https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Targeted Spectral Threshold Tests for Reducing Cloud Omission in the GOSAT-2 CAI-2 Cloud Mask

Mohammadreza Safabakhshpachehkenari[1], Tsuneo Matsunaga[1]

[1]Satellite Observation Center, Earth System Division, National Institute for Environmental Studies (NIES), 16-2 Onogawa, Tsukuba, Ibaraki, 305-8506, Japan

Correspondence to: Mohammadreza Safabakhshpachehkenari (safabakhsh.reza@nies.go.jp)

Abstract. Cloud detection is a common problem in optical remote sensing, but its consequences vary by mission. In GOSAT-2, missed clouds matter because they can affect greenhouse gas retrievals. This research introduces a targeted omission-recovery approach for the GOSAT-2 CAI-2 Level 2 cloud mask over water. Rather than replacing the operational mask, the method identifies and corrects its errors. True-positive (TP) and false-negative (FN) cloud footprints are mapped into two-dimensional spectral feature spaces proposed from native CAI-2 L1B ultraviolet, visible, and near-infrared radiances. Polynomial logistic regression was used to estimate the probability that a cloud footprint belonged to the detected-cloud group. A Kolmogorov–Smirnov test then compared false negatives at or above the weakest detected TP level with those below it, and the three feature spaces that separated these groups most clearly across granules were converted into probability-threshold tests. Finally, these tests were applied only to pixels labeled no-cloud in the original L2 mask, with additional controls such as sun-glint filtering. Among four overwater granules around Japan, the revised mask reduced cloud omission by 64–83%, while overestimation increased by 0 to 4 percentage points. Cohen's kappa increased by 5–14 points, reaching 0.88 to 0.93. In the independent granule, the tests recovered 55.4% of the clouds missed by the L2 mask while introducing false positives over only 2.07% of cloud-free pixels. Also, an indicator semivariogram showed that recovering the missed clouds improves the mask most at the scales of edges and small and scattered clouds. These results show that missed clouds in CAI-2 contain recoverable spectral structure, which can be used to improve the operational CAI-2 cloud mask over the ocean.

Key words: Satellite observation, GOSAT-2 TANSO-CAI-2 L2, Cloud detection, Spectral tests development, Polynomial logistic regression

1. Introduction

An effective response to climate change rests on the ability to measure it. Satellite remote sensing records radiation emitted or reflected by the atmosphere, surface, and shallow subsurface without physical contact, and it yields continuous, globally consistent observations on timescales relevant to both weather and climate policy. Among the variables it records are concentrations of carbon dioxide and methane in the atmosphere, which together account for more than 80% of anthropogenic radiative forcing (IPCC, 2021). Mapping the spatial distribution of these two gases, their temporal variability, and the long-

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

term trends in their concentrations remains essential for elucidating their atmospheric budgets (UNFCCC, 2015). The United Nations Environment Programme, in its 2026–2029 medium-term strategy, identifies remote sensing as an instrument that will increasingly provide data insight for better decision-making in climate action.

The first spacecraft specifically built to measure atmospheric carbon dioxide, and methane was the Greenhouse gases Observing Satellite (GOSAT), launched by Japan in 2009, operates in a sun-synchronous orbit at an altitude of 666 km with a 3-day repeat cycle. Its spectrometer, the Thermal And Near-infrared Sensor for carbon Observation–Fourier Transform Spectrometer (TANSO-FTS), measures reflected sunlight in the short-wavelength infrared and thermal emission in the long-wave. From these spectra, it retrieves the column-averaged concentrations of both gases. Although designed for a five-year mission, GOSAT has operated for more than fifteen years. That has an advantage. A single instrument under a consistent onboard calibration that can track year-by-year global and regional concentration changes. Continuity of this measurement standard motivated the successor mission. GOSAT-2, launched in October 2018 into a 613 km sun-synchronous orbit with a six-day revisit, provides continuous monitoring of the column-averaged dry-air mole fractions of carbon dioxide and methane, XCO_2 and XCH_4. The second-generation spectrometer, TANSO-FTS-2, raises the signal-to-noise ratio across both the short-wavelength and thermal infrared and extends its spectral coverage to carbon monoxide near 2.3 μm, so that carbon dioxide, methane, and carbon monoxide are retrieved within a single sounding. Such precision depends on a viewing path free of cloud and aerosol. GOSAT-2 therefore carries a second instrument for this purpose, the Cloud and Aerosol Imager-2 (TANSO-CAI-2). It is a forward- and backward-viewing imager and covers a broad (1000 km) cross-track swath to identify the observations suitable for retrieval. TANSO-CAI-2 has ten bands, covering ultraviolet, visible (red and blue), near-infrared, and shortwave-infrared wavelengths; five are observed forward and five backward, at viewing angles of up to +20° and −20°, respectively.

The importance of cloud screening in remote sensing is usually explained by two related factors. One is the annual cloud frequency. The International Satellite Cloud Climatology Project (ISCCP) puts the global mean cloud fraction at about 66%. The other is retrieval bias caused by clouds. Accurate retrieval of surface and atmospheric properties usually requires cloud-contaminated observations to be removed, such as XCO_2 estimation from GOSAT-2. Although these two facts explain the need for cloud screening, the literature points to a broader context. Existing cloud-detection algorithms include dynamic thresholding and Bayesian classification for multichannel imagers (Bulgin et al., 2014); tiered threshold tests for aerosol-oriented retrievals (Mei et al., 2017); fuzzy-logic spectral threshold masking (Frey et al., 2020); spatiotemporal thresholding with object-based refinement (Wang et al., 2021); visible–infrared cloud-probability masking (Hünerbein et al., 2023); Function of Mask algorithm (Fmask) cloud and shadow masking (Zhu et al., 2017); parallax-based Cloud Displacement Index refinement (Frantz et al., 2018); and multitemporal blue-reflectance and neighborhood-correlation detection (Hagolle et al., 2010). Other schemes are instrument-specific, designed for hyperspectral sensors, cloud radars, and individual cloud types such as cirrus and low stratus. The breadth of these approaches reflects the inherent difficulty of cloud detection, because no single algorithm performs uniformly across sensors, surfaces, and cloud regimes.

Cloud detection is well-suited to Machine Learning (ML) because it often involves multivariate classification. Early ML cloud-masking studies often paired learned classifiers with spatial or geometric checks. For example, Hughes and Hayes (2014)

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

developed a neural network for automated removal of cloud and shadow in a Landsat scene. Relative to Fmask, the method reduced cloud-shadow omission errors, with only slightly higher cloud-omission errors. Use of Support Vector Machines (SVMs) and Random Forests (RFs) later expanded across sensors and platforms, as they are expected to be less prone to overfitting. Bai et al. (2016) used an SVM with spectral, textural, and contextual features, followed by object-oriented post-processing, and reported more than 91% overall accuracy for high-resolution cloud detection. Using a different set of features, Joshi et al. (2019) applied an SVM based on tasseled-cap and shortwave-infrared information. Their method outperformed CFmask by about 3.8% in F-measure. Thampi et al. (2017) used RF to classify CERES field-of-view observations as clear or cloudy, while Haynes et al. (2019) used RF with GOES ABI data to improve low-cloud detection in multilayer cloud scenes. Reference data, however, remain a limiting factor for supervised cloud detection. Wang et al. (2020) partly addressed this by training RF models on CALIOP lidar labels to retrieve VIIRS cloud presence and phase, while noting that label quality and sampling still require attention. Deep Learning (DL) reduces manual feature design by learning useful patterns from the data. Zi et al. (2018) used a PCANet-based model for Landsat 8 masking. Mateo-García et al. (2018) treated masking as multitemporal change detection against a clear reference image and outperformed Fmask. Later gains came largely from improved DL architectures. Chen et al. (2018) coupled superpixel segmentation with multiple convolutional neural networks (CNNs) to classify cloud, shadow, and clear pixels. Puigdollers et al. (2021) benchmarked DL detectors against operational threshold methods on Landsat-8 and Sentinel-2, finding strong within-dataset accuracy but weaker cross-dataset and cross-sensor transfer. Pang et al. (2023) used an extended U-Net, UNmask, across visible to shortwave-infrared bands, improving cloud detection over bright surfaces, and Dong et al. (2024) used change-guided, multiscale DL toward the same end.

Despite major progress, DL is still constrained by its data demand, computational cost, and model stability. With limited data, deep networks can easily overfit; with many parameters, they are difficult to tune. These constraints matter for GOSAT-2 CAI-2, which needs a stable, transparent operational rule. GOSAT-2 documents describe two CAI-2 cloud-discrimination algorithms. The first, Cloud and Aerosol Unbiased Decision Intellectual Algorithm 1 (CLAUDIA1), uses conventional threshold tests. The second, CLAUDIA3, was developed using ML; it replaces sequential tests with an SVM that automatically estimates an optimized clear–cloud boundary (Ishida et al., 2018). This approach carries two limits. One is that, because it is trained on MODIS, it depends on external training data. The other is that it requires different support vectors, and without this adjustment, it may over-detect clouds in some conditions and miss them in others (Oishi et al., 2018). Beyond the algorithm, CAI-2 lacks thermal-infrared bands, weakening sensitivity to cirrus, thin clouds, and small ocean clouds. FTS-2 adds thermal-infrared and 2 μm water-vapor information, but coverage gaps and difficult surfaces still leave missed clouds and false alarms. All this matters because one cloudy CAI-2 pixel can invalidate an FTS-2 footprint for XCO_2 retrieval. The main challenges are single-pixel sensitivity, thin-cloud omission, and uneven error patterns, and they overlap. The targeted reduction of CAI-2 cloud-omission error, using only the sensor's own bands and no external reference data at the application stage, has therefore remained unresolved. The structure of the problem points away from building another general classifier and toward targeted spectral tests for the known error regimes of the current mask.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

This study develops a set of targeted spectral tests to reduce cloud omission in the GOSAT-2 CAI-2 cloud mask over water, where missed clouds are the dominant error. We propose five spectral feature spaces from band ratios, using only CAI-2's L1B band radiances (ultraviolet 0.380 µm, visible 0.674 µm, and near-infrared 0.869 µm) and no external reference data. Correctly flagged (true-positive) and missed (false-negative) cloud footprint samples from L2 are collected across three granules and mapped into these spaces, where a polynomial logistic regression estimates the probability that a footprint is a true positive. The aim is to recover the false negatives that are at least as cloud-like as the weakest detected true positive; anchoring the decision boundary to that weakest cloud keeps the criterion conservative and focused on omission. The Kolmogorov–Smirnov statistic identifies which feature spaces best separate recoverable false negatives, and the three that remain stable across all granules are retained and used to define feature-space probability thresholds for omission recovery. The thresholds, trained on those three granules, are tested on all four, including an independent fourth. The goal of this study is to detect clouds missed by CAI-2 Level 2 (L2) and compare the result against the original CAI-2 L2 detection.

The structure of this paper is as follows. Section 2 details the GOSAT-2 CAI-2 L1B and L2 samples, the construction of spectral feature spaces from L1B radiances, and the polynomial-logistic framework used to estimate cloud probability. It also explains the separability test used to retain stable feature spaces and accuracy assessment methods. Section 3 presents the recovery results and evaluates the revised mask against the original CAI-2 L2 detection. Section 4 evaluates the results, interprets spectral patterns associated with L2 cloud detection and omission in the retained feature spaces, and presents the limitations and conclusions.

2. Materials and Methods

2.1. Study Area and Datasets

The study area covers four CAI-2 granules over the ocean around Japan, spanning roughly 28°N–41°N and 130°E–150°E (Figure 1). The granules were selected to span different latitudes, acquisition dates, cloud fields, and sun-glint geometries over the ocean, while keeping the surface type and viewing direction fixed. This way, recovery could be tested across varied conditions rather than a single scene. Specifically, the granules were acquired on three dates across two seasons. Two share the same date, 29 August 2023, but differ in location: Granule 1 lies off the Pacific coast of central-eastern Honshu, and Granule 2 to its south. Granule 3 (16 November 2023) lies on the Sea of Japan side, and Granule 4 (28 December 2023) to the southwest. All scenes were taken in the forward-looking (FWD) view, and the analysis is restricted to water pixels, where cloud omission is the dominant error and surface effects such as sun-glint complicate discrimination.

