The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume L-4/W1-2026
Free and Open Source Software for Geospatial (FOSS4G) 2026 – Academic Track "Bridging Geospatial Technology and Humanity",
30 August–5 September 2026, Hiroshima, Japan

Mapping Nearshore Cladophora in Lake Ontario: An Automated Open-Source Sentinel-2 Workflow and Experimental Index

Shichao Wang[1], Jianqiao Liu[2], Sean J. Bennett[3]
[1] Department of Computer Science, Western University, London, ON N6A 5B7, Canada; swan347@uwo.ca
[2] Department of Geography, University at Buffalo, Buffalo, NY 14261, USA; jliu222@buffalo.edu
[3] Department of Geography, University at Buffalo, Buffalo, NY 14261, USA; seanb@buffalo.edu

Keywords: Sentinel-2; local cloud screening; quicklook; homography; automated workflow; Cladophora

Abstract

Acquiring high-quality Sentinel-2 images is a key first step for nearshore remote-sensing studies, but scene-level cloud metadata can be misleading when the research window is small. This paper presents an open-source, local cloud-aware workflow for two Lake Ontario regions centered on the Olcott (OOL) and Irondequoit (OIR) stations. The workflow first queries Sentinel-2 Level-2A scenes from the Copernicus Data Space STAC catalogue, then downloads only the quicklook image and footprint GeoJSON for each candidate scene. The station ROI is mapped to quicklook image space with a homography transform, cropped, and scored with a local HSV cloud rule. Full spectral bands are downloaded only when the local ROI cloud rate meets the user threshold. Homography improved ROI localization, increasing IoU from 0.2279 to 0.8352 at OIR and from 0.4722 to 0.8706 at OOL. A 500-scene water-land comparison also showed that the mean land cloud rate was about 30 percentage points higher than the lake-water rate. The paper also outlines a controlled tank experiment and retains two candidate Cladophora indices for future validation.

1. Introduction

Sentinel-2 Data has become an important data source for geographical studies. It provides researchers with 10-meter-high-resolution, multi-spectral bands covering both visible and invisible wavelengths (ESA, 2015; Wulder et al., 2016). The datasets provide long-term, repeated-observation data. Repetitive observations are useful because nearshore algal conditions can change rapidly in response to light, nutrients, water clarity, wave exposure, and detachment events (Howell, 2018; Auer et al., 2021). For Lake Ontario, one important target is Cladophora, a filamentous green alga that grows attached to hard substrate in shallow water. It is part of the natural ecosystem, but nuisance growth can cause beach fouling, odor, recreation problems, and management concern (Kuczynski et al., 2016; Howell, 2018).

The mapping of Cladophora is not a main topic in this paper; rather, we focus on the data-acquisition problem. An automated, open-source workflow is also an important problem to solve. The two research areas are station-centered nearshore ROIs. Each ROI is a 6 km by 6 km square, with two observation stations in its center, constructed by the USGS. A Sentinel-2 tile is much larger than these small areas. A tile-level cloud value can therefore describe the wrong area. A scene may be cloudy in most places but clear over the research area. The opposite can also happen: the whole scene may look acceptable, but the small station crop may be cloudy. In both cases, the data quality will be seriously affected.

This problem occurs very often when we download large amounts of Sentinel-2 data. The Copernicus Data Space Ecosystem provides open Sentinel data, but access is still managed through request, processing, and transfer limits (Copernicus Data Space Ecosystem, 2026a). The STAC catalog is useful because it provides a standardized way to search for items and inspect asset links (STAC Contributors, 2024; Copernicus Data Space Ecosystem, 2026b). However, the algorithm on the website does not account for the cloud rate issues we just mentioned.

Many cloud-masking methods operate after the full image is downloaded. Fmask, Sen2Cor-based masks, MAJA, and learning-based methods are useful for final analysis, but they do not remove the cost of downloading many scenes that will later be rejected (Zhu et al., 2015; Baetens et al., 2019; Wright et al., 2024). The workflow in this paper uses preview imagery and footprint geometry to estimate cloud cover inside the station ROI before the expensive band download step.

The paper also tests a hypothesis. Because a mixed Sentinel-2 tile contains both water and land, and because land cloud patterns can differ from open-lake areas, land pixels are expected to be cloudier than lake-surface pixels for the selected Lake Ontario image set. The workflow tests this hypothesis on 500 Sentinel-2 scenes. The mean land cloud rate was about 30 percentage points higher than the lake-water cloud rate. This result supports local cloud screening for small-lake targets rather than relying solely on broad scene metadata.

