AVIRIS images. Thresholds adapted to Landsat-OLI and VIIRS sensors were used in this study. Of the two cloud masks obtained for each image, only the one with the highest F_1 score was selected for performance display.

Unlike CHIWAWA, which uses ground reflectance, Fmask and CDAG use apparent reflectance. This compensates for solar illumination, but does not correct for atmospheric effects. As the thresholds for these methods have been constructed and validated on multispectral images, we proceeded as follows. For each hyperspectral image, the spectral bands of the radiance image were used to synthesize multispectral images (Landsat-OLI, VIIRS, Sentinel2). The apparent reflectance was then calculated for each of the images obtained.

Finally, for PRISMA images, a cloud optical thickness (COT) map is provided as a L2C product. This map is estimated from a Look-up Table covering a large number of possible scenarios [36]. By identifying the pixels in this map whose optical cloud thickness is not zero, we can build a cloud mask that will be used for performance evaluation. This method will hereafter be referred to as COTLUT.

C. Evaluation metrics

For each image, overall accuracy OA, precision P, recall R and F_1 score are computed to assess performance and expressed as a percentage:

- OA = 100 × (TP + TN) / (TP + FP + TN + FN),
- P = 100 × TP / (TP + FP),
- R = 100 × TP / (TP + FN),
- $F_1 = 2/(P^{-1} + R^{-1})$.

TP (True Positive) is the number of cloudy pixels detected as cloudy, TN (True Negative) is the number of non-cloudy pixels detected as non-cloudy, FP (False Positive) is the number of non-cloudy pixels detected as cloudy, FN (False Negative) is the number of cloudy pixels detected as non-cloudy. Precision and recall are often called user accuracy and producer accuracy respectively. F_1 score reflects the compromise between precision and recall. The color code shown in Fig. 11 is used to display detection masks.

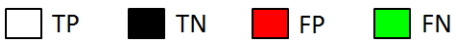

Fig. 11. Colors used on the cloud detection masks.

Cloud Cover CC is defined as the ratio between the number of pixels detected as cloudy and the total number of pixels in the image, expressed in percentage. The metric ΔCC evaluates the difference between estimated cloud cover CC^{est} and ground truth cloud cover CC^{ref}: $\Delta CC = |CC^{est} - CC^{ref}|$. As CC^{ref} could be computed as 100 × (TP + FN) / (TP + FP + TN + FN) and CC^{est} could be computed as 100 × (TP + FP) / (TP + FP + TN + FN), we obtain: ΔCC = |FP − FN| / (TP + FP + TN + FN).

It's important to note that performance metrics such as P, R and F_1 for images with low cloud cover are more difficult to analyze. Indeed, cloud boundaries are difficult to delineate, particularly when creating ground truths, and this can have a significant impact on the performance displayed. OA is not relevant either, as it is generally very close to 100%. Only ΔCC is taken into account when evaluating performance on images containing no clouds.

D. Values of CHIWAWA parameters for PRISMA and AVIRIS-NG images

Parameter values are different for PRISMA and AVIRIS-NG images (see Table IV). This is due in particular to the different spatial resolutions. Only one set of parameters is used for PRISMA images (CHIP). For AVIRIS-NG images, for which spatial resolution is in the range 3.4-8.4 m, two sets are defined. The threshold values T_{VNIR} and T_{SWIR} have been estimated with the process described in section II-B. T_{VNIR} has a significant impact on detection performance.A value of 0.10 will detect most clouds, but can lead to a large number of false alarms, particularly in urban areas. Roads and some buildings may have a significantly lower WV value than the environment and may be mistakenly detected as clouds by M^C method (see Fig. 16). Using a value of 0.15 somewhat reduces cloud detection (especially thin clouds) but greatly improves accuracy. The same reasoning applies to T^{vb}_{SWIR}. The relaunch option (see section III-H) may increase false detections for the same structures. In the end, two CHIWAWA versions are used: CHIU is dedicated to urban areas, CHIC is dedicated to other types of landscapes. The r_3 parameter, used to remove long, thin structures such as roads, is not essential when T_{VNIR} = 0.15. This parameter is therefore used for CHIC but not for CHIU. For PRISMA images (30 m spatial resolution), the problems that can arise in urban environments with AVIRIS-NG images do not arise, hence the unique parameterization proposed.

TABLE IV
PARAMETER VALUES FOR APPLYING CHIWAWA TO PRISMA IMAGES (CHIP), NON-URBAN AVIRIS-NG IMAGES (CHIC) AND URBAN AVIRIS-NG IMAGES (CHIU).

Parameter (section)	CHIP	CHIC	CHIU
T_{VNIR} (II-B)	0.07	0.10	0.15
T_{SWIR} (II-B)	0.07	0.03	0.03
T^{vb}_{VNIR} (II-C)	0.40	0.40	0.40
T^{vb}_{SWIR} (II-C)	0.12	0.12	0.15
L (III-D)	41	101	101
r_1 (III-G)	15	15	15
r_2 (III-G)	25	25	25
r_3 (III-G)	0	5	0
relaunch (III-H)	yes	yes	no

To apply the CHIWAWA method to images acquired with other sensors, it is necessary to modify the parameters. These

TABLE V
DETECTION PERFORMANCE ON PRISMA IMAGES P_1–P_{12}. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD. IMAGES P_{13}, P_{14} AND P_{15} ARE PROCESSED IN SECTIONS IV-G AND IV-H (NO PERFORMANCE HERE).

	Images	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
	CC (%)	2.2	2.0	60.0	12.1	0.1	21.1	13.4	38.2	8.2	59.4	1.5	34.6
	N_{cloud}	22k	20k	600k	121k	1.2k	211k	134k	382k	82k	594k	15k	346k
OA	Fmask	58.8	98.2	68.0	94.8	99.7	83.6	37.6	72.7	94.0	59.7	99.3	92.8
	CDAG	98.4	98.7	87.8	96.8	99.9	91.6	97.0	84.7	95.9	93.4	99.1	86.4
	COTLUT	97.6	98.1	76.8	91.6	99.9	79.1	88.7	74.7	92.2	61.8	98.7	84.3
	CHIWAWA	**99.2**	**99.6**	**96.9**	**98.8**	**100**	**99.7**	**98.8**	**97.8**	**97.9**	**98.2**	**99.7**	**99.8**
Precision	Fmask	5.3	52.2	65.3	70.2	29.0	56.4	17.6	58.4	60.7	59.6	70.3	83.1
	CDAG	73.9	79.5	94.8	**97.3**	55.5	98.9	99.0	**99.4**	74.1	**98.5**	95.0	**99.9**
	COTLUT	10.4	84.5	**96.0**	94.6	27.7	**99.9**	**100**	**99.4**	64.9	97.0	**99.6**	**99.9**
	CHIWAWA	**97.1**	**86.3**	95.6	93.0	**97.3**	98.9	93.3	97.5	**97.2**	98.0	96.6	**99.9**
Recall	Fmask	**98.8**	95.5	**99.8**	**98.9**	**95.0**	98.6	**99.8**	**98.9**	76.3	**100**	**95.7**	**99.5**
	CDAG	51.0	45.4	84.3	75.6	37.0	60.9	78.1	60.4	**77.1**	90.3	42.9	59.7
	COTLUT	0.5	6.0	64.1	32.1	6.5	1.0	15.4	34.3	9.7	36.9	10.2	54.6
	CHIWAWA	71.7	**96.9**	99.5	98.1	91.5	**99.5**	97.8	96.9	76.1	99.0	84.4	99.4
F_1 score	Fmask	10.1	67.	78.9	82.1	44.4	71.7	30.0	73.5	67.6	74.7	81.0	90.6
	CDAG	60.3	57.	89.3	85.1	43.8	75.4	87.3	75.1	75.5	94.2	59.1	74.8
	COTLUT	0.9	11.	76.8	47.9	10.6	2.0	26.8	51.0	16.9	53.5	18.4	70.6
	CHIWAWA	**82.5**	**91.3**	**97.5**	**95.4**	**94.4**	**99.2**	**95.7**	**97.2**	**85.4**	**98.5**	**90.1**	**99.7**
ΔCC	Fmask	41.2	1.6	31.8	5.0	0.3	15.8	62.3	26.5	2.1	40.3	0.5	6.8
	CDAG	0.7	0.8	6.7	2.7	**0**	8.1	2.8	15.0	0.3	5.0	0.8	13.6
	COTLUT	2.2	1.8	19.9	8.0	0.1	20.9	11.3	25.1	6.9	36.8	1.3	15.7
	CHIWAWA	**0.7**	**0.2**	**2.4**	**0.7**	**0**	**0.1**	**0.6**	**0.2**	**1.8**	**0.6**	**0.2**	**0.2**

are highly dependent on spatial resolution. Thus, for images with decametric resolution, the PRISMA parameterization can be used initially, then fine tuned with a dedicated sensitivity study. Similarly, for metric resolution images, the parameterization used for AVIRIS-NG can be used first and then tuned.