Two GOSAT-2 products were used, both obtained from the NIES GOSAT-2 data portal (https://prdct.gosat-2.nies.go.jp/index.html.en). The first is the TANSO-CAI-2 Level 1B radiance product (V03.13), which provides calibrated, geolocated spectral radiances in the CAI-2 ultraviolet (UV), visible (VIS), and near-infrared (NIR) bands; these radiances are the sole input to the proposed spectral tests. The second is the TANSO-CAI-2 Level 2 cloud-discrimination product (V01.05), which provides the per-pixel cloud flag used in this study.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

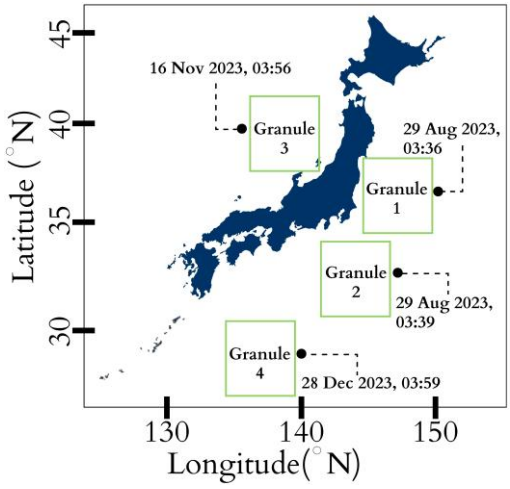


Figure 1: Study area and selected GOSAT-2 CAI-2 granules over Japan.

2.2 Methodology

2.2.1 Overview

The proposed method recovers clouds missed by the operational CAI-2 L2 mask over water, using only CAI-2's own radiances. Figure 2 summarizes the workflow. Each step name in parentheses below is a box in that figure. The CAI-2 L1B radiance bands and the L2 cloud-discrimination layer were first converted in MATLAB into rasters on a common WGS84 grid for subsequent analysis in the ArcGIS Pro. The L2 cloud-confidence classes were binarized into a mask (hereafter, "L2 cloud mask") using the operational integrated clear-sky confidence level (CCL) threshold of 0.33. The 0.33 is the operational threshold specified in the CAI-2 L2 / GOSAT-2 cloud-discrimination product documentation (Oishi et al., 2020). True-positive (TP) and false-negative (FN) cloud footprints were then identified by visual inspection in ArcGIS Pro (Sample collection). Each cloud footprint was converted to a geodesic buffer, and its UV, VIS, and NIR signals were summarized as the 95th-percentile radiance to capture the bright, cloud-sensitive response. These samples were mapped into five candidate spectral feature spaces built from band ratios (Initial candidate features). In each space, a polynomial logistic regression assigned every footprint a probability of being a true positive, and the decision boundary was anchored to the weakest detected true positive at this same step, before the separability test. The Kolmogorov–Smirnov statistic (Tang et al., 2011) then measured how well each space separated recoverable misses, reducing the five candidates to the three most stable tests.

The logistic scores from the retained feature-space models were evaluated over cloud-free ocean reference pixels, including dark water and bright glint-affected water. The resulting shift in predicted cloud probability between strong- and weak-glint subsets was then used to define the glint-adjusted boundary (Cone angle-dependent glint effect extraction). These subsets were separated using a cone-angle layer, which is defined as the angular separation between the satellite viewing direction and the specular reflection direction of sunlight over the ocean. This layer was calculated from the L1B FWD geometry fields by

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

combining the solar and satellite zenith angles and the solar and satellite azimuth angles using the CAI-2 cone-angle equation (Oishi et al., 2020).

In parallel, a sigmoidal "large" fuzzy membership function in the VIS band assigned each pixel a membership value, defining a compact range for glint-contaminated pixels (Fuzzy-based glint filtering). Pixels at or below the maximum of this range were removed as glint. A minimum radiance was also estimated for the NIR and VIS bands over ocean pixels. For each band, this limit was derived from the mean and standard deviation of the darkest ocean subset.

In the final application, the three tests were run only on pixels left labeled no-cloud by L2 and above these radiance limits, with the decision boundary shifted by cone angle under glint-prone geometry (Logit boundary to threshold test and Cloud omission detection). In the strongest glint, detections were kept only where the fuzzy VIS membership supported a cloud-like signal. Recovered clouds were merged with the L2 mask, and the optimized mask was evaluated against L2 using visually verified points, Cohen's kappa, recovery-density maps, and variograms (Performance and error pattern assessment).

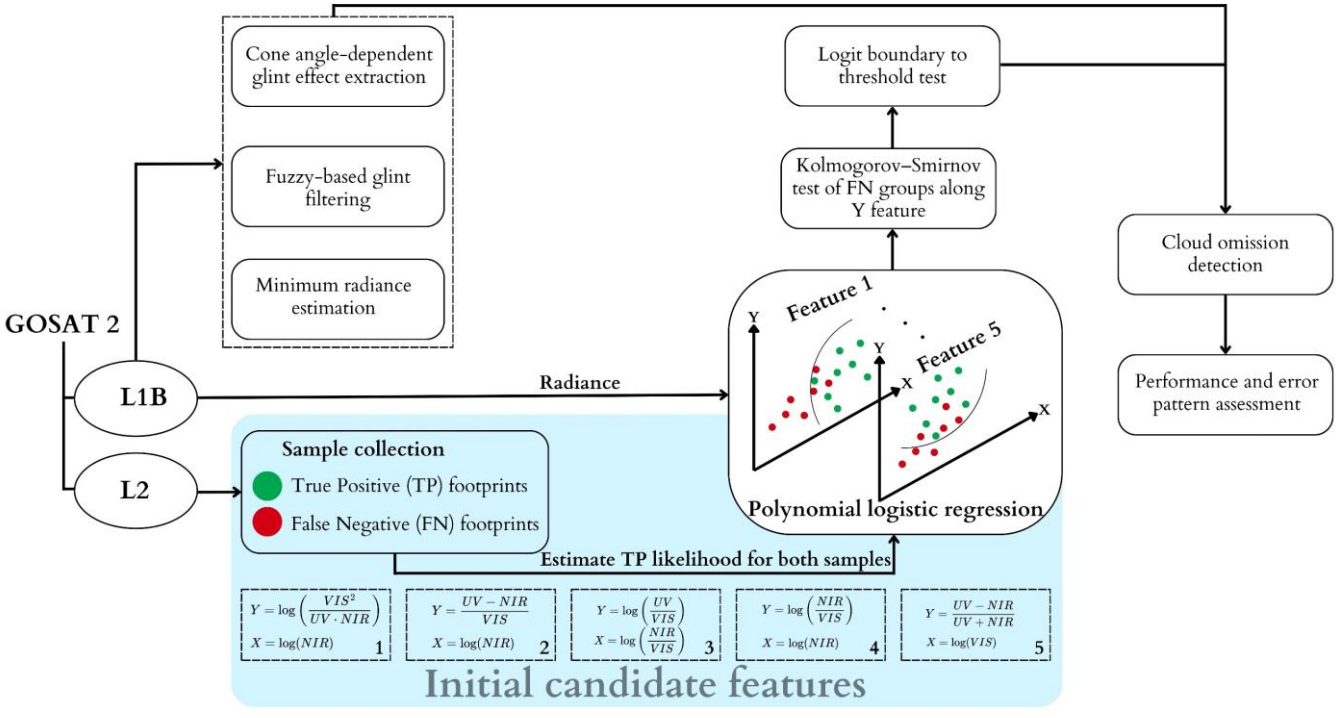


Figure 2: Workflow for targeted spectral threshold tests to reduce cloud omission over water in the GOSAT-2 CAI-2 mask.

2.2.2 Spectral Features

Addressing the omission of thin and scattered clouds first requires understanding its underlying cause. Many cloud-detection methods rely on several spectral tests, each based on a single scalar signal such as a band reflectance or a band ratio. When that signal is weak, and especially in scenes where sunglint must be controlled, a single-band or ratio threshold sits on a fine

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

line between cloud and clear; it cannot isolate thin and scattered cloud without raising large overestimations. Cloud edges pose the same problem. For GOSAT-2 over water, the CAI-2 L2 threshold tests use NIR reflectance, the NIR/VIS reflectance ratio, and the normalized difference vegetation index. To overcome this, a bivariate spectral test is proposed. The aim is to build a space in which these subtle differences become separable. Indeed, each bivariate spectral test is defined by a polynomial logistic regression in a 2D spectral feature space (X, Y, and X^2) and is applied to each pixel as a threshold on the resulting cloud probability. In principle, this added dimension modulates the infiltration of weak signals of thin and scattered clouds into the decision. The feature spaces were constructed from physically interpretable combinations of the available CAI-2 UV, VIS, and NIR radiances. This design yielded five candidate feature spaces, shown in Figure 2 and described below. In each, the bands are combined as ratios or differences of spectral radiances, and the natural logarithm is applied to non-linearly expand the dynamic range of the low, closely spaced radiances of thin cloud and clear ocean.

Feature 1 uses NIR brightness and VIS-weighted three-band contrast to test whether thin cloud signals exhibit a stronger VIS response than UV and NIR. Feature 2 compares the UV–NIR difference after scaling by VIS to test whether thin cloud signals show a distinct UV-to-NIR contrast. Feature 3 uses VIS as a common reference band, comparing UV/VIS and NIR/VIS to test whether thin cloud signals are better described by relative spectral contrast than by absolute radiance. Feature 4 combines NIR brightness with NIR/VIS contrast to test whether thin cloud signals are better identified when the absolute NIR response is paired with its contrast against VIS. Feature 5 uses VIS brightness and the normalized UV–NIR difference to test whether thin cloud signals can be described by visible signal strength and UV–NIR spectral balance.

2.2.3 Threshold Derivation

Satellite data classification can be treated as a probabilistic mapping from measured radiance to an assigned class. In this framework, the behavior of a cloud mask is analyzed by relating the radiances a sensor records to the decisions it produces. Logistic regression (Sun et al., 2020) is a reliable and efficient statistical strategy for this task. It estimates the relationship between a binary dependent variable and a set of predictors, and its simplicity keeps it stable on small datasets. The fitted model yields a logit, which can serve as the basis for defining threshold tests for targeting cloud-specific spectral signatures.

For this purpose, the L2 cloud mask is first assessed by visual inspection, and footprints are labeled as TP or FN. The center coordinate of each CAI-2 footprint is extracted, and the L1B radiances at those coordinates are retrieved. Each labeled footprint is projected into the five feature spaces defined in the previous section. A logistic model fitted to the TP and FN samples assigns every footprint a predicted TP probability. Processing is done in ArcGIS Pro with Python, with the L1B and L2 layers already imported. Mapping these probabilities back onto the feature spaces exposes the overlap between the two samples. The decision boundary is therefore tied to the weakest detected TP, the smallest predicted probability among correctly identified clouds. Footprints at least as cloud-like as this threshold are marked for recovery to cloud class. Because the missed footprints concentrate in optically thin, small scattered, and cloud-edge cases, the regression isolates the spectral signature that separates these recoverable clouds from clear water. Failure-prone L2 footprints are thus characterized through their L1B spectral-ratio signatures.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

The polynomial logistic regression (Bishop, 2006) is written as:

$$\eta_i = \beta_0 + \beta_1 X_i + \beta_2 Y_i + \beta_3 X_i^2, \quad (1)$$

Here η_i is the linear predictor for footprint i, and X_i and Y_i are the paired band-radiance features used to define each spectral test space. Also, β_0 is the intercept and β_i is the fitted regression coefficients. The quadratic term X_i^2 admits a curved decision boundary. The predicted TP probability follows from the logistic (Lee et al., 2017; Sun et al., 2020) link:

$$p_i = P(y_i = 1 \mid X_i, Y_i) = \frac{1}{1+\exp(-\eta_i)}, \quad (2)$$

Where $y_i \in \{0,1\}$ is the label, with $y_i = 1$ for a TP and $y_i = 0$ for an FN. Equivalently, in logit form:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i + \beta_2 Y_i + \beta_3 X_i^2, \quad (3)$$

p_i is the predicted probability that footprint i is a TP, with $p_i = P(y_i = 1 \mid X_i, Y_i)$.