The main contribution is a reproducible, free, and open-source software workflow. It searches candidate Sentinel-2 scenes, saves only quicklook and footprint files, registers the ROI to the quicklook, calculates a local cloud rate, and downloads bands only for retained scenes. A second contribution is the geometric test. The paper compares a simple geographic scaling method with a homography method that maps the footprint polygon to quicklook image space (Hartley and Zisserman, 2004; Luo et al., 2023). Also, a tank experiment was conducted to simulate how a green object reflects light to satellite sensors.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-L-4-W1-2026-391-2026

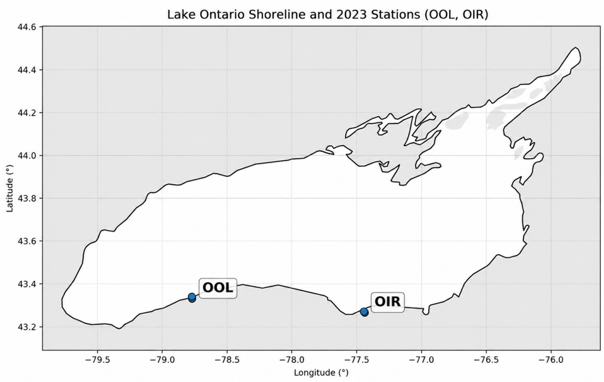

Figure 1. Lake Ontario shoreline and the two station locations used in the workflow. OOL is near Olcott, and OIR is near Irondequoit.

2. Study Area and System Requirements

2.1 Study Area and ROI Geometry

The study area is the southern nearshore of Lake Ontario. The two focal stations are OOL and OIR. Each station ROI is stored as a WGS84 GeoJSON polygon. For this paper, each ROI is a 6 km by 6 km square centered on the station coordinate. This size is large enough to include the local nearshore setting but small enough to show the mismatch between tile-level cloud metadata and the true study window.

The ROI GeoJSON is used in two ways. First, its bounding box is used for the Sentinel-2 catalogue query. Second, the full polygon is later projected into the quicklook image so that cloud cover is measured only inside the local station window. Keeping the ROI in GeoJSON also keeps the workflow portable and easy to audit (Butler et al., 2016).

2.2 Software Architecture

The software follows a pipeline design with clear boundaries between services. The input contract is simple: a user provides an ROI GeoJSON file, a date range, a candidate-scene limit, a local cloud threshold, and an output folder. Authentication is only needed for the final band download step. The system then produces audit artifacts for each candidate scene. These artifacts include the item identifier, footprint GeoJSON, quicklook image, projected ROI image, ROI crop, cloud mask, local cloud score, download flag, and error message when a step fails.

This design uses separation of concerns. The discovery service searches the catalogue. The registration service maps the ROI into image coordinates. The cloud scoring service turns a local crop into a numerical score. The decision service applies the threshold. The download service retrieves JP2 bands only after a scene passes the local gate. This is a software-engineering choice as much as a remote-sensing choice. It makes the workflow easier to test, rerun, and debug.

Table 1 summarizes the functional requirements without tying the paper to long source-code file names. The repository contains GUI programs for these functions, and the full pipeline calls the same core logic for automated processing (f-jgp, 2026).

ID	Requirement	Required output
FR1	Search candidate scenes by ROI and date.	Item list, quicklooks, footprints.
FR2	Register the ROI to quicklook space.	Marked image and ROI crop.
FR3	Score local cloud cover.	Mask and local cloud percent.
FR4	Apply the download gate.	Retained/rejected scene record.
FR5	Download retained bands.	JP2/GeoTIFF band folders.
FR6	Support validation runs.	Accuracy and water-land summaries.

Table 1. Functional requirements for the local cloud-aware acquisition system.

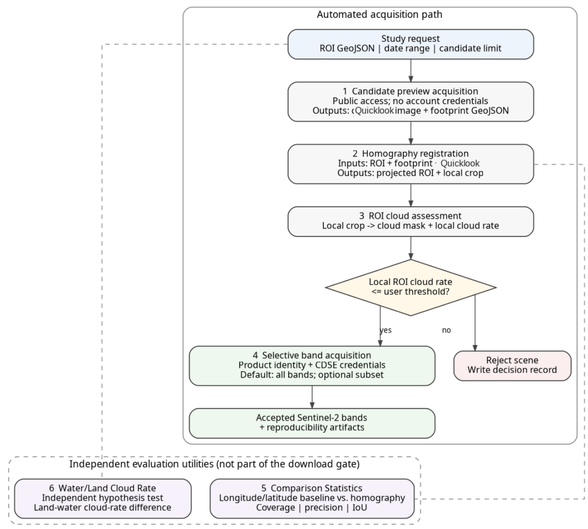

Figure 2. Software architecture for the local cloud-aware download workflow.