E. Detection performance on PRISMA images

Performance of Fmask, CDAG, COTLUT and CHIWAWA methods over PRISMA images are shown in Table V. In the following, $\overline{m}$ is the mean value of metric m over all PRISMA images. Overall accuracy is very high for CHIWAWA for all images ($\overline{OA}$ = 99%) and quite significantly exceeds CDAG ($\overline{OA}$ = 94%), COTLUT ($\overline{OA}$ = 87%) and Fmask ($\overline{OA}$ = 80%). OA is a global metric that erases detection performance when cloud cover is low (TN then has a dominant impact on this metric). The P, R and F_1 metrics, more focused on clouds, vary greatly from one method to another. Fmask is not very accurate but misses few cloudy pixels ($\overline{P}$=52%, $\overline{R}$=96%, $\overline{F_1}$=64%). In contrast, CDAG is quite accurate but misses many cloudy pixels ($\overline{P}$=89%, $\overline{R}$=64%, $\overline{F_1}$=73%). COTLUT is relatively accurate but misses most cloudy pixels ($\overline{P}$=81%, $\overline{R}$=23%, $\overline{F_1}$=32%). CHIWAWA detects most cloud pixels with great accuracy, ($\overline{P}$=96%, $\overline{R}$=94%, $\overline{F_1}$=95%) clearly outperforming other methods, especially for detecting thin clouds (see Fig. 12).

Let's look at CHIWAWA in more detail. The greater the cloud cover, the better the performance. High cloud cover is often associated with extensive and dense clouds, which are easier to detect. F_1 score ranges in 94–100% for PRISMA images with a cloud cover higher than 10%. For the images studied, low cloud cover implies the presence of clouds of limited extent (see P_1 in Fig. 15). P_1, P_2 and P_{11} contain many small clouds of only a few pixels each. Cloud edges being difficult to locate in practice, especially when the ground truth is built, the comparison between ground truth and detection mask is tricky for small clouds, and this has an impact on the computation of performance. However, performance is satisfactory for these images (F_1 ranges in 83–91%). If all images are combined to form a single image (approximately 2.35 millions cloud pixels), F_1 score is 98%. For illustration, cloud masks for PRISMA images P_1–P_{12} are shown in Fig. 13.

In terms of cloud cover assessment, the values estimated by CHIWAWA are very consistent with the ground truth: $\Delta CC = 0.6\% \pm 0.7\%$. CHIWAWA remains accurate for images that do not contain clouds: the number of false positive pixels is less than 200 for images P_{16} , P_{17} and P_{18}, corresponding to a very accurate estimate of cloud cover ($\Delta CC \leq 0.02\%$).

Fig. 15 shows specific cases where CHIWAWA fails to detect clouds. In images P_1 and P_9, some clouds are not detected because their WV values are not low enough and are not contrasted with their environment.

F. Detection performance on AVIRIS-NG images

The performance obtained with AVIRIS-NG and PRISMA images cannot simply be compared from one sensor to the other. Indeed, for AVIRIS-NG images, only images with low or moderate cloud cover were selected: one reached 27% and the others ranged from 0.4% to 14.5%, while PRISMA's cloud cover varied from 0% to 60%.

As cloud cover is generally low, the average overall accuracy $\overline{OA}$ is quite high for all methods: 96% for Fmask, 95% for CDAG, 98% for CHIC and 98% for CHIU. When it comes to clouds more precisely, Fmask ($\overline{F_1}$=63%) and CDAG ($\overline{F_1}$=47%) do not combine high precision and high recall. CHIC ($\overline{F_1}$=87%) and CHIU ($\overline{F_1}$=88%) offer a good compromise between precision and recall on all the images. When a distinction is made between urban and non-urban areas, F_1 score is even higher for CHIC ($\overline{F_1}$=91%) and CHIU ($\overline{F_1}$=91%).

TABLE VI
DETECTION PERFORMANCE OF FMASK, CDAG, CHIC AND CHIU METHODS ON AVIRIS-NG IMAGES A_1–A_9. IMAGES CONTAINING MAINLY URBAN AREAS ARE MARKED WITH (U). THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

	Images	A_1^u	A_2^u	A_3	A_4	A_5	A_6	A_7^u	A_8^u	A_9
	CC (%)	1.8	16.5	29.4	5.6	11.6	1.8	0.2	1.9	20.6
	N_{cloud}	20k	814k	1097k	267k	359k	40k	9k	90k	1206k
OA	Fmask	97.6	94.9	95.1	96.0	95.5	98.2	99.7	98.7	91.7
	CDAG	98.4	94.9	87.6	95.4	**99.5**	98.0	99.6	98.1	86.7
	CHIC	99.3	97.1	**97.2**	**98.5**	98.8	**99.6**	99.7	99.6	**98.6**
	CHIU	**99.6**	**98.5**	95.2	98.4	97.3	99.4	**99.9**	**99.8**	97.3
Precision	Fmask	41.2	78.0	91.4	**100**	99.5	49.7	29.8	69.5	**99.1**
	CDAG	61.2	**96.6**	93.0	**100**	**99.6**	40.8	14.0	48.5	98.1
	CHIC	77.5	84.6	93.4	96.5	95.1	86.7	56.2	83.7	97.1
	CHIU	**93.4**	95.5	**97.6**	96.5	96.1	**91.0**	**86.1**	**97.2**	**99.1**
Recall	Fmask	70.7	96.5	92.0	29.1	61.2	47.5	36.5	54.8	60.4
	CDAG	36.8	75.7	62.8	17.8	48.6	27.5	16.6	17.5	36.1
	CHIC	**86.8**	**97.8**	**96.4**	**73.6**	**93.3**	**86.7**	**87.5**	**94.1**	**93.4**
	CHIU	82.5	94.0	84.4	70.5	77.2	68.1	**86.8**	90.8	81.9
F_1 score	Fmask	52.1	86.3	91.7	45.1	75.8	48.6	32.8	61.3	75.0
	CDAG	46.0	82.3	74.9	30.1	65.3	32.9	15.2	25.8	52.8
	CHIC	81.9	90.7	**94.9**	**83.5**	**94.2**	**86.7**	68.4	88.6	**95.2**
	CHIU	**87.6**	**94.7**	90.5	81.5	85.6	77.9	**86.5**	**93.9**	89.7
ΔCC	Fmask	1.3	3.9	**0.2**	4.0	4.5	0.1	0.1	0.4	8.0
	CDAG	0.7	2.4	9.6	4.6	5.9	0.6	**0**	1.2	13.0
	CHIC	**0.2**	2.3	0.9	**1.2**	**0.2**	**0**	0.2	0.2	**0.5**
	CHIU	**0.2**	**0.2**	3.7	1.5	1.9	0.5	**0**	**0**	**0.5**

As for the PRISMA images, the greater the cloud cover, the better the performance. For studied AVIRIS-NG images, when the cloud cover is above 10% (A_2, A_3, A_5 and A_9), many dense and large clouds are present. If the appropriate set of parameters is used (CHIU for A_2, CHIC for A_3, A_5 and A_9), the detection is quite accurate (F_1 in 94–95%).

For lower cloud cover (mainly in urban areas for the images used), the clouds are often thin. The number of false alarms can be significant when using the CHIC setting, and are mainly located on roads and certain buildings. For urban scenes, the CHIU setting is more appropriate: it provides high precision ($P > 90\%$ except for image A_7), the recall is $88\% \pm 4\%$ and the average F_1 score is $91\% \pm 4\%$ (whereas it is $82\% \pm 9\%$ for CHIC).

If all the images are combined into a single image (approximately 3.911 millions cloud pixels), F_1 score is 93% for CHIC and 90% for CHIU. For illustration, cloud masks for several regions of interest are shown in Fig. 14.

Fig. 16 shows some special cases where CHIU and CHIC can perform differently: detection of thin clouds with CHIC but not with CHIU, false cloud detections over buildings and roads with CHIC but not with CHIU. It also shows the impossibility of detecting cloudy pixels for which the WV values have little contrast with the background, as in A_4 image. Fmask and CDAG do not detect these clouds either (see Fig. 12). The distribution of cloud altitudes or thicknesses can induce strong variations on the WV map histogram; it is then possible that clouds associated with the highest WV values remain undetected even after two successive CHIWAWA launches (see Fig. 16, first row). Such cases are rare, however, and would correspond to very thin clouds.