Because some missed footprints carry probabilities comparable to detected clouds while others do not, the FN footprints cannot be treated as a single group. Their separability is therefore tested. The FN footprints are divided into two groups by the fitted model, and the distribution of their Y feature values within each group is summarized by the empirical cumulative distribution function:

$$\hat{F}_n(y) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}(y_k \le y), \quad (4)$$

where n is the group size and $\mathbf{1}(\cdot)$ is the indicator function, so that $\hat{F}_n(y)$ gives the fraction of footprints with a value at or below y. The separation between the two groups is measured by the two-sample Kolmogorov–Smirnov (KS) statistic, the largest vertical gap between their distributions:

$$D = \sup_y \left|\hat{F}_1(y) - \hat{F}_2(y)\right|, \quad (5)$$

with significance assessed from the asymptotic Kolmogorov distribution (Stephens, 1970). A large D indicates that the two groups occupy distinct ranges of Y, so the feature space draws a sharp line between recoverable and remaining FN footprints. The spaces with consistently large D are retained. In each retained space, the fitted logit η_i becomes the test variable. two cutoffs on η_i define a ramped Clear Confidence Level, which falls from 1 over clear pixels to 0 over cloud (Fig. 3).

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

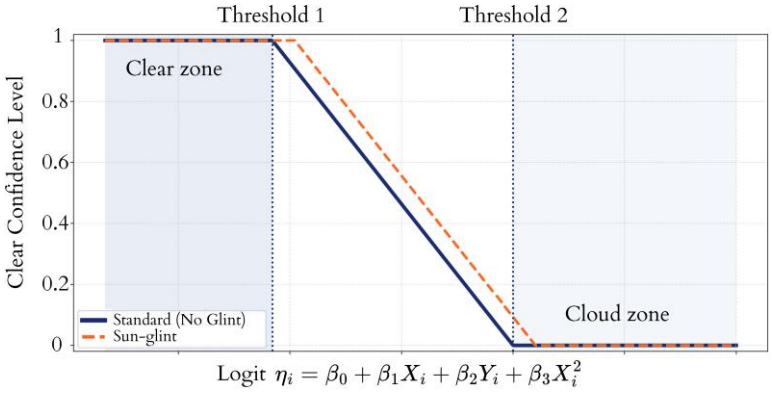


Figure 3: Clear confidence level as a function of the logit derived from the feature values.

Threshold 1 is set to the data-driven cutoff η_{opt} at the weakest detected TP. Threshold 2 is set by the desired cloud probability, which the logit link converts into a matching cutoff on η. A probability of 0.5 corresponds to $\eta_i = 0$, so, threshold 2 is set at $\eta_i = 0$ and requiring a higher probability, such as 0.7, shifts the cutoff to a correspondingly higher value of η. The threshold can therefore be tuned to the confidence level required. The logit η_i is computed for each pixel from the spectral features X, Y, and X^2 through the fitted coefficients, so the thresholds act on the spectral model rather than on the radiances directly. Likewise, the threshold tests are derived by relating CAI-2 L1B radiances to the L2 cloud-mask outcomes at the footprint scale.

2.2.4 Performance Assessment

Performance is measured against visual-inspection reference labels. For each granule, 500 points are distributed across the scene and classified as cloud or clear by visual inspection of the L1B imagery. These points form the truth set for the point-based metrics, comparing both the original CAI-2 L2 mask and the optimized mask against the same labels. In visual inspection, Cohen's kappa is calculated. Cohen's kappa κ corrects for agreement expected by chance, providing a stricter measure of classification agreement than accuracy alone.

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{6}$$

where p_o is the observed proportion of agreement and p_e i is the proportion of agreement expected by chance from the marginal totals of the confusion matrix. Two error types are then tracked separately. Overlook is the fraction of true clouds the mask misses; overestimation is the fraction of clear points it flags as cloud. They were reported together because recovering missed clouds is expected to increase overestimation, so reporting both shows the size of this trade-off.

Also, spatial patterns are examined across the entire raster by mapping the recovered missed clouds and residual false positives as separate pixel sets. Each set is smoothed with a Gaussian kernel to form a continuous density field:

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

$$\lambda(\mathbf{x}) = \sum_k G_\sigma\,(\mathbf{x} - \mathbf{x}_k), \tag{7}$$

where $\mathbf{x}_k$ are the pixel locations of the selected set and G_σ is a Gaussian of width σ. The resulting maps show whether recovery and overestimation are diffuse across the scene or concentrated in particular regions (Section 3.6, Figure 9).

Moreover, an indicator semi variogram characterizes the spatial structure of the masks. For a binary class field, this measures how often two pixels separated by a lag distance h fall in different classes:

$$\gamma_I(h) = \frac{1}{2N(h)} \sum_{(i,j)\,\in\,N(h)} \mathbf{1}\left(z_i \neq z_j\right), \tag{8}$$

where z_i and z_j are the class labels of a pixel pair, $\mathbf{1}(\cdot)$ is the indicator function, and $N(h)$ is the set of pairs separated by lag h. Pairs are drawn from stratified anchor pixels spread across the scene, so the estimate is not dominated by any one region. Low semi-variance at short lags reflects spatially coherent cloud, while higher values indicate frequent class changes over short distances. Comparing the original and optimized masks against the reference semi variogram shows whether the optimized mask reproduces the spatial texture of the truth at the scales of scattered cloud (small h) and cloud edges (intermediate h).

3. Results

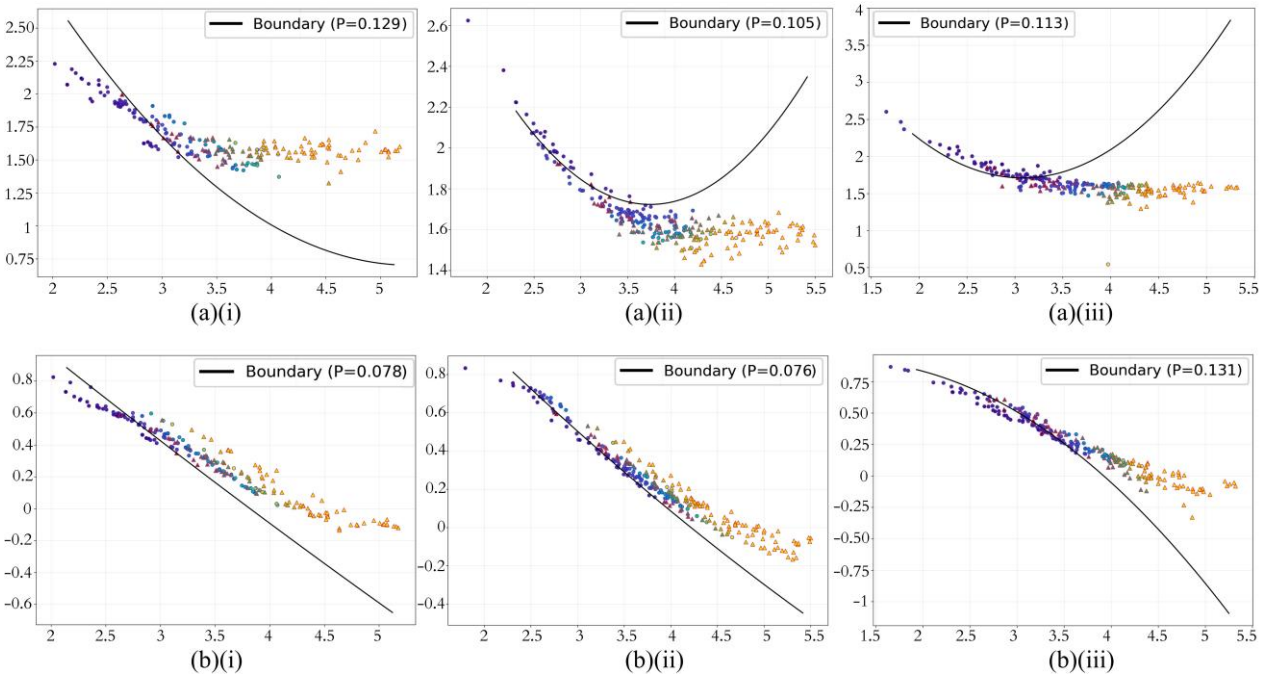
3.1 Mapping L2 Detection Outcomes in L1B-Derived Feature Spaces

For each granule, approximately 130 FN and 130 TP footprints were manually extracted, and their spectral patterns were then analyzed in five two-dimensional feature spaces. Figure 4 presents the five candidate feature spaces (rows a–e), each fitted independently across the three training granules (columns i–iii). In each panel, footprints are plotted according to their feature values and coloured by the TP probability assigned by the fitted model (Eq. 2). Triangles represent footprints detected by the L2 mask (TP), whereas circles represent missed footprints (FN). The curve also shows the fitted boundary defined by the weakest detected TP footprint in each panel. The decision boundary is also a fixed-probability contour of this surface. The panels share a common structure. In each feature space, the footprints form an elongated band along which the fitted TP probability generally changes steadily. So that one end is dominated by high TP probability detections, most of which are observed TPs, while the other contains mainly low-probability footprints, most of which are observed FNs. But the transition between them is gradual rather than a sharp separation. This gradient is the signal that provides the basis for the threshold tests in Section 3.2, as a footprint's position in the feature space reflects its degree of cloud-like behaviour. At any spectral position, it returns the detectability of a cloud there, calibrated against the footprints the operational mask has already classified.

Two aspects of this pattern are worth noting. First, the FN footprints are not confined to the low-probability end. In several feature spaces (e.g., Figure 4c(i) and 4c(ii)), they extend well into the region occupied by detected TPs, meaning that some overlooked footprints have TP probabilities as high as those assigned to footprints flagged by the mask. These are the recoverable cases, because they are spectrally similar to the detected ones, yet were missed in the mask. This is because, in the single-band or single-ratio tests, their weak radiance contrasts remained close to the no-cloud background; only the combined

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

two-dimensional spectral pattern separated them. Second, for a given granule, the same footprints can form a nonlinear distribution in one feature space (Figure 4a(i)) and an almost linear and compact distribution in another (Figure 4e(i)). So, the feature space's ability to separate footprints is not the same when the underlying differences are subtle. The next section first
280 examines these differences in separation to select feature spaces for targeting recoverable FNs.