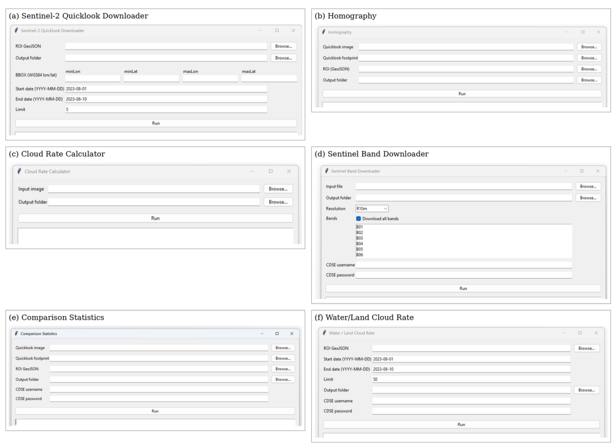

Figure 3. Compact view of the six GUI interfaces used for candidate search, ROI registration, local cloud scoring, band download, quality statistics, and water-land cloud testing.

https://doi.org/10.5194/isprs-archives-L-4-W1-2026-391-2026

2.3 Decision Logic

The workflow is built around one rule: full Sentinel-2 bands are downloaded only when the local station ROI is clear enough. The local decision is made after homography registration and after the quicklook crop is created. This order is important. A cloud score is meaningful only if the crop represents the correct station area. If the crop is shifted toward land or away from the nearshore target, the score describes the wrong place.

The same rule handles two common cases. If a scene has a high overall cloud value but a clear lake ROI, the workflow keeps the scene and downloads bands. If a scene has an acceptable overall cloud value but a cloudy station crop, the workflow rejects the scene. In this way, the system changes the decision from a tile-level question to a station-level question.

2.4 Data Artifacts and Testability

The workflow uses a file-based state model. Each candidate scene receives a folder with a stable set of artifacts: product metadata, footprint, quicklook, marked ROI image, local crop, cloud mask, local cloud value, decision file, and downloaded bands if the scene is accepted. These files are part of the method, not just temporary outputs. They make the scene decision visible.

This design also supports reruns. If the local cloud threshold changes, the workflow can reuse the saved quicklook crop and cloud mask instead of repeating the catalogue search. If authentication or network transfer fails during band download, the pipeline can retry only the accepted scenes. This is useful under quota limits and makes the processing more robust when the same method is run on another computer.

The modules can also be tested separately. The discovery service is correct when it returns a valid item list and saves a footprint. The registration service is correct when the ROI appears in the expected part of the quicklook. The cloud service is correct when the mask matches the bright, low-saturation cloud pixels in the crop. These are simple software checks, but they protect the scientific quality of the final image stack.

2.5 Failure States and Audit Records

The pipeline also treats failure as a normal output state rather than as an unreported exception. A candidate scene can fail because the catalogue item does not expose a usable quicklook, the footprint cannot be reduced to a stable corner order, the transformed ROI falls outside the image, the cloud crop is empty, or the final band download fails during authentication or network transfer. In each case, the system writes a short status message in the scene record. This design keeps rejected scenes visible and prevents a silent bias toward only the easiest products.

From a software-engineering view, the workflow is idempotent at the scene level. If the quicklook and footprint are already saved, the registration and cloud steps can be rerun without a new catalogue query. If only the final band download failed, the earlier screening outputs can be reused and the system can retry the download step. This matters for open data services because failed network calls and quota limits are common in long acquisition jobs. It also makes threshold testing easier. A user can run the same candidate set with a 10%, 20%, or 30% local cloud threshold and compare the retained-scene list without rebuilding all intermediate products.

The audit record is intentionally simple. It stores inputs, intermediate artifacts, and the final decision in ordinary folders. This choice is less elegant than a database, but it is easy to inspect, copy, and archive with a project archive or conference supplement. It also helps users who are not software specialists. They can open the marked quicklook and the mask image to understand why the program kept or rejected a scene.

3. Methods

3.1 Candidate Discovery and Quicklook Cache

The first stage reads the station ROI and extracts its WGS84 bounding box. The search is sent to the Copernicus Data Space STAC endpoint for the Sentinel-2 Level-2A collection. The query uses the ROI bounding box, a start date, an end date, and a user-defined limit. For each returned item, the workflow saves the item geometry as a footprint GeoJSON file and searches the item assets for a quicklook-like image such as a thumbnail, preview, overview, or quicklook (Copernicus Data Space Ecosystem, 2026b; f-jgp, 2026).