Cloud Cover is quite accurately estimated by CHIC and CHIU: mean ΔCC is $0.6\% \pm 0.7\%$ for CHIC and $0.9\% \pm 1.2\%$ for CHIU. Performance is also assessed for images without any clouds. Ten ROIs from A_6, A_7, A_8 (of about one million pixels each) were selected. Among these, three images are very urbanized and mean ΔCC is $3.8\% \pm 2.4\%$ for CHIC and $0.4\% \pm 0.5\%$ for CHIU. For the other images that are not very urbanized, mean ΔCC is $0.4\% \pm 0.5\%$ for CHIC and $0.1 \pm 0.1\%$ for CHIU. Overall, therefore, CHIU appears to be more accurate than CHIC for estimating cloud cover, whatever the type of cloud cover. CHIWAWA can be applied to an AVIRIS-NG image as follows. CHIU is first applied to estimate the cloud detection mask and cloud cover. If estimated cloud cover is lower than a given threshold, e.g. 10%, then CHIU detection mask is retained. If it exceeds this threshold, the CHIC method is applied to improve the cloud detection mask.

Another way of choosing between CHIC and CHIU is to evaluate the level of urbanization of the scene observed. For instance, an unsupervised classification method such as CHRIPS [33] could be used to assess the distribution of classes in the image (asphalt, house roofs, different types of vegetation, etc.). The same image could thus be divided into different zones to which CHIC or CHIU will be applied according to the level of urbanization detected.

G. Detection performance on mountain images

The application of the regression method described in section III-B is illustrated here on mountain images (see Figs. 17, 18 and 19). In these images, the WV maps are represented with RGB colors, which makes it easier to see the changes obtained after correction of altitude variations. In order to equalize the distribution of pixels according to altitude (some altitudes have many more pixels than others), the altitude axis is divided into 25 m slices and the average WV value is calculated for all the pixels within each slice (red triangles in the figures below). The fit between WV and altitude is calculated using these average values.

In Fig. 17 (ROI from A_9), altitude correction allows to reduce the number of false detections and to better detect thin clouds: F_1 score is 90% without correction and reaches 99% with correction. In Fig. 18 (another ROI from A_9), studied image is non-cloudy but a local area with a significantly higher altitude than elsewhere is associated with much lower WV values, leading to erroneous cloud detection if no correction is made. After altitude correction, the CHIWAWA method no longer detects clouds in the area.

Fig. 19 deals with the PRISMA Cuprite image P_{13}. In this scene, the WV distribution is not only correlated with altitude: WV increases gradually from the left to the right of the image. Altitude correction does not alter the WV map very much. The image contains many rocky pixels which are retained in the mask of potential cloud pixels m_0 when the value of T_{VNIR} is 0.07 (CHIP), leading to a final detection mask with many false detections (P = 6%, R = 99%, F_1 = 11%). Rocky pixels can be largely eliminated from m_0 using a value of 0.15 for T_{VNIR}, resulting in very accurate cloud detection (P $\approx$ 100%, R = 95%, F_1 = 97%). In general, in the event of unsatisfactory performance for the user following the execution of CHIWAWA with proposed standard parameters, it is very often possible to improve detection by simply modifying a few key thresholds such as T_{VNIR} or T_{SWIR}.

H. Detection performance on images containing snow or ice

In the PRISMA image P_{14} image, the surface is mostly covered with ice (see Fig. 20). Constructing the ground truth for this type of image is very complex, as the scene contains many small, relatively thin clouds. There are many uncertainties regarding the location of the boundaries of each cloud, which show very little contrast with the environment whatever the spectral band in the VNIR and SWIR ranges. The obtained performances should therefore be interpreted with caution.

The average WV level is very low (between 0.25 and 0.36 g/cm^2) and WV variations are difficult to detect. The M^T and M^C methods are based on WV map processing and fail to detect most scattered clouds except for the cloud at bottom left. Cloud pixels are mostly detected as very bright pixels (see section II-C). Detection performance is quite accurate: P = 95%, R = 82%, F_1 = 88%.

Ice spectra include absorption around 1030 nm, reflectance growth in the range [1500-1750nm] and a local maximum around 2250 nm as shown in the section II-D. Pixels containing clouds also check these properties (see Fig. 20), but with a spectral offset. The detection of this offset can be used directly to identify clouds, particularly in the SWIR range at 2350 nm, where ice reflectance should be very low (typically less than 0.1) when there are no clouds as it is done with the criteria dedicated to very bright pixels detection (see section II-C). A simple cloud detection method for this kind of ice-covered image could just be to threshold ρ_{2350}: the reflectance should be close to 0 when the pixel contains only ice, but becomes higher in the presence of clouds. Tests have shown that using a threshold between 0.09 and 0.11 gives the following detection performance, which is very similar to those of CHIWAWA: P = 95%, R = 80%, F_1 = 87%.

Fig. 21 deals with the PRISMA image P_{15} which is quite complex to process: it contains mountains, snow and clouds with different thicknesses. Altitude correction is performed on it. Detection performance is quite accurate: P = 89%, R = 99%, F_1 = 94%. False detections can easily be removed by deleting clouds that are too narrow (containing fewer than 10 pixels). Accuracy increases significantly and recall is little impacted: P = 96%, R = 97% , F_1 = 96%. On the whole, there is no confusion between clouds and snow. For illustration, CHRIPS classification method [33] is applied to obtain a classification map including clouds.

V. Conclusion

The CHIWAWA cloud detection method has been presented: it exploits the complementarity of the ground reflectance image and the WV map to detect thick and thin clouds. A method for detecting snow and ice is also proposed, as well as a method for automatically correcting the altitude dependency of the WV map in the case of scenes with relief.

CHIWAWA has been assessed on images acquired with two different sensors with very different spatial resolutions: PRISMA and AVIRIS-NG. It clearly outperforms the other methods tested, especially on thin clouds. For both sensors, overall accuracy ranges in 97–100% and the cloud cover estimation error averages 1%. For PRISMA images, the F_1 score ranges in 95–100% when cloud cover exceeds 10%, and in 83–94% for smaller cloud covers. For AVIRIS-NG images, the F_1 score ranges in 94–95% when cloud cover exceeds 10%. For lower cloud cover (corresponding to urban images), the F_1 score ranges in 84–91%.

Some limitations of the CHIWAWA method have been observed. Thin clouds are difficult to detect in urban areas, as they can be confused with roads and buildings. Low clouds can also pose a problem, as the associated WV values contrast poorly with their surroundings. The presence of numerous bright pixels can significantly increase the false alarm rate but this problem can be highly reduced by modifying a few key thresholds.

Other ways of improvement can also be explored. Anthropogenic structures or classes of material that could cause problems (roads, buildings, water, rocks, etc.) could first be detected in areas that are not potentially cloudy. The resulting detection masks could then be used by CHIWAWA to detect thin clouds in potentially cloudy areas. Spatial characterization of clouds (texture) could also be used to better delineate cloud edges. The construction of indices combining the WV map and reflectance could be considered (initial tests proved unsuccessful but the idea still has potential). Finally, the simultaneous exploitation of ground reflectance, WV, the 760 nm oxygen A-band and the 1380 nm water vapor band (cirrus) could also be explored.

References

[1] A. Lyapustin, Y. Wang, and R. Frey, "An automatic cloud mask algorithm based on time series of modis measurements," *Journal of Geophysical Research: Atmospheres*, vol. 113, no. D16, 2008. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009641

[2] O. Hagolle, M. Huc, V. P. David, and G. Dedieu, "A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images," *Remote Sensing of Environment*, vol. 114, no. 8, pp. 1747–1755, Aug. 2010. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00489793

[3] H. Tang, K. Yu, O. Hagolle, K. Jiang, X. Geng, and Y. Zhao, "A cloud detection method based on a time series of modis surface reflectance images," *International Journal of Digital Earth*, vol. 6, no. sup1, pp. 157–171, 2013. [Online]. Available: https://doi.org/10.1080/17538947.2013.833313

[4] C.-H. Lin, B.-Y. Lin, K.-Y. Lee, and Y.-C. Chen, "Radiometric normalization and cloud detection of optical satellite images using invariant pixels," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 106, pp. 107–117, Aug. 2015.

[5] X. Zhu and E. H. Helmer, "An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions," *Remote Sensing of Environment*, vol. 214, pp. 135–153, 2018.

[6] F. Sedano, P. Kempeneers, P. Strobl, J. Kucera, P. Vogt, L. Seebach (Reithmaier), and J. San-Miguel-Ayanz, "A cloud mask methodology for high resolution remote sensing data combining information from high and medium resolution optical sensors." *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, pp. 588–596, 09 2011.