https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

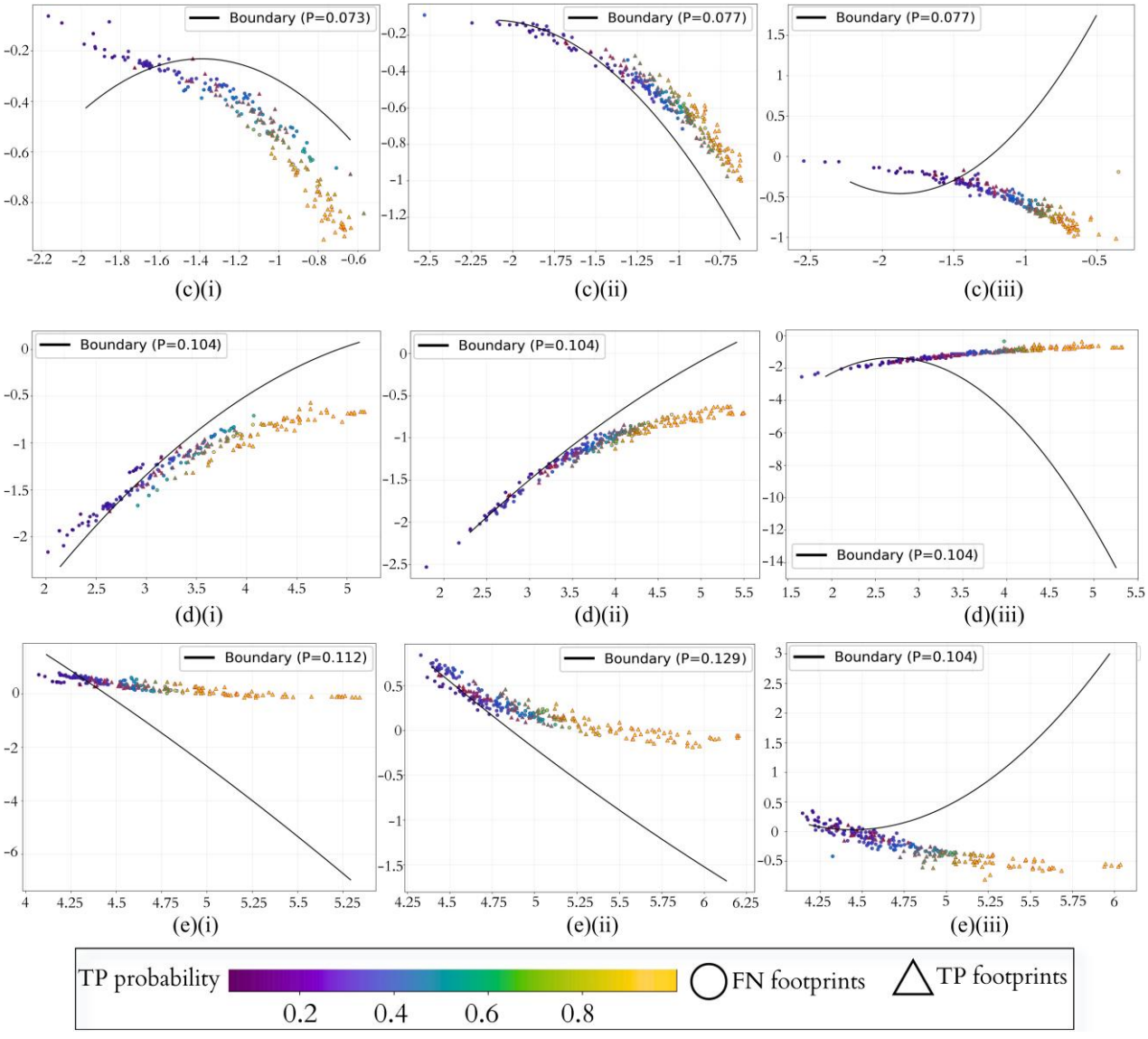


Figure 4: Polynomial logistic fits (black lines) for the five candidate feature spaces across the three training granules, where columns i–iii correspond to granules 1–3, and rows a–e correspond to feature spaces 1–5 defined in Section [methods]. In each panel, the horizontal and vertical axes are the two spectral features (X and Y) of the corresponding feature space, and footprints are colored by their predicted TP probability (color bar).

3.2 Kolmogorov–Smirnov Separation of Overlooked-Footprint Groups

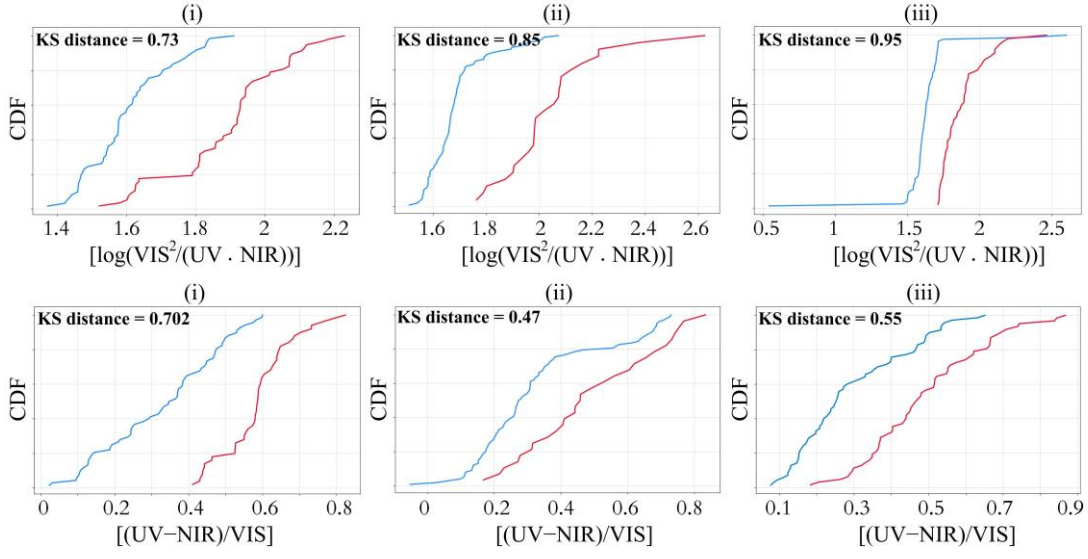
This section evaluates the five candidate feature spaces by how well each separates two groups of FN footprints. The first group contains FNs whose TP probability is at or above the weakest detected TP level; these are treated as potentially recoverable. The second contains FNs below that level. The most suitable feature spaces are those that separate the two groups

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

most clearly. A clearer separation means the feature space does not treat all FN footprints alike but responds to differences in their signals.

Within each feature space, (X) helps set the position and curvature of the fitted boundary, while (Y) provides the common axis for comparing the potentially recoverable and remaining FN groups. The KS test compares their empirical cumulative distributions along (Y), and the KS distance records the largest vertical gap between the two curves. A larger distance indicates less overlap and stronger separation between the groups.

Figure 5 compares the five feature spaces across the three granules. In each panel, the blue curve shows the cumulative distribution function (CDF) of FNs above the weakest detected TP level, while the red curve shows the CDF of those below it. The corresponding KS distance is reported in each panel. The results show clear differences among the feature spaces. Features 1, 3, and 4 produced consistently high KS distances across all three granules. For log ($VIS^2/(UV \cdot NIR)$), the KS distances were 0.73, 0.85, and 0.95 for Granules 1, 2, and 3, respectively. The log (UV/VIS) gave complete separation in Granule 1 (D = 1.00), followed by 0.85 and 0.89 in Granules 2 and 3. The log (NIR/VIS) feature also performed consistently well, with KS distances of 0.70, 0.76, and 0.93. A feature space that separates the groups well in one granule but poorly in another is unlikely to generalize across unseen scenes. Features 1, 3, and 4 maintained clear separation in all three granules and were therefore carried forward to the threshold tests, whereas Features 2 and 5 were excluded because their separation was weaker.



https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

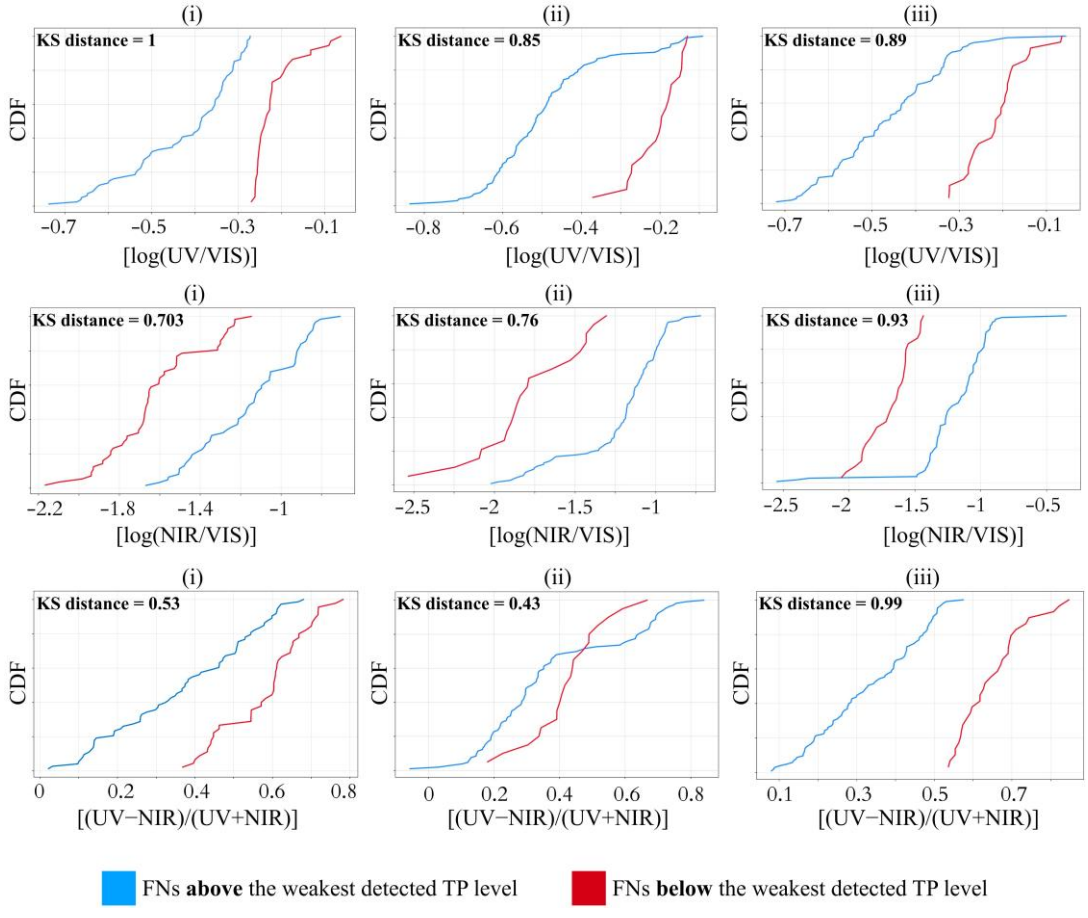


Figure 5: Empirical CDFs of the two FN groups for the five candidate feature spaces, with columns i–iii corresponding to granules 1–3.

3.3 Determination of Spectral Thresholds

After this selection step, the threshold tests are applied in the full two-dimensional feature space. For each retained feature space (Features 1, 3, and 4), the fitted polynomial logistic model defines the decision boundary over both X and Y. As described in Section 2.2.3, Threshold 1 is anchored to the TP probability of the weakest detected TP, while Threshold 2 can be set to any chosen probability. For example, if the weakest real TP has a probability of about 0.1, Threshold 1 falls at a logit of roughly −2; Threshold 2 might then be set at a probability of 0.5, which corresponds to a logit of 0. Table 1 lists the fitted coefficients and the two thresholds for the three retained feature spaces across the three granules. The coefficients β_0 to β_3 set the shape and position of each boundary, while Threshold 1 and Threshold 2 mark the limits of the CCL ramp described in Section 2.2.3 (Figure 3).

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1. Fitted coefficients and the two thresholds for the three retained feature spaces (F1, F3, F4) across granules 1–3.

Feature	Granule	β_0	β_1	β_2	β_3	Threshold 1	Threshold 2
F1	G1	-42.3974	13.5495	6.79325	-1.285	-1.913	0.847
	G2	76.8316	-27.1919	-16.3071	3.6336	-2.145	0.405
	G3	25.3133	-12.705	-4.56558	2.0621	-2.060	-0.847
F3	G1	-21.4324	-22.4179	-14.2551	-8.064	-2.544	0.847
	G2	34.3404	32.9257	14.4474	7.72088	-2.477	0.405
	G3	6.43871	10.7133	-2.4275	2.85927	-2.477	-0.847
F4	G1	-58.074	18.9414	-9.65214	-1.5385	-2.157	0.847
	G2	-47.6255	12.4156	-9.69243	-0.6971	-2.129	0.405
	G3	9.02515	-7.68806	0.720387	1.42656	-2.157	-0.847

It is worth noting that different logit coefficients do not imply different boundary behavior. The coefficients vary because the fitted boundary takes a different geometric form in each granule, adapting its curvature and orientation to the local point cloud. The signs and magnitudes, therefore, differ between fits, yet each boundary orders the footprints in the same direction of increasing cloud probability, so the separation it produces remains consistent.