This stage is deliberately lightweight. It does not download every spectral band. Footprints and quicklooks are small enough for fast screening but still support local ROI localization and first-pass cloud estimation. The approach is useful when many dates must be checked under bandwidth, storage, and API limits.

3.2 Baseline Localization by Geographic Scaling

The baseline method uses direct geographic scaling. It assumes that the scene footprint can be represented by western, eastern, southern, and northern bounds and that longitude and latitude scale linearly to image columns and rows. Let lon_r and lat_r be the longitude and latitude of an ROI vertex. Let $minlon_f$, $maxlon_f$, $minlat_f$, and $maxlat_f$ be the footprint bounds. Let W and H be the quicklook width and height. The baseline pixel coordinates are:

$$x = \frac{lon_r - minlon_f}{maxlon_f - minlon_f} W \tag{1}$$

$$y = \left(1 - \frac{lat_r - minlat_f}{maxlat_f - minlat_f}\right) H \tag{2}$$

Here, x is the image column and y is the image row. The subtraction from 1 is needed because image rows increase downward while latitude increases northward. The method is easy to implement, but it ignores projective distortion and the actual shape of the footprint-image relationship.

3.3 Homography Registration

The preferred method uses a homography transformation. A homography is a 3 by 3 projective matrix that maps points from one plane to another. It can represent translation, rotation, scale, shear, and perspective effects in one operation (Hartley and Zisserman, 2004; Luo et al., 2023). In this workflow, the four footprint corners are ordered as top-left, top-right, bottom-right, and bottom-left. These geographic points are paired with the four quicklook image corners: $(0,0)$, $(W,0)$, (W,H), and $(0,H)$. OpenCV estimates the matrix from those pairs (Bradski, 2000).

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \mathbf{H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3)$$

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (4)$$

$$x_{img} = \frac{x'}{w'}, \qquad y_{img} = \frac{y'}{w'} \quad (5)$$

In these equations, x and y are geographic ROI coordinates after they have been placed in the same coordinate system as the footprint, x' and y' are homogeneous image coordinates, and w' is the homogeneous scale term. The values h_{11} to h_{33} are the homography coefficients. The perspective terms h_{31} and h_{32} are kept because they are part of the original projective model. Removing them would reduce the method to a simpler affine-like mapping.

After transformation, the projected ROI vertices are clipped to the quicklook image bounds and converted to a crop window. The crop becomes the input for the local cloud detector.

3.4 Local Cloud Rate from the Crop

The cropped quicklook is converted from RGB to HSV color space. HSV separates brightness from saturation. In the quicklooks used here, many opaque cloud pixels are bright and weakly saturated. The default rule marks a pixel as cloud when its normalized value V is greater than 0.75 and its normalized saturation S is less than 0.20:

$$M_c(i,j) = \begin{cases} 1, & V(i,j) > 0.75 \ \wedge \ S(i,j) < 0.20 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The local ROI cloud percentage is:

$$C_{ROI} = 100 \times \frac{\sum_{i=1}^{H_c} \sum_{j=1}^{W_c} M_c(i,j)}{N_{ROI}} \quad (7)$$

Here, $M_c(i,j)$ is the binary cloud mask at crop pixel (i,j), V is value, S is saturation, H_c and W_c are crop height and width, and N_{ROI} is the number of valid pixels in the crop. This rule is not a final cloud mask for reflectance analysis. It is a pre-download screen that prevents large downloads when the station crop is clearly cloudy.

3.5 Water-Land Cloud-Rate Comparison

The water-land test checks the hypothesis that land is cloudier than lake water in the selected image set. The workflow uses the Sentinel-2 Level-2A Scene Classification Layer. Classes 8, 9, and 10 are treated as medium-probability cloud, high-probability cloud, and cirrus (Copernicus Data Space Ecosystem, 2026c). A Global Surface Water occurrence layer is used to separate persistent water from non-water background (Pekel et al., 2016). Pixels with water occurrence greater than or equal to 50 are treated as water.

$$C_{water,k} = \frac{N_{cloud \cap water,k}}{N_{water,k}} \quad (8)$$

$$C_{land,k} = \frac{N_{cloud \cap land,k}}{N_{land,k}} \quad (9)$$

$$\Delta C_k = C_{land,k} - C_{water,k}, \qquad \overline{\Delta C} = \frac{1}{500} \sum_{k=1}^{500} \Delta C_k \quad (10)$$

In these expressions, k is the scene number, $N_{cloud \cap water,k}$ and $N_{cloud \cap land,k}$ are cloud-pixel counts over water and land, and $N_{water,k}$ and $N_{land,k}$ are valid water and land pixel counts. The mean difference $\overline{\Delta C}$ is reported across 500 scenes.