[7] N. Kussul, M. Korbakov, A. Kravchenko, and A. Shelestov, "Cloud mask extracting from meteosat data with use of parallel markovian segmentation algorithm," in *2005 IEEE Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, 2005, pp. 204–207.

[8] S. Le Hégarat and C. André, "Use of markov random fields for automatic cloud/shadow detection on high resolution optical images," *Isprs Journal of Photogrammetry and Remote Sensing - ISPRS J PHOTOGRAMM*, vol. 64, pp. 351–366, 07 2009.

[9] C. Latry, C. Panem, and P. Dejean, "Cloud detection with svm technique," in *2007 IEEE International Geoscience and Remote Sensing Symposium*, 2007, pp. 448–451.

[10] J. dong Jang, A. A. Viau, F. Anctil, and E. Bartholomé, "Neural network application for cloud detection in spot vegetation images," *International Journal of Remote Sensing*, vol. 27, no. 4, pp. 719–736, 2006.

[11] M. Segal Rozenhaimer, A. Li, K. Das, and V. Chirayath, "Cloud detection algorithm for multi-modal satellite imagery using convolutional neural-networks (cnn)," *Remote Sensing of Environment*, vol. 237, p. 111446, 02 2020.

[12] Y. Li, W. Chen, C. Tao, R. Xiao, and Y. Tan, "Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning," *Remote Sensing of Environment*, vol. 250, 08 2020.

[13] W. B. Rossow and L. C. Garder, "Cloud detection using satellite measurements of infrared and visible radiances for isccp," *Journal of Climate*, vol. 6, no. 12, 1993.

[14] L. L. Stowe, P. A. Davis, and E. P. McClain, "Scientific basis and initial evaluation of the clavr-1 global clear/cloud classification algorithm for the advanced very high resolution radiometer," *Journal of Atmospheric and Oceanic Technology*, vol. 16, no. 6, pp. 656 – 681, 1999.

[15] K. T. Kriebel, G. Gesell, M. Kästner, and H. Mannstein, "The cloud analysis tool apollo: Improvements and validations," *International Journal of Remote Sensing*, vol. 24, no. 12, pp. 2389–2408, 2003.

[16] R. Irish, "Landsat 7 automatic cloud cover assessment," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 4049, 08 2000.

[17] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in landsat imagery," *Remote Sensing of Environment*, vol. 118, pp. 83–94, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425711003853

[18] Z. Zhu, S. Wang, and C. E. Woodcock, "Improvement and expansion of the fmask algorithm: cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images," *Remote Sensing of Environment*, vol. 159, pp. 269–277, 2015.

[19] L. Sun, X. Mi, J. Wei, J. Wang, X. Tian, H. Yu, and P. Gan, "A cloud detection algorithm-generating method for remote sensing data at visible to short-wave infrared wavelengths," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 124, pp. 70–88, Feb. 2017.

[20] Y. Lin, L. He, Y. Zhang, and Z. Wu, "Cloud detection of gaofen-2 multi-spectral imagery based on the modified radiation transmittance map," *Remote Sensing*, vol. 14, no. 17, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/17/4374

[21] M. D. King, W. P. Menzel, P. S. Grant, J. S. Myers, G. T. Arnold, S. E. Platnick, L. E. Gumley, S.-C. Tsay, C. C. Moeller, M. Fitzgerald, K. S. Brown, and F. G. Osterwisch, "Airborne scanning spectrometer for remote sensing of cloud, aerosol, water vapor, and surface properties," *Journal of Atmospheric and Oceanic Technology*, vol. 13, no. 4, pp. 777 – 794, 1996.

[22] L. Guanter, R. Richter, and H. Kaufmann, "On the application of the modtran4 atmospheric radiative transfer code to optical remote sensing," *International Journal of Remote Sensing*, vol. 30, no. 6, pp. 1407–1424, 2009.

[23] R. B. A. Koelemeijer, P. Stammes, J. W. Hovenier, and J. F. de Haan, "A fast method for retrieval of cloud parameters using oxygen a band measurements from the global ozone monitoring experiment," *Journal of Geophysical Research: Atmospheres*, vol. 106, no. D4, pp. 3475–3490, 2001. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD900657

[24] B. Gao, A. F. H. Goetz, and W. J. Wiscombe, "Cirrus cloud detection from airborne imaging spectrometer data using the 1.38 μm water vapor band," *Geophysical Research Letters*, vol. 20, no. 4, p. 301 – 304, 1993.

[25] B.-C. Gao and A. F. H. Goetz, "Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data," *Journal of Geophysical Research: Atmospheres*, vol. 95, no. D4, pp. 3549–3564, 1990. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JD095iD04p03549

[26] D. Schläpfer, 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, 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425798000443

[27] R. Loizzo, R. Guarini, F. Longo, T. Scopa, R. Formaro, C. Facchinetti, and G. Varacalli, "Prisma: The italian hyperspectral mission," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 175–178.

[28] J. W. Chapman, D. R. Thompson, M. C. Helmlinger, B. D. Bue, R. O. Green, M. L. Eastwood, S. Geier, W. Olson-Duvall, and S. R. Lundeen, "Spectral and radiometric calibration of the next generation airborne visible infrared spectrometer (aviris-ng)," *Remote Sensing*, vol. 11, no. 18, p. 2129, Sep 2019. [Online]. Available: http://dx.doi.org/10.3390/rs11182129

[29] D. Hall, J. Foster, D. Verbyla, A. Klein, and C. Benson, "Assessment of snow-cover mapping accuracy in a variety of vegetation-cover densities in central alaska," *Remote Sensing of Environment*, vol. 66, no. 2, pp. 129–137, 1998.

[30] S. Singh, A. Kulkarni, and B. Chaudhary, "Hyperspectral analysis of snow reflectance to understand the effects of contamination and grain size," *Annals of Glaciology*, vol. 51, no. 54, p. 83–88, 2010.

[31] R. N. Clark and T. L. Roush, "Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications," *Journal of Geophysical Research*, vol. 89, no. B7, pp. 6329–6340, 1984.

[32] R. N. Clark, G. Swayze, K. E. Livo, R. Kokaly, S. J. Sutley, J. Dalton, R. R. McDougal, and C. A. Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the usgs tetracorder and expert systems," *Journal of Geophysical Research*, vol. 108, 2003.

[33] A. Alakian and V. Achard, "Classification of hyperspectral reflectance images with physical and statistical criteria," *Remote Sensing*, vol. 12, no. 14, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/14/2335

[34] R. Kokaly, R. Clark, G. Swayze, K. Livo, T. Hoefen, N. Pearson, R. Wise, W. Benzel, H. Lowers, R. Driscoll, and A. Klein, "USGS spectral library version 7 data," *U.S. Geological Survey data release*, 2017.

[35] Y. Zhang, B. Guindon, and J. Cihlar, "An image transform to characterize and compensate for spatial variations in thin cloud contamination of landsat images," *Remote Sensing of Environment*, vol. 82, pp. 173–187, 10 2002.

[36] "Prisma algorithm theoretical basis document (atbd)," Tech. Rep., 2021.