3.4 Spectral-Threshold Test Outcomes

For each granule, the three threshold tests applied were those derived from that same granule's labels. The tests built from the granule 1 logits were run on granule 1, the granule 2 tests on granule 2, and the granule 3 tests on granule 3, so that each granule was evaluated with the three tests trained on its own footprints. Although the logit coefficients were fitted at the footprint scale, detection was carried out pixel by pixel. For every pixel, the L1B and L2 products were compared and a pixel not flagged as cloud in L2, and whose radiance exceeded the minimum-radiance threshold, was passed to the three tests for evaluation. Their outputs were combined so that a pixel was classified as newly cloudy if any one of the three tests was positive. The resulting detections are shown in Figure 6, where blue represents the original GOSAT-2 mask, and yellow represents new threshold-test detections.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

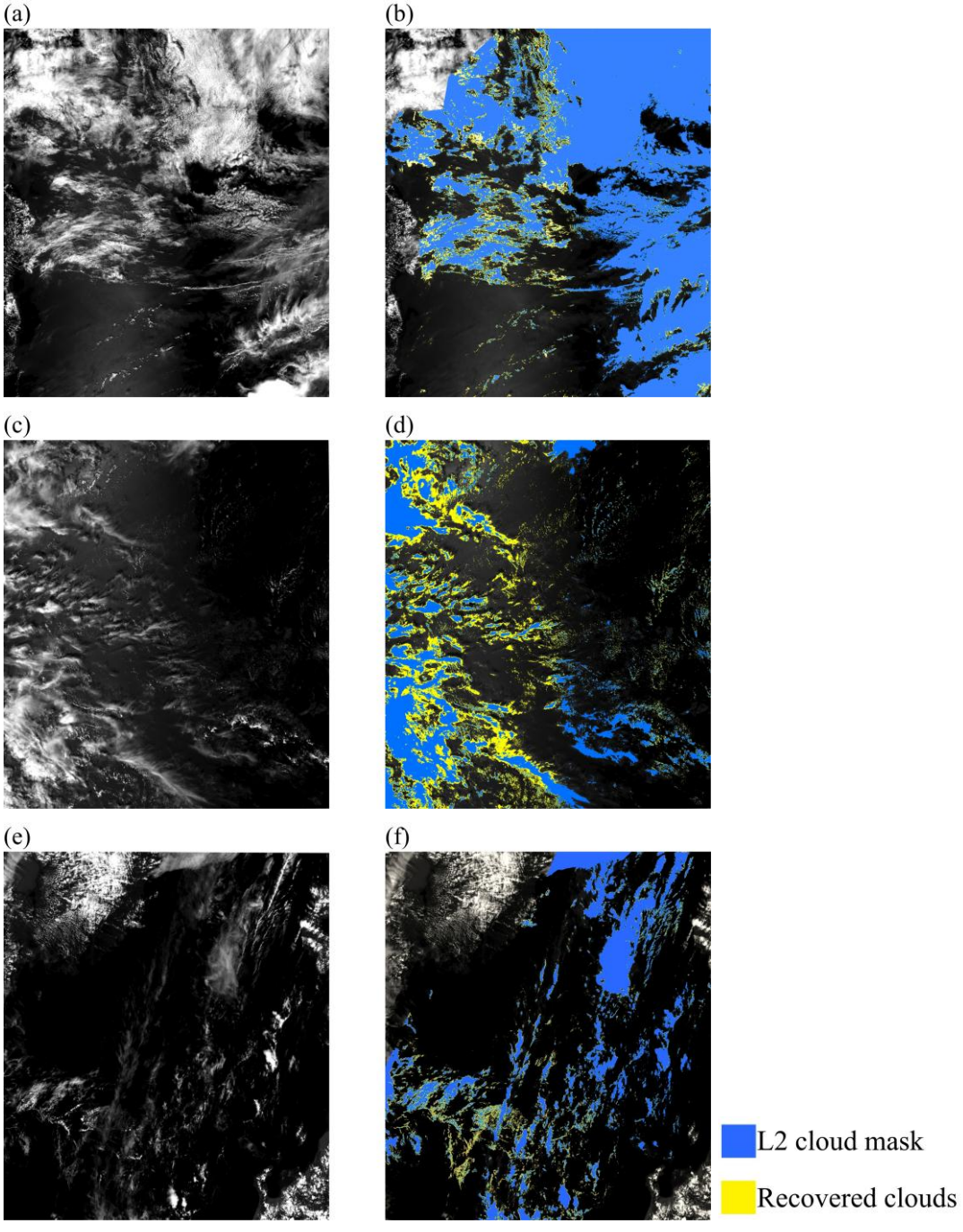
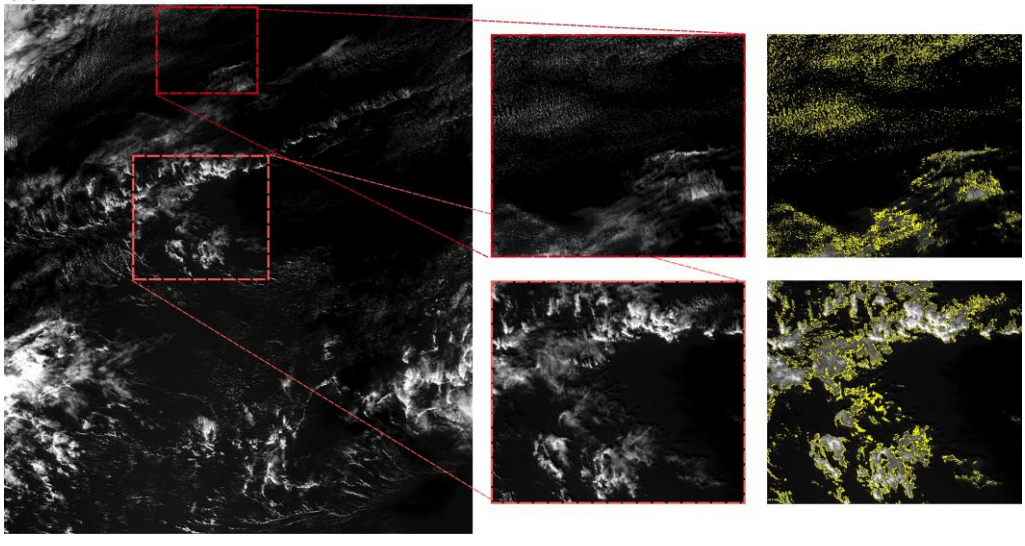


Figure 6: Cloud recovery on the three granules used for fitting. Left column (a, c, e): L1B scenes for granules 1–3. Right column (b, d, f): the corresponding detection results, with the original L2 cloud mask in blue and the clouds recovered by the combined application of the three retained threshold tests (F1, F3, F4) in yellow.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Small, scattered clouds, optically thin clouds, and cloud edges were recovered. These are the cases that dominate L2 omissions. The recovered cloud pixels represented 8.1%, 40.0%, and 19.0% of the original L2 cloud-pixel counts in Granules 1–3, respectively. The variation mainly reflects differences in cloud-field characteristics, as scenes containing more thin, fragmented, and edge clouds provide more opportunities for additional detection. This spatial pattern is visible in all three scenes, showing that the recovery is structured rather than random. Cloud edges were recovered more often than the other cloud types, most likely because they have stronger spectral contrast. Small scattered clouds were also detected well. Recovery of thin clouds varied between scenes, largely because the tests were deliberately conservative at small cone angles, where sunglint is strongest. The three granules differ in acquisition and in the difficulty they present. Granules 1 and 2 were acquired on the same date but at different times and locations, whereas granule 3 was acquired on a different date, time, and location. These differences are reflected in the original L2 performance, so its quality varies from one granule to another. Because the tests are applied only to pixels not already flagged as cloud, the number of candidate pixels varies across granules. Granule 2, for example, showed a wide range of new detections spanning scattered, thin, and edge clouds, because more pixels were available for testing, and sunglint affected a smaller part of the study area. Overall, the recovered clouds are spatially consistent with the existing cloud field, extending the same systems identified by the L2 mask.

(a)



https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

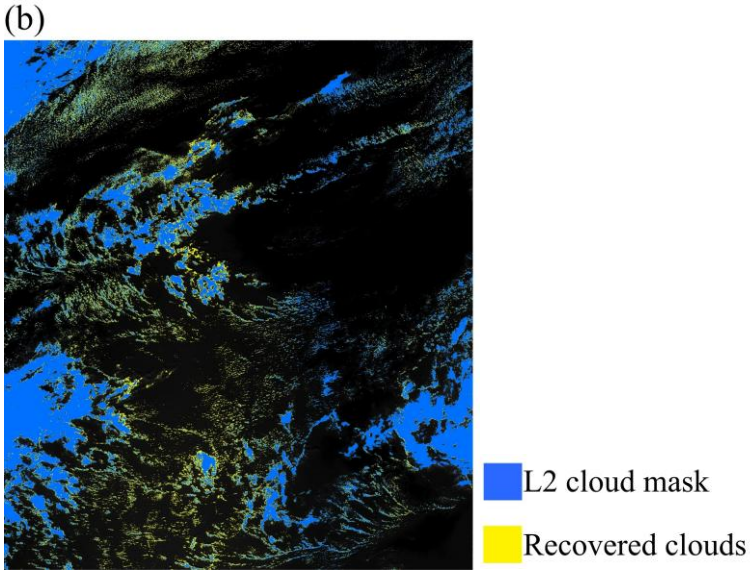


Figure 7: Cloud recovery on the independent scene (granule 4). (a) L1B grayscale scene, with two regions outlined and enlarged to show the recovered clouds in detail. (b) Detection result over the full scene, with the original L2 cloud mask in blue and the clouds recovered by the threshold tests in yellow.

The tests were then applied to an independent scene, granule 4, whose footprints were not used to derive any of the logits. For this scene, each retained spectral test had three fitted versions, one from each training granule. In practice, all nine logits were evaluated. A pixel passed a test when at least two of its three versions flagged it as cloud and was retained as cloud if any one of the three tests passed. Thus, two positive logits within the same spectral test were sufficient to classify a pixel as cloud. This result is shown in Figure 7, together with the original L2 mask, the new detections from the tests, and two zoomed insets highlighting the recovered clouds. The independent Granule 4 supports the results beyond the training set. Specifically, recovery in this independent scene suggests the method can generalize beyond the training granules.

3.5 Accuracy Metrics Across All Granules

Figure 8 compares the original CAI-2 L2 mask with the optimized mask produced by the new spectral thresholds across the four granules. Performance was evaluated using the overlook rate, overestimation rate, and Cohen's kappa against the visually verified reference mask. A total of 1,000 random-point evaluations were conducted per granule, with 500 for each of the original CAI-2 L2 and optimized masks, giving 4,000 evaluations overall. Based on the results, the optimized mask sharply reduced cloud omission while increasing the overestimation rate by a smaller margin. In Granule 1, the overlook rate decreased by 7 percentage points (9% to 2%), while overestimation increased by about 4 points. The trade-off was more favourable in the other granules. Overlook decreased by 20 points in Granule 2 (24% to 4%), 22 points in Granule 3 (32% to 10%), and 18 points in Granule 4 (28% to 10%). The corresponding increases in overestimation were 3 points in Granule 2 (1% to 4%), no change in Granule 3, and 2 points in Granule 4 (1% to 3%). Cohen's kappa increased in all four granules, showing that the reduction in missed clouds outweighed the additional false detections. After optimization, kappa exceeded 0.90 in the first

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

three granules. It increased by 5 percentage points in Granule 1 (0.88 to 0.93), 11 in Granule 2 (0.81 to 0.92), 14 in Granule 3 (0.77 to 0.91), and 8 in Granule 4 (0.80 to 0.88). The largest improvements occurred in the granules with the highest original omission rates, suggesting that the proposed threshold tests mainly corrected scenes where the L2 mask missed the most cloud.