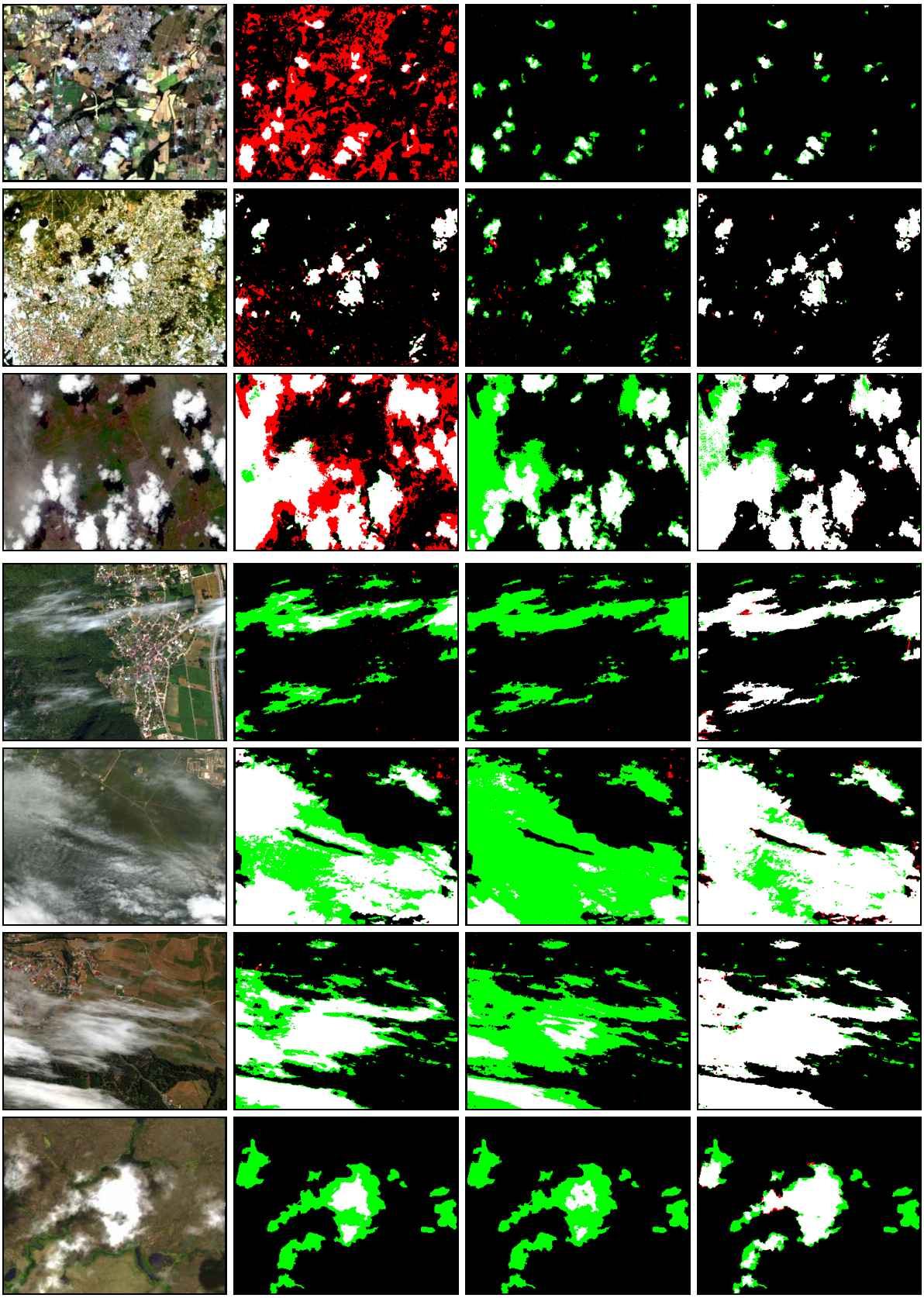

Fig. 12. Comparison of thin cloud detection performance with the Fmask, CDAG and CHIWAWA methods on PRISMA images P_1 (first row), P_2 (second row), P_8 (third row) and AVIRIS-NG images A_8 (fourth row), A_3 (fifth row), A_9 (sixth row) and A_4 (seventh row). Thin clouds are mainly detected by CHIWAWA.

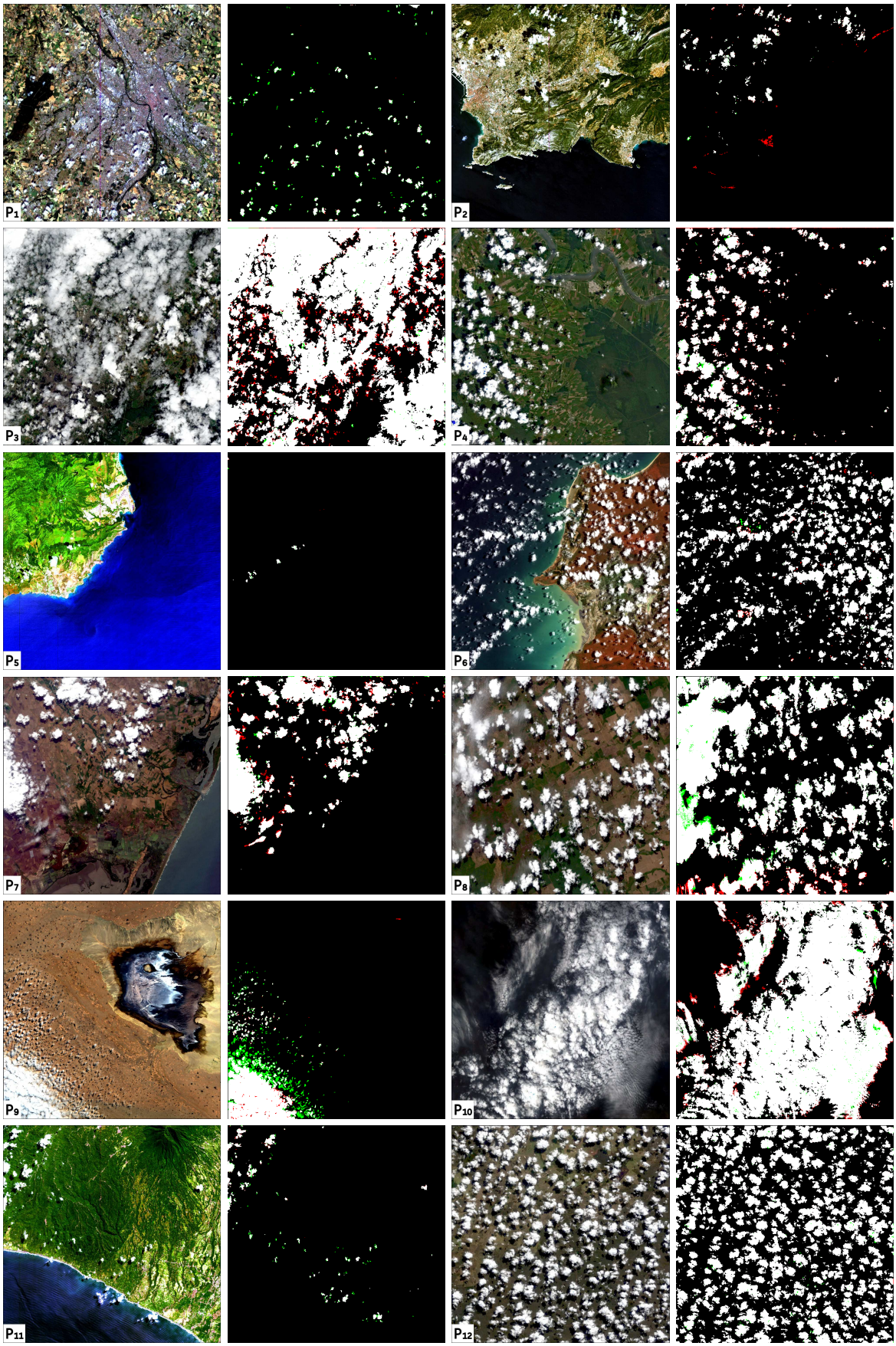

Fig. 13. Cloud masks estimated with the CHIWAWA method for PRISMA P_1–P_{12} images.

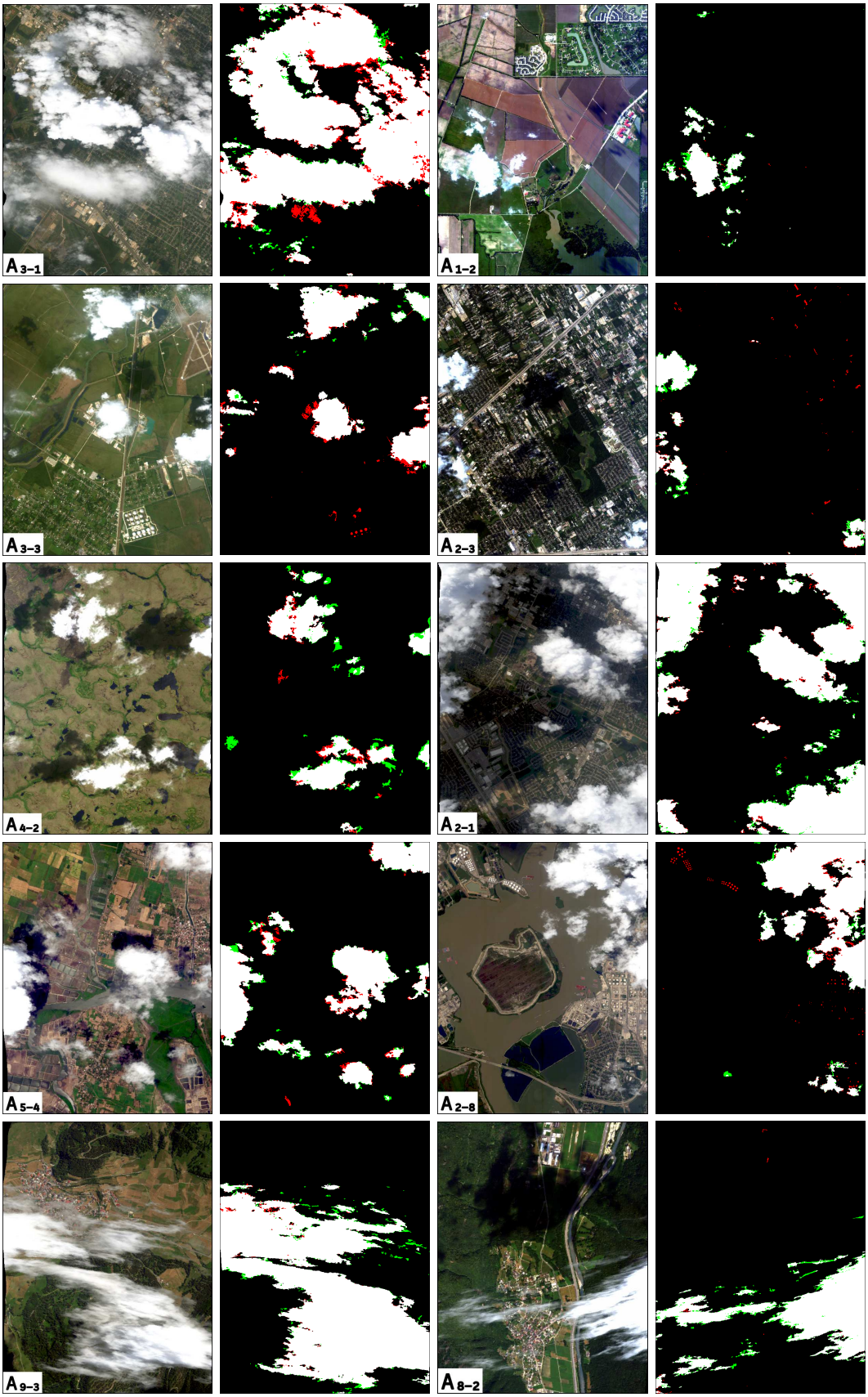

Fig. 14. Cloud masks estimated with CHIC (non-urban areas) and CHIU (urban areas) for regions of interest from different AVIRIS-NG images. For notation purposes, A_{6-3} is ROI number 3 of image A_6.

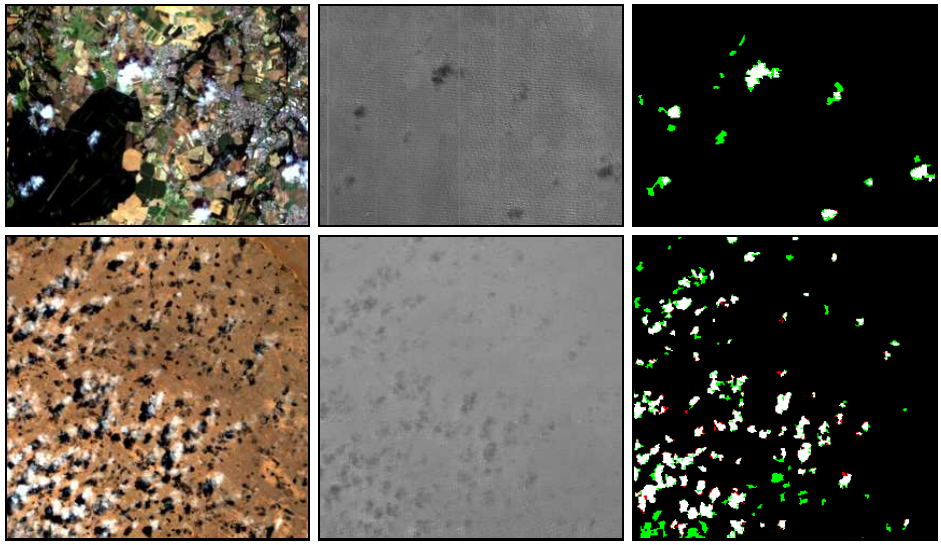

Fig. 15. Problems that may arise when applying CHIWAWA to PRISMA images. Two regions of interest from images P_1 (first row) and P_9 (second row) are shown. In both images, small clouds or cloud parts with insufficiently low WV values (values not contrasted with surrounding pixels) are not detected by CHIWAWA.

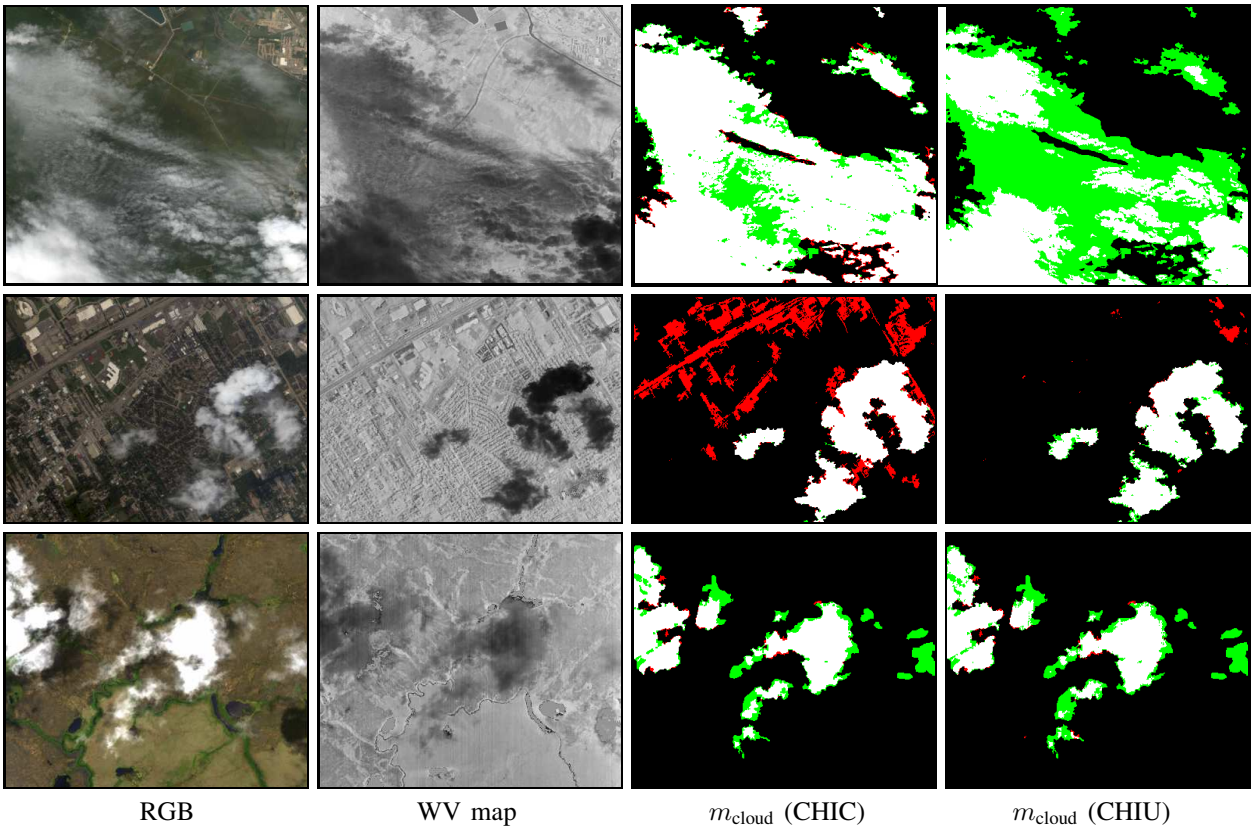

Fig. 16. Problems that may arise when applying CHIWAWA to AVIRIS-NG images: comparison between CHIC and CHIU. Three regions of interest from images A_3 (first row), A_2 (second row) and A_4 (third row) are shown. In A_3, thin clouds are quite well detected by CHIC (with the exception of very thin clouds) but only partially detected by CHIU. In A_2, some buildings and roads are mistaken for clouds by CHIC, but not by CHIU. in A_4, the thin parts of the clouds have WV values that contrast poorly with the background: CHIC and CHIU are unable to detect them correctly.