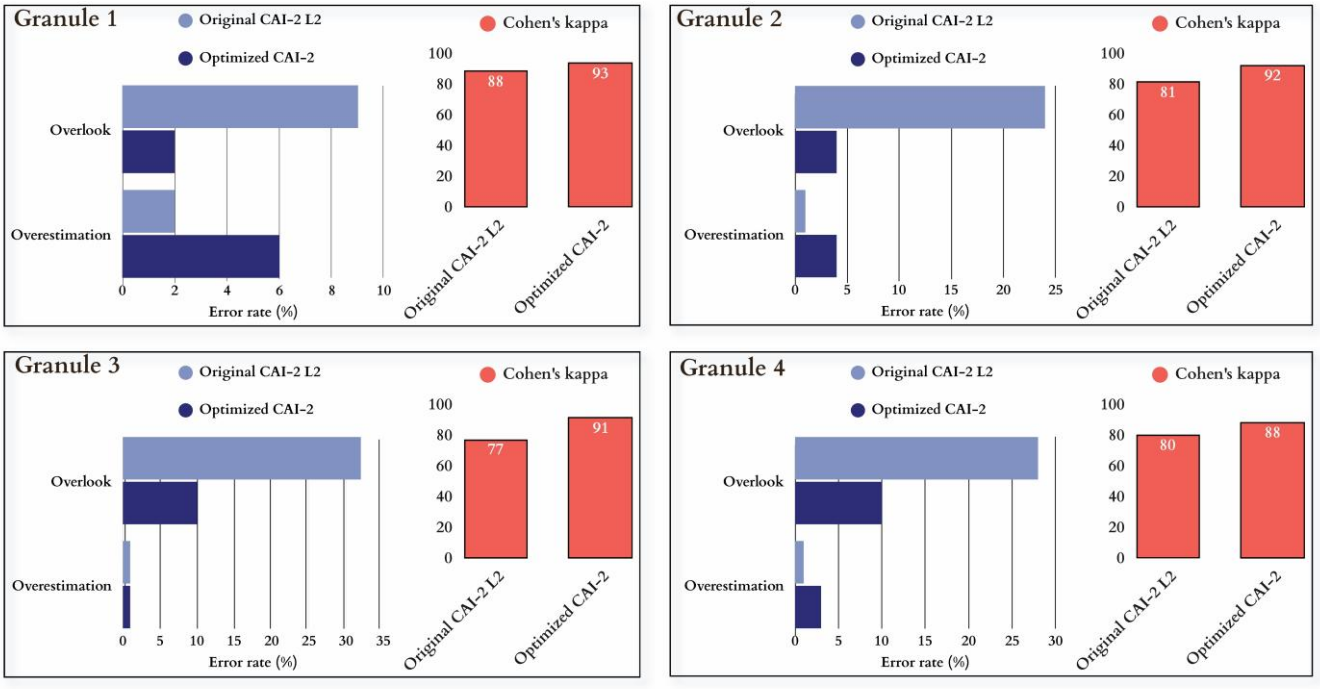


Figure 8: Accuracy of the original and optimized CAI-2 masks per granule. Within each panel, overlook and overestimation error rates (%) are shown on the left, and Cohen's kappa against the visually verified reference is shown on the right.

3.6 Spatial Error Assessment for the Independent Granule

For Granule 4, a binary evaluation reference mask was prepared for spatial comparison using supervised classification, followed by detailed visual inspection of the entire scene and manual correction of misclassified pixels. To locate where overlook is reduced and where overestimation rises in the optimized mask, the two Gaussian-smoothed density maps in Figure 9 show both effects across Granule 4. The optimized tests recover 55.4% of the clouds missed by the L2 mask, while introducing false positives over only 2.07% of truly clear pixels, giving a recovery-to-cost ratio of about 27 to 1. Recovery and overestimation also show different spatial patterns. Recovery in panel (a) occurs throughout the scene but is strongest in the south and center, where the 95th-percentile contours form broad, connected areas with local recovery fractions above 0.5. This concentration mainly reflects the large number of recovered cloud-edge pixels. In the eastern part of the scene, recovery is less dense and is associated more with small and optically thin clouds. By contrast, the false positives in panel (b) are concentrated in narrower patches in the north, mainly between 23° and 25°N. Their local fractions generally remain below 0.20, suggesting

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

that overestimation is largely confined to low-cone-angle or cloud-clear mixed-pixel areas, where cloud–clear distinction is difficult, rather than spread across the scene.

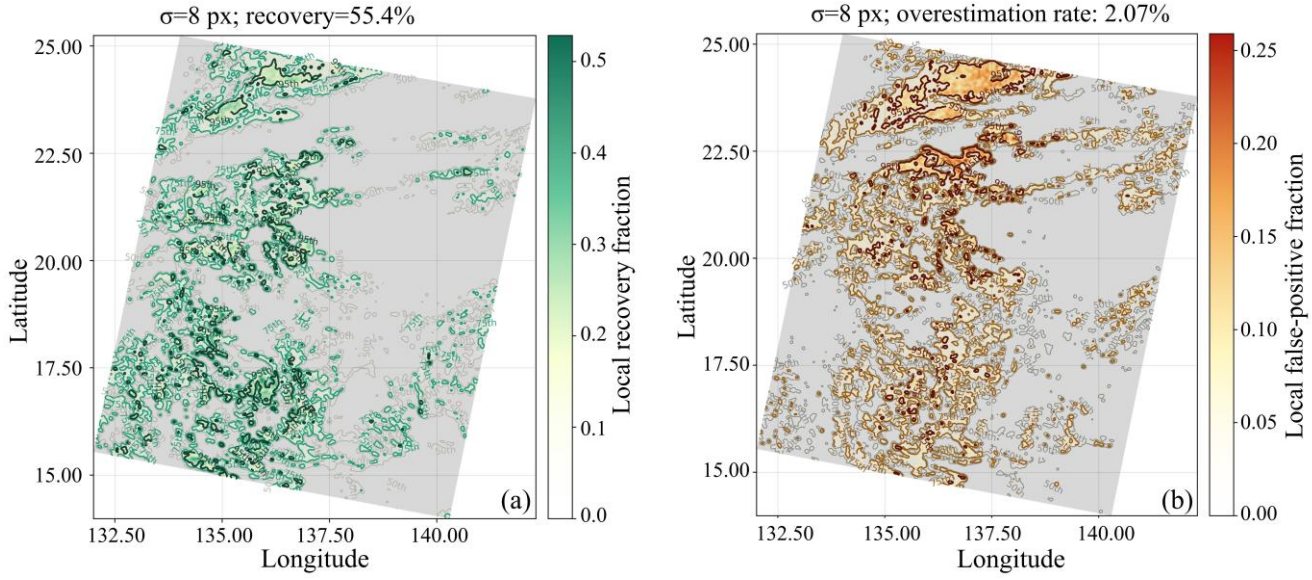


Figure 9: Spatial distribution of recovered clouds in panel (a) and false positives in panel (b), expressed as local fractions across the scene in granule 4.

To examine which cloud patterns show improvement, an indicator semivariogram was used (Figure 10). The estimate was built from 10,000 stratified anchor pixels spread evenly across the scene in 25 × 25 sampling cells. Each anchor was paired with 50 neighboring pixels at distances of up to 200 pixels (≈ 77 km), and the resulting pairs were grouped into 40 lag intervals. At each lag, semi variance increases as class disagreement between paired pixels becomes more common. It therefore describes the spatial pattern of the cloud field rather than simply its total cloud amount. At short lags, the original CAI-2 mask departs from the evaluation reference mask, and the difference remains as the curves level off near 0.09 and 0.12, respectively. The short-lag difference is associated with fine, closely spaced cloud structure, while the optimized mask remains close to the reference across all lags. Agreement is tightest in the two ranges the tests target, scattered clouds below h ≈ 8 pixels and cloud edges from 8 to 50 pixels, where the optimized and reference curves nearly coincide. These results suggest that recovering the missed clouds by the new spectral tests improves both cloud representation and spatial arrangement, bringing the optimized mask closer to the reference at the scales of small and broken clouds.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

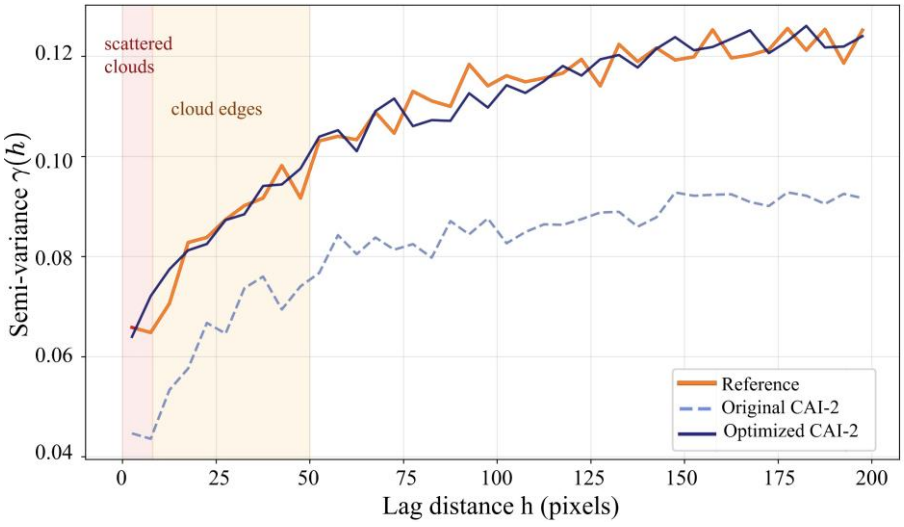


Figure 10: Indicator semi-variogram of the binary cloud masks for granule 4. Shaded bands mark the scattered-cloud (h < 8 px) and cloud-edge (8–50 px) ranges.

4 Discussion

4.1 Main Finding and Implications

The present results confirm that an operational cloud mask's performance is also shaped by an important factor besides its detection rate, the spectral structure of the errors it produces. Over water, CAI-2 L2 cloud mask uses three threshold tests derived from the same VIS and NIR bands. This shared spectral basis can produce correlated cloud omissions, so performance depends on the sensitivity of the individual tests and also the independent information each contributes. To address this, the study converted omission errors in the original CAI-2 cloud mask into a targeted detection problem. The method learned the spectral difference between the clouds the original algorithm detected and the clouds it missed, using native CAI-2 L1B radiances from three bands, UV, VIS, and NIR, in relative two-dimensional X–Y feature spaces. The threshold tests, derived in this way, were used as an independent omission recovery. The improvement in overall detection suggests that the recovery drew on new spectral information. Since the new tests are less likely to repeat the same errors as the original CAI-2 tests, their errors should be less correlated when combined. The CAI-2 cloud mask determines which observations enter the GOSAT-2 gas retrievals, so reducing missed-cloud errors helps limit cloud-related bias in the retrieved CO_2 and CH_4 column concentration.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

4.2 Cloud signals in the selected feature spaces

The results in each granule are best understood by viewing the feature spaces as cloud-detectability fields. Because both groups are clouds, this surface models the conditional probability that a cloud at a given spectral position is detected. Three features contributed to the improved result. Feature 1 combines NIR radiance with the degree to which VIS stands out from UV and NIR. Probability falls toward low NIR with high VIS excess and rises toward higher NIR with lower VIS excess. It is therefore useful for weak, low-NIR clouds that still retain a distinct VIS contrast relative to UV and NIR. Feature 3 is useful for identifying clouds whose signal is clearer in band ratios than in absolute radiance. Probability increases as samples move toward higher NIR/VIS and lower UV/VIS. Feature 4 is useful for detecting clouds with strong NIR radiance and higher NIR contrast relative to VIS. Probability increases as samples move toward higher NIR and higher NIR/VIS. In all three features, the natural log transform converts band ratios into additive spectral contrasts, so small relative differences among low-radiance targets are more clearly evident. Together, these spaces show a curved progression from weak to stronger cloud signals. This curvature separates subtle cloud differences that brightness alone cannot resolve.