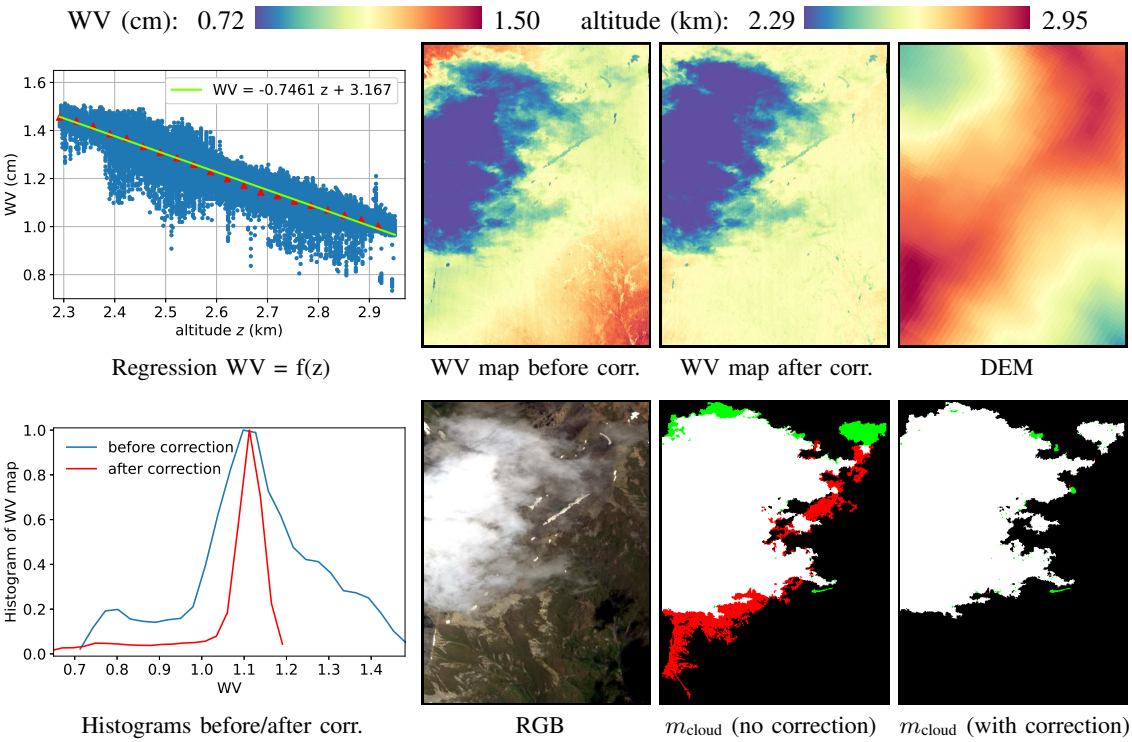

Fig. 17. WV map altitude correction for a cloudy ROI in image A_9 (Swiss mountains) and impact on cloud detection mask. Distribution of couples (altitude z, WV value) for pixels that are not potentially cloudy (regression line is green) and histograms of WV values before and after altitude correction are shown, as well as WV maps before and after altitude correction with associated digital elevation model, and cloud detection masks without or with altitude correction of WV map. In this scene, correction allows to reduce the number of false detections (spatial extension of the big cloud) and to better detect thin clouds: P = 85%, R = 94%, F_1 = 90% before correction, P $\approx$ 100%, R = 99%, F_1 = 99% after correction.

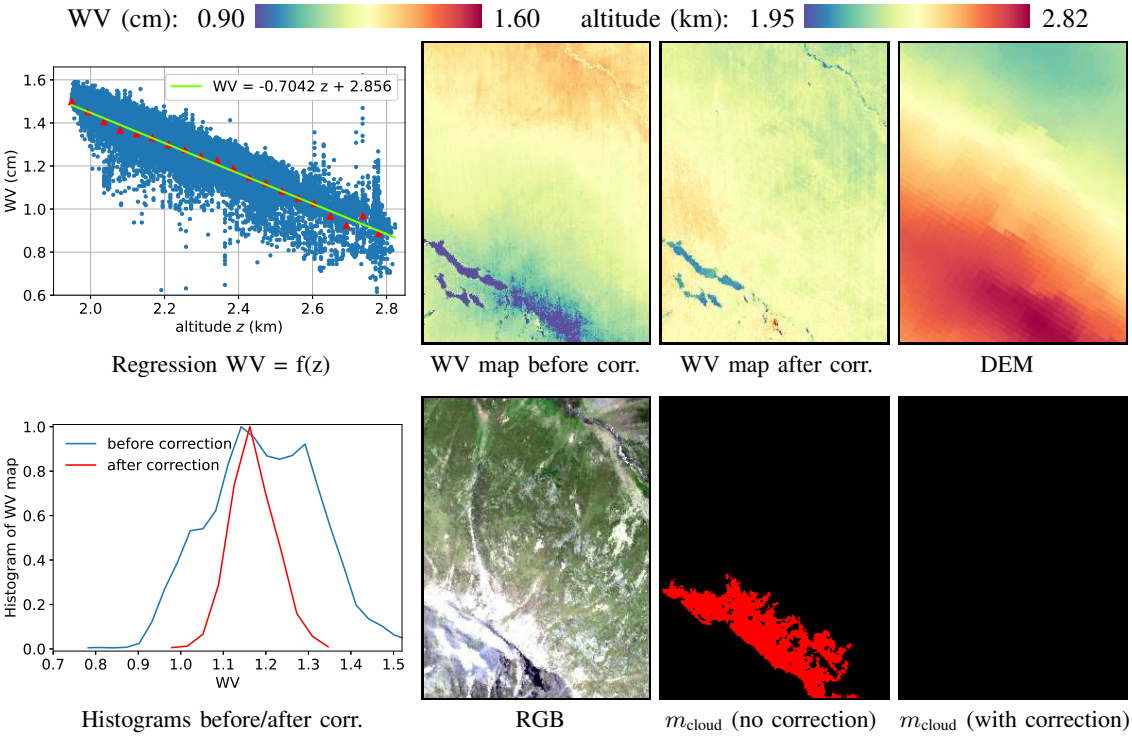

Fig. 18. WV map altitude correction for a non-cloudy ROI in image A_9 (Swiss mountains) and impact on cloud detection mask. Distribution of couples (altitude z, WV value) for pixels that are not potentially cloudy (regression line is green) and histograms of WV values before and after altitude correction are shown, as well as WV maps before and after altitude correction with associated digital elevation model, and cloud detection masks without or with altitude correction of WV map. In this scene, altitude correction avoids false cloud detections in a mountainous area.

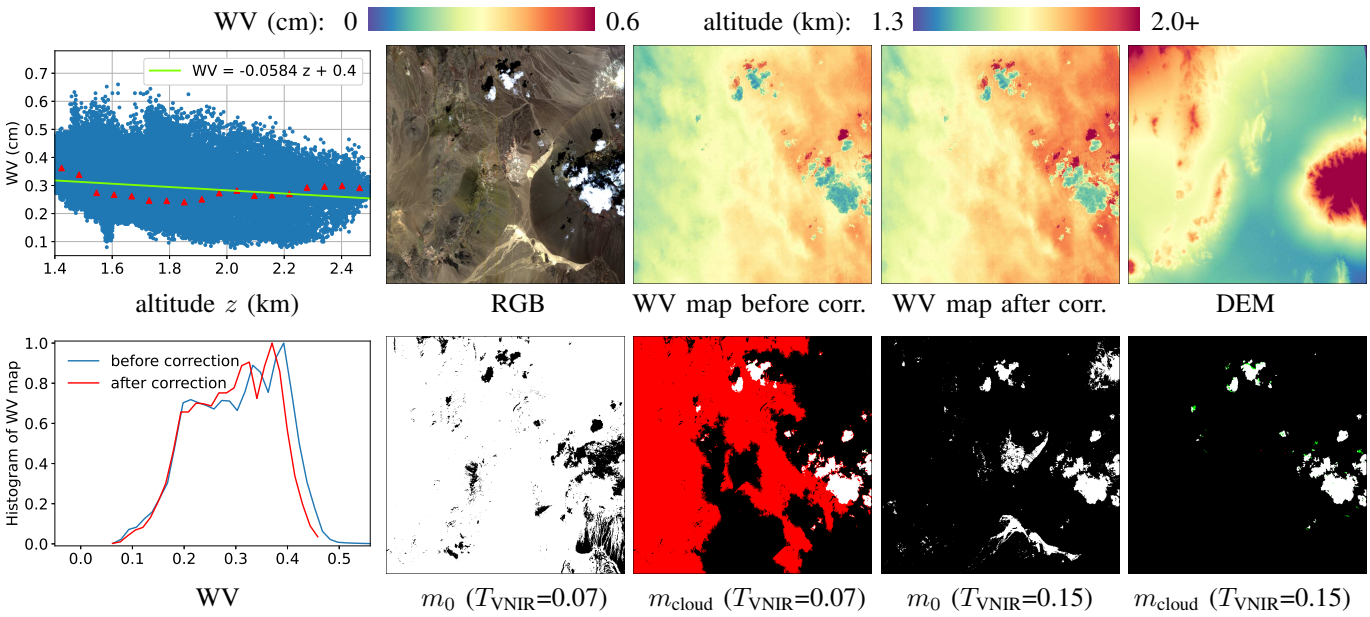

Fig. 19. Altitude correction of the WV map for the P_{13} cloudy image (Cuprite, USA) and impact on the cloud detection mask. Top, distribution of pairs (altitude z, WV value) for pixels that are not potential clouds and histograms of WV values before and after altitude correction. In the middle, masks of WV values before and after altitude correction with the associated digital elevation model: here, altitude correction has little impact on WV distribution and does not really help for cloud detection. However, the value of the threshold T_{VNIR} has a significant impact. A value of 0.07 implies a large number of potential cloud pixels (rocky surfaces) and a poor quality detection mask (P = 6%, R = 99%, F_1 = 11%). Increasing T_{VNIR} to 0.15 removes most of the rocky pixels from m_0 and enables fairly accurate cloud detection (P $\approx$ 100%, R = 95%, F_1 = 97%).

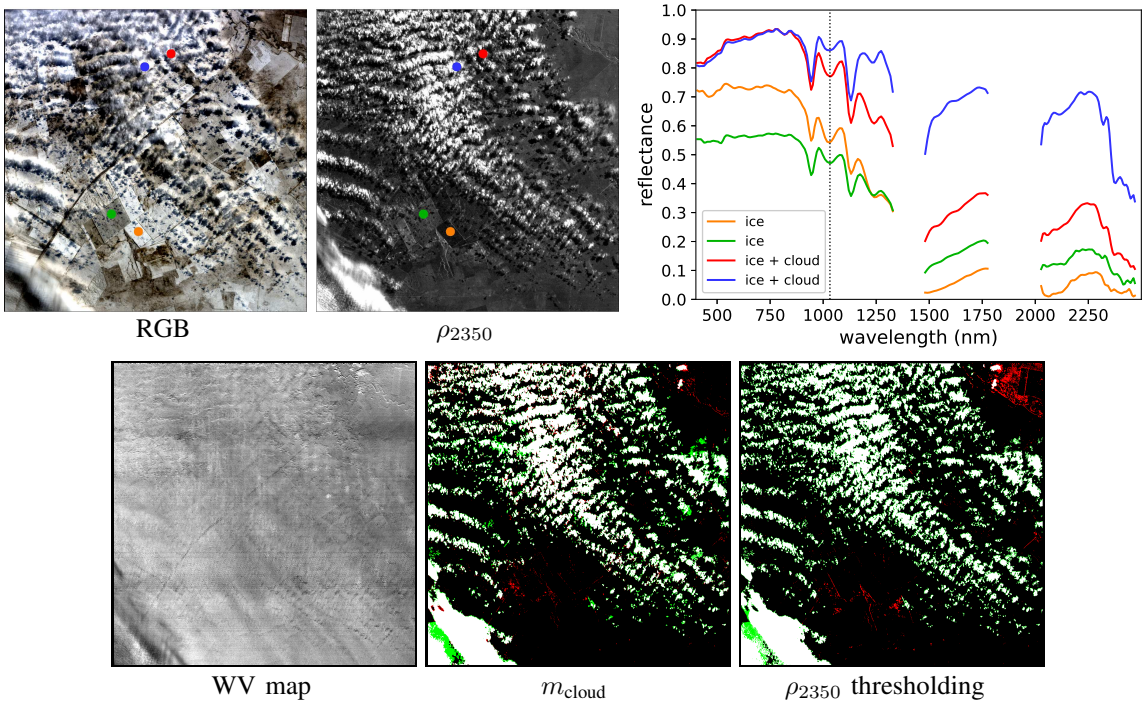

Fig. 20. Reflectance spectra of the ice-covered PRISMA P_{14} image and cloud detection maps. The spectral shapes are similar whether clouds are present or not, the amplitude increases significantly in the presence of clouds in full spectral range. The presented cloud detection maps are obtained by applying CHIWAWA or by simply thresholding ρ_{2350} (threshold is 0.10). Respective performances are (P = 95%, R = 82%, F_1 = 88%) and (P = 95%, R = 80%, F_1 = 87%).

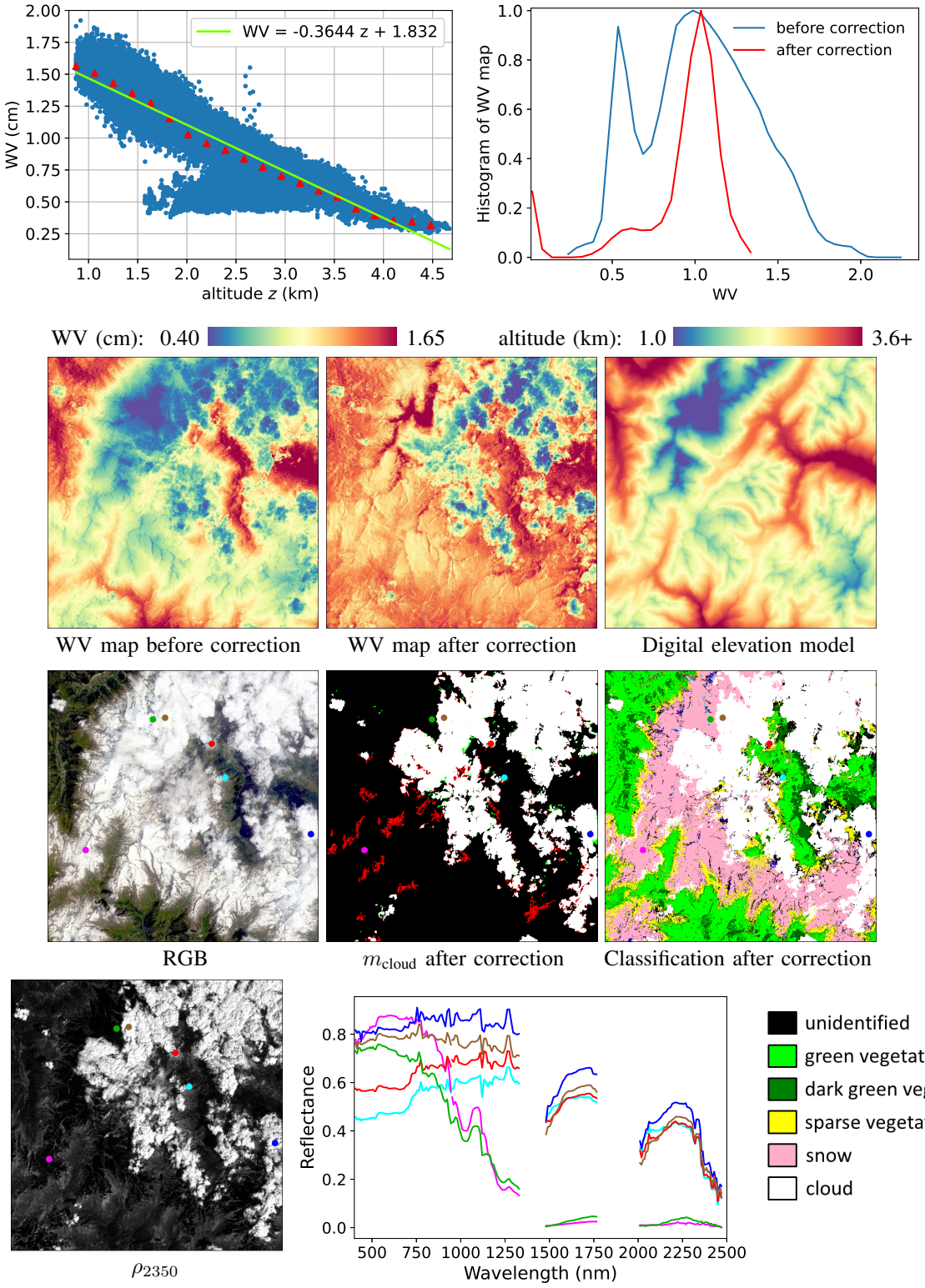

Fig. 21. WV map altitude correction for the PRISMA image P_{15} and impact on cloud detection mask. This image contains mountains, snow and clouds. At the bottom, cloud detection mask and a classification map performed with CHRIPS [33] are shown. Reflectance spectra of snowy pixels and pixels of clouds with different thicknesses are also shown. Specific absorption around 1030 nm can be observed for snow pixels (pink and green spectra). The red spectrum corresponds to a pixel detected as very bright and then identified as cloudy: without this processing (see section II-C), it would not be detected as cloudy because its WV value is too high. Detection performance is quite accurate: P = 89%, R = 99%, F_1 = 94%. It can be seen that, as with the ice-covered image P_{14}, ρ_{2350} easily distinguishes clouds from snow and could be used as a discriminative feature to detect clouds in this type of scene.