4.3 Interpretation of Results

Among the four granules, kappa increased by 5 to 14 percentage points, although the threshold tests reduced a broadly similar share of overlook error in each case. The kappa gains were therefore divergent. Granule 1 shows this most clearly, as it had the smallest kappa gain but removed a large share of its overlook (78%), comparable to the others (64% to 83%). The cloud detection threshold setting does not explain these outcomes. The cloud-side threshold is the cloud probability above which a pixel is classified as cloud. This value was varied from 0.7 in Granule 1 to 0.6 in Granule 2 and 0.3 in Granule 3, yet it did not track the resulting overestimation. In Granule 1, the strictest threshold produced the largest increase in overestimation; in Granule 3, the most permissive threshold added almost no such errors. The overestimation instead came from the scenes themselves. These included mixed pixels that were ambiguous even under visual inspection, a few in the glint area at low cone angle, and the weak contrast between black ocean and white cloud along cloud edges. The four granules differed in cloud-covered water area, time, date, and location, and their sun-glint conditions also varied. It is necessary to note that the threshold tests were derived from footprint-based L2 and L1B training data but were applied at the pixel level for cloud detection. In the last granule, they recovered both cloud edges and thin/scattered clouds, the same targets they were built to detect. These results have two implications. First, a lower probability threshold can recover more overlooked pixels in absolute terms, but the recovered fraction and the false increase depend strongly on scene circumstances. Second, the fact that the same threshold tests recovered comparable spectral targets in a distinct granule suggests the diagnostic information they captured is not specific to the training scenes.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

4.4 Conclusion

Cloud detection based mainly on thresholds from a single band or a band ratio has a limited trade-off between error rates, as improving omission recovery usually requires moving the same threshold that also controls overestimation. This study treated missed clouds over oceans in CAI-2 L2 cloud mask not as random errors, but as a recoverable spectral class. Their distributions with true detections were mapped onto new two-dimensional feature spaces built from the UV, VIS, and NIR bands. And these spaces were then converted into spectral threshold tests based on coefficients for omission recovery. The new threshold tests recovered 40% to 61% of the remaining gap to perfect agreement, about half of the disagreement left by the L2 mask and gave only a modest number of new false detections. These results will contribute to the reduction of GOSAT-2 greenhouse gas concentration errors due to previously undetected clouds over oceans.

4.5 Limitations and Future Work

This study is limited to oceanic granules around Japan and to data from a single year, 2023. The contribution of each threshold test was also not assessed individually. These limitations restrict generalizability without further calibration and validation, and obscure which tests contributed most to omission recovery and overestimation. Future work should extend the analysis to other regions and years, broaden the training stage to include true-negative and false-positive samples with the TP and FN samples. Also, apply a multinomial logistic function for multi-class analysis. The same spectral-test approach could be applied to other satellite cloud masks to assess the method's applicability.

Data availability

The raw data used in this study are available to download from the GOSAT-2 Product Archive (https://prdct.gosat-2.nies.go.jp/index.html.en). The processed rasters for one of the granules analysed here, together with the recovered cloud mask, are archived at https://doi.org/10.5281/zenodo.22246678 (Safabakhshpachehkenari and Matsunaga, 2026).

Code availability

The ArcGIS Pro scripts for the spectral tests, the sunglint checks, and the mask recovery are archived at https://doi.org/10.5281/zenodo.22246678 (Safabakhshpachehkenari and Matsunaga, 2026).

Author contributions

MS and TM conceived and designed the study. MS developed the model code and performed the analyses. MS wrote the manuscript. TM supervised the work and reviewed and edited the manuscript.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Competing interests

The authors declare no conflict of interest

Acknowledgements

This study was supported and funded by NIES GOSAT-2 Project and the Ministry of the Environment of Japan.

Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

References

Bai, T., Li, D., Sun, K., Chen, Y., and Li, W.: Cloud detection for high-resolution satellite imagery using machine learning and multi-feature fusion, Remote Sens., 8, 715, https://doi.org/10.3390/rs8090715, 2016.

Ball, J. E., Anderson, D. T., and Chan, C. S.: Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community, J. Appl. Remote Sens., 11, 042609, https://doi.org/10.1117/1.JRS.11.042609, 2017.

Banks, D. L. and Fienberg, S. E.: Statistics, Multivariate, in: Encyclopedia of Physical Science and Technology, 3rd Edn., edited by: Meyers, R. A., Academic Press, New York, 851–889, https://doi.org/10.1016/B0-12-227410-5/00731-6, 2003.

Barr, A. G., Landgraf, J., Martinez-Velarte, M., Vrekoussis, M., Sussmann, R., Morino, I., Strong, K., Zhou, M., Velazco, V. A., Ohyama, H., Warneke, T., Hase, F., and Borsdorff, T.: Five years of GOSAT-2 retrievals with RemoTeC: XCO_2 and XCH_4 data products with quality filtering by machine learning, Atmos. Meas. Tech., 18, 6093–6123, https://doi.org/10.5194/amt-18-6093-2025, 2025.

Bishop, C. M.: Pattern Recognition and Machine Learning, Springer, New York, 738 pp., 2006.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Bulgin, C. E., Sembhi, H., Ghent, D., Remedios, J. J., and Merchant, C. J.: Cloud-clearing techniques over land for land-surface temperature retrieval from the Advanced Along-Track Scanning Radiometer, Int. J. Remote Sens., 35, 3594–3615, https://doi.org/10.1080/01431161.2014.907941, 2014.

Chen, Y., Fan, R., Bilal, M., Yang, X., Wang, J., and Li, W.: Multilevel cloud detection for high-resolution remote sensing imagery using multiple convolutional neural networks, ISPRS Int. J. Geo-Inf., 7, 181, https://doi.org/10.3390/ijgi7050181, 2018.

Congalton, R. G.: A review of assessing the accuracy of classifications of remotely sensed data, Remote Sens. Environ., 37, 35–46, https://doi.org/10.1016/0034-4257(91)90048-B, 1991.

Diggle, P.: A kernel method for smoothing point process data, J. Roy. Stat. Soc. C-App., 34, 138–147, https://doi.org/10.2307/2347366, 1985.

Dong, J., Wang, Y., Yang, Y., Yang, M., and Chen, J.: MCDNet: Multilevel cloud detection network for remote sensing images based on dual-perspective change-guided and multi-scale feature fusion, Int. J. Appl. Earth Obs., 129, 103820, https://doi.org/10.1016/j.jag.2024.103820, 2024.

Espinar, B., Ramírez, L., Drews, A., Beyer, H. G., Zarzalejo, L. F., Polo, J., and Martín, L.: Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations, Sol. Energy, 83, 118–125, https://doi.org/10.1016/j.solener.2008.07.009, 2009.

Fitzgerald, R. W. and Lees, B. G.: Assessing the classification accuracy of multisource remote sensing data, Remote Sens. Environ., 47, 362–368, https://doi.org/10.1016/0034-4257(94)90103-1, 1994.

Frantz, D.: FORCE – Landsat + Sentinel-2 analysis ready data and beyond, Remote Sens., 11, 1124, https://doi.org/10.3390/rs11091124, 2019.

Frey, R. A., Ackerman, S. A., Holz, R. E., Dutcher, S., and Griffith, Z.: The continuity MODIS-VIIRS cloud mask, Remote Sens., 12, 3334, https://doi.org/10.3390/rs12203334, 2020.

Hagolle, O., Huc, M., Villa Pascual, D., and Dedieu, G.: A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images, Remote Sens. Environ., 114, 1747–1755, https://doi.org/10.1016/j.rse.2010.03.002, 2010.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Haynes, J. M., Noh, Y.-J., Miller, S. D., Haynes, K. D., Ebert-Uphoff, I., and Heidinger, A.: Low cloud detection in multilayer scenes using satellite imagery with machine learning methods, J. Atmos. Ocean. Tech., 39, 319–334, https://doi.org/10.1175/JTECH-D-21-0084.1, 2022.

Hughes, M. J. and Hayes, D. J.: Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing, Remote Sens., 6, 4907–4926, https://doi.org/10.3390/rs6064907, 2014.

Hünerbein, A., Bley, S., Horn, S., Deneke, H., and Walther, A.: Cloud mask algorithm from the EarthCARE Multi-Spectral Imager: the M-CM products, Atmos. Meas. Tech., 16, 2821–2836, https://doi.org/10.5194/amt-16-2821-2023, 2023.

Imasu, R., Matsunaga, T., Nakajima, M., Yoshida, Y., Shiomi, K., Morino, I., Saitoh, N., Niwa, Y., Someya, Y., Oishi, Y., Hashimoto, M., Noda, H., Hikosaka, K., Uchino, O., Maksyutov, S., Takagi, H., Ishida, H., Nakajima, T. Y., Nakajima, T., and Shi, C.: Greenhouse gases Observing SATellite 2 (GOSAT-2): mission overview, Prog. Earth Planet. Sci., 10, 33, https://doi.org/10.1186/s40645-023-00562-2, 2023.

IPCC: Summary for Policymakers, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 3–32, https://doi.org/10.1017/9781009157896.001, 2021.

Ishida, H., Oishi, Y., Morita, K., Moriwaki, K., and Nakajima, T. Y.: Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions, Remote Sens. Environ., 205, 390–407, https://doi.org/10.1016/j.rse.2017.11.003, 2018.

Janardanan, R., Maksyutov, S., Wang, F., Nayagam, L., Yoshida, Y., Lan, X., and Matsunaga, T.: High-resolution inversion of GOSAT-2 retrievals for sectoral methane emission estimates during 2019–2022: A consistency analysis with GOSAT inversion, Remote Sens., 17, 2932, https://doi.org/10.3390/rs17172932, 2025.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

JAXA and NIES: GOSAT-2/IBUKI-2 Data Users Handbook, 1st Edn., Japan Aerospace Exploration Agency and National Institute for Environmental Studies, https://prdct.gosat-2.nies.go.jp/documents/pdf/GOSAT-2_Data_Users_Handbook_1stEdition_en.pdf (last access: 18 June 2026), 2020.

JAXA, NIES, and Ministry of the Environment: GOSAT/IBUKI Data Users Handbook, 1st Edn., Japan Aerospace Exploration Agency, National Institute for Environmental Studies, and Ministry of the Environment, https://data2.gosat.nies.go.jp/doc/GOSAT_HB_E_1stEdition_for_HP.pdf (last access: 17 June 2026), 2011.

Joshi, P. P., Wynne, R. H., and Thomas, V. A.: Cloud detection algorithm using SVM with SWIR2 and tasseled cap applied to Landsat 8, Int. J. Appl. Earth Obs., 82, 101898, https://doi.org/10.1016/j.jag.2019.101898, 2019.

Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Appl. Optics, 48, 6716–6733, https://doi.org/10.1364/AO.48.006716, 2009.

Lee, S., Han, H., Im, J., Jang, E., and Lee, M.-I.: Detection of deterministic and probabilistic convection initiation using Himawari-8 Advanced Himawari Imager data, Atmos. Meas. Tech., 10, 1859–1874, https://doi.org/10.5194/amt-10-1859-2017, 2017.

Li, Y., Chen, W., Zhang, Y., Tao, C., Xiao, R., and Tan, Y.: Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning, Remote Sens. Environ., 250, 112045, https://doi.org/10.1016/j.rse.2020.112045, 2020.

Li, Z., Shen, H., Weng, Q., Zhang, Y., Dou, P., and Zhang, L.: Cloud and cloud shadow detection for optical satellite imagery: Features, algorithms, validation, and prospects, ISPRS J. Photogramm. Remote Sens., 188, 89–108, https://doi.org/10.1016/j.isprsjprs.2022.03.020, 2022.

López-Puigdollers, D., Mateo-García, G., and Gómez-Chova, L.: Benchmarking deep learning models for cloud detection in Landsat-8 and Sentinel-2 images, Remote Sens., 13, 992, https://doi.org/10.3390/rs13050992, 2021.

Lu, K., Kluger, D. M., Bates, S., and Wang, S.: Regression coefficient estimation from remote sensing maps, Remote Sens. Environ., 330, 114949, https://doi.org/10.1016/j.rse.2025.114949, 2025.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Mateo-García, G., Gómez-Chova, L., Amorós-López, J., Muñoz-Marí, J., and Camps-Valls, G.: Multitemporal cloud masking in the Google Earth Engine, Remote Sens., 10, 1079, https://doi.org/10.3390/rs10071079, 2018.

Mei, L., Vountas, M., Gómez-Chova, L., Rozanov, V., Jäger, M., Lotz, W., Burrows, J. P., and Hollmann, R.: A cloud masking algorithm for the XBAER aerosol retrieval using MERIS data, Remote Sens. Environ., 197, 141–160, https://doi.org/10.1016/j.rse.2016.11.016, 2017.

Nakajima, T., Sekiguchi, M., Momoi, M., Shi, C., Hashimoto, M., Babu, S. S., Gogoi, M. M., Goto, D., Madhavan, B. L., Nakajima, T. Y., Shiomi, K., Takenaka, H., Tang, C., Yao, R., Imasu, R., and Matsunaga, T.: Aerosol retrievals from the GOSAT-2/CAI-2 imager with the 2nd version of the multi-wavelength and multi-pixel method, J. Quant. Spectrosc. Radiat. Transf., 360, 109939, https://doi.org/10.1016/j.jqsrt.2026.109939, 2026.

Noël, S., Reuter, M., Buchwitz, M., Borchardt, J., Hilker, M., Bovensmann, H., Burrows, J. P., Di Noia, A., Suto, H., Yoshida, Y., Buschmann, M., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Podolske, J. R., Pollard, D. F., Sha, M. K., Shiomi, K., Sussmann, R., Té, Y., Velazco, V. A., and Warneke, T.: XCO_2 retrieval for GOSAT and GOSAT-2 based on the FOCAL algorithm, Atmos. Meas. Tech., 14, 3837–3869, https://doi.org/10.5194/amt-14-3837-2021, 2021.

Oishi, Y., Ishida, H., Nakajima, T. Y., Nakamura, R., and Matsunaga, T.: The impact of different support vectors on GOSAT-2 CAI-2 L2 cloud discrimination, Remote Sens., 9, 1236, https://doi.org/10.3390/rs9121236, 2017.

Oishi, Y., Ishida, H., Nakajima, T. Y., Nakamura, R., and Matsunaga, T.: Preliminary verification for application of a support vector machine-based cloud detection method to GOSAT-2 CAI-2, Atmos. Meas. Tech., 11, 2863–2878, https://doi.org/10.5194/amt-11-2863-2018, 2018.

Oishi, Y., Ishida, H., and Nakajima, T. Y.: GOSAT-2 TANSO-CAI-2 L2 Cloud Discrimination Processing Algorithm Theoretical Basis Document, NIES-GOSAT2-ALG-20191008-009-00, National Institute for Environmental Studies, GOSAT-2 Project, https://prdct.gosat-2.nies.go.jp/documents/pdf/ATBD_CAI-2_L2_CLDD_en_00.pdf (last access: 18 June 2026), 2020.

Pang, S., Sun, L., Tian, Y., Ma, Y., and Wei, J.: Convolutional neural network-driven improvements in global cloud detection for Landsat 8 and transfer learning on Sentinel-2 imagery, Remote Sens., 15, 1706, https://doi.org/10.3390/rs15061706, 2023.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Parker, R. J., Webb, A., Boesch, H., Somkuti, P., Barrio Guillo, R., Di Noia, A., Kalaitzi, N., Anand, J. S., Bergamaschi, P., Chevallier, F., Palmer, P. I., Feng, L., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Kivi, R., Morino, I., Notholt, J., Oh, Y.-S., Ohyama, H., Petri, C., Pollard, D. F., Roehl, C., Sha, M. K., Shiomi, K., Strong, K., Sussmann, R., Té, Y., Velazco, V. A., Warneke, T., Wennberg, P. O., and Wunch, D.: A decade of GOSAT Proxy satellite CH_4 observations, Earth Syst. Sci. Data, 12, 3383–3412, https://doi.org/10.5194/essd-12-3383-2020, 2020.

Reuter, M., Buchwitz, M., Schneising, O., Noël, S., Bovensmann, H., and Burrows, J. P.: A fast atmospheric trace gas retrieval for hyperspectral instruments approximating multiple scattering – Part 2: Application to XCO_2 retrievals from OCO-2, Remote Sens., 9, 1102, https://doi.org/10.3390/rs9111102, 2017.

Safabakhshpachehkenari, M. and Matsunaga, T.: Code and data for "Targeted spectral threshold tests for reducing cloud omission in the GOSAT-2 CAI-2 cloud mask", Zenodo [data set], https://doi.org/10.5281/zenodo.22246678, 2026.

Shao, Z., Deng, J., Wang, L., Fan, Y., Sumari, N. S., and Cheng, Q.: Fuzzy autoencode based cloud detection for remote sensing imagery, Remote Sens., 9, 311, https://doi.org/10.3390/rs9040311, 2017.

Shi, Z., Li, P., Jin, H., Tian, Y., Chen, Y., and Zhang, X.: Improving super-resolution mapping by combining multiple realizations obtained using the indicator-geostatistics based method, Remote Sens., 9, 773, https://doi.org/10.3390/rs9080773, 2017.

Someya, Y., Yoshida, Y., Ohyama, H., Nomura, S., Kamei, A., Morino, I., Mukai, H., Matsunaga, T., Laughner, J. L., Velazco, V. A., Herkommer, B., Té, Y., Sha, M. K., Kivi, R., Zhou, M., Oh, Y. S., Deutscher, N. M., and Griffith, D. W. T.: Update on the GOSAT TANSO–FTS SWIR Level 2 retrieval algorithm, Atmos. Meas. Tech., 16, 1477–1501, https://doi.org/10.5194/amt-16-1477-2023, 2023.

Stephens, M. A.: Kolmogorov-type tests for exponentiality when the scale parameter is unknown, Technical Report SU-TR-154, Department of Statistics, Stanford University, Stanford, CA, https://apps.dtic.mil/sti/tr/pdf/AD0701308.pdf (last access: 18 June 2026), 1970.

Sun, Q., Zhang, P., Wei, H., Liu, A., You, S., and Sun, D.: Improved mapping and understanding of desert vegetation-habitat complexes from intraannual series of spectral endmember space using cross-wavelet transform and logistic regression, Remote Sens. Environ., 236, 111516, https://doi.org/10.1016/j.rse.2019.111516, 2020.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.

Suto, H., Kataoka, F., Kikuchi, N., Knuteson, R. O., Butz, A., Haun, M., Buijs, H., Shiomi, K., Imai, H., and Kuze, A.: Thermal and near-infrared sensor for carbon observation Fourier transform spectrometer-2 (TANSO-FTS-2) on the Greenhouse gases Observing SATellite-2 (GOSAT-2) during its first year in orbit, Atmos. Meas. Tech., 14, 2013–2039, https://doi.org/10.5194/amt-14-2013-2021, 2021.

Tang, Y., Zhang, L., and Huang, X.: Object-oriented change detection based on the Kolmogorov–Smirnov test using high-resolution multispectral imagery, Int. J. Remote Sens., 32, 5719–5740, https://doi.org/10.1080/01431161.2010.507263, 2011.

Thampi, B. V., Wong, T., Lukashin, C., and Loeb, N. G.: Determination of CERES TOA fluxes using machine learning algorithms. Part I: Classification and retrieval of CERES cloudy and clear scenes, J. Atmos. Ocean. Tech., 34, 2329–2345, https://doi.org/10.1175/JTECH-D-16-0183.1, 2017.

UNEP: Solutions for a healthy, prosperous and resilient people and planet: the United Nations Environment Programme medium-term strategy for 2026–2029, UNEP/EA.7/3, United Nations Environment Programme, https://docs.un.org/en/UNEP/EA.7/3 (last access: 17 June 2026), 2025.

UNFCCC: Paris Agreement, FCCC/CP/2015/10/Add.1, United Nations Framework Convention on Climate Change, https://unfccc.int/sites/default/files/resource/parisagreement_publication.pdf (last access: 17 June 2026), 2015.

Wan, X.: The influence of polynomial order in logistic regression on decision boundary, IOP Conf. Ser. Earth Environ. Sci., 267, 042077, https://doi.org/10.1088/1755-1315/267/4/042077, 2019.

Wang, C., Platnick, S., Meyer, K., Zhang, Z., and Zhou, Y.: A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations, Atmos. Meas. Tech., 13, 2257–2277, https://doi.org/10.5194/amt-13-2257-2020, 2020.

Wang, J., Yang, D., Chen, S., Zhu, X., Wu, S., Bogonovich, M., Guo, Z., Zhu, Z., and Wu, J.: Automatic cloud and cloud shadow detection in tropical areas for PlanetScope satellite images, Remote Sens. Environ., 264, 112604, https://doi.org/10.1016/j.rse.2021.112604, 2021.

Wang, Z., Zhao, L., Meng, J., Han, Y., Li, X., Jiang, R., Chen, J., and Li, H.: Deep learning-based cloud detection for optical remote sensing images: A survey, Remote Sens., 16, 4583, https://doi.org/10.3390/rs16234583, 2024.

https://doi.org/10.5194/egusphere-2026-5309
Preprint. Discussion started: 16 November 2026
© Author(s) 2026. CC BY 4.0 License.



Wolfe, J., Jin, X., Bahr, T., and Holzer, N.: Application of softmax regression and its validation for spectral-based land cover mapping, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-1/W1, 455–459, https://doi.org/10.5194/isprs-archives-XLII-1-W1-455-2017, 2017.

Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., and Maksyutov, S.: Global concentrations of CO_2 and CH_4 retrieved from GOSAT: First preliminary results, SOLA, 5, 160–163, https://doi.org/10.2151/sola.2009-041, 2009.

Yoshida, Y., Ota, Y., Eguchi, N., Kikuchi, N., Nobuta, K., Tran, H., Morino, I., and Yokota, T.: Retrieval algorithm for CO_2 and CH_4 column abundances from short-wavelength infrared spectral observations by the Greenhouse gases observing satellite, Atmos. Meas. Tech., 4, 717–734, https://doi.org/10.5194/amt-4-717-2011, 2011.

Yousaf, R., Rehman, H. Z. U., Khan, K., Khan, Z. H., Fazil, A., Mahmood, Z., Qaisar, S. M., and Siddiqui, A. J.: Satellite imagery-based cloud classification using deep learning, Remote Sens., 15, 5597, https://doi.org/10.3390/rs15235597, 2023.

Zhao, S., Liu, M., Tao, M., Zhou, W., Lu, X., Xiong, Y., Li, F., and Wang, Q.: The role of satellite remote sensing in mitigating and adapting to global climate change, Sci. Total Environ., 905, 166820, https://doi.org/10.1016/j.scitotenv.2023.166820, 2023.

Zhu, X. and Helmer, E. H.: An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions, Remote Sens. Environ., 214, 135–153, https://doi.org/10.1016/j.rse.2018.05.024, 2018.

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., and Fraundorfer, F.: Deep learning in remote sensing: A comprehensive review and list of resources, IEEE Geosci. Remote Sens. Mag., 5, 8–36, https://doi.org/10.1109/MGRS.2017.2762307, 2017.

Zi, Y., Xie, F., and Jiang, Z.: A cloud detection method for Landsat 8 images based on PCANet, Remote Sens., 10, 877, https://doi.org/10.3390/rs10060877, 2018.