3.6 Localization Metrics

The projected ROI area P was compared with a reference ROI area G. Three metrics were used: coverage, precision, and intersection over union (IoU). Coverage measures how much of the true ROI is captured. Precision measures how much of the predicted crop is correct. IoU measures overall spatial agreement (Congalton and Green, 2019).

$$Coverage = \frac{|G \cap P|}{|G|} \quad (11)$$

$$Precision = \frac{|G \cap P|}{|P|} \quad (12)$$

$$IoU = \frac{|G \cap P|}{|G \cup P|} \quad (13)$$

Here, $|G \cap P|$ is the overlap area, $|G|$ is the reference area, $|P|$ is the predicted area, and $|G \cup P|$ is the union area.

3.7 Pipeline Execution Model

The full automatic chain can be described as a deterministic state machine. For each STAC item, the program creates a scene folder and writes all intermediate results into that folder. The controller then calls the discovery, geometry, quality, and acquisition functions in order. The important rule is that the acquisition function is not allowed to run until a local cloud value exists and passes the threshold. This makes the expensive operation, the band download, the final step rather than the first step.

The execution model can be stated in plain language. Read the ROI. Search scenes by date and bounding box. For each scene, save the footprint and quicklook. Estimate the footprint-to-image homography. Project the ROI into the quicklook. Crop the local window. Calculate C_{ROI}. If C_{ROI} is less than or equal to the threshold, download the selected bands and clip them to the ROI. If not, write a rejection record and continue to the next item. This simple ordering is the main reason the workflow can overcome unnecessary API and bandwidth use.

The pipeline is also designed to be idempotent. If a quicklook or footprint already exists, the program can reuse it instead of downloading it again. If a scene has already failed at the local cloud gate, its rejection record remains in the output folder. This behavior is useful when the user changes only one parameter, such as the local cloud threshold, or when a network failure interrupts a long run.

4. Workflow Results

4.1 ROI Localization

The visual comparison shows that simple longitude-latitude scaling drifted away from the reference ROI, while the homography method stayed close to the station square. This difference was clear at both stations. At OOL, the rectangle crop shifted relative to the nearshore station window. At OIR, the shift was larger and moved a substantial part of the crop away from the true target. Homography performed better because it used the footprint-to-image relationship rather than only footprint bounds.

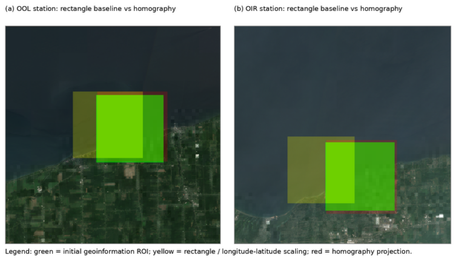

Figure 4. ROI localization comparison. Green is the reference ROI, yellow is rectangle scaling, and red is homography projection.

Case	Coverage	Precision	IoU
OIR rectangle	0.3759	0.3666	0.2279
OIR homography	0.9981	0.8366	0.8352
OOL rectangle	0.6497	0.6336	0.4722
OOL homography	0.9678	0.8966	0.8706

Table 2. Quantitative comparison of rectangle scaling and homography localization.

At OIR, rectangle scaling captured only 37.59% of the reference ROI and produced an IoU of 0.2279. Homography captured 99.81% of the ROI and raised IoU to 0.8352. At OOL, rectangle scaling captured 64.97% with an IoU of 0.4722, while homography captured 96.78% with an IoU of 0.8706. These gains are important because the local cloud rate is only useful when the local crop is correctly placed.

4.2 Automatic Scene Retention

The pipeline turns the individual steps into one automatic download decision. For each candidate scene, it downloads the quicklook and footprint, organizes them into a scene folder, runs homography cropping, calculates local cloud percentage, compares the percentage with the user threshold, and downloads JP2 bands only if the local crop passes. It also writes a summary file with the product ID, cloud percentage, download flag, and error message when a step fails (f-jgp, 2026).

This is the main operational result. The user no longer has to download a large product before knowing whether the station window is useful. The program can retain a scene when overall cloud metadata is high, if the lake ROI is clear. It can also reject a scene with acceptable overall metadata if the station crop is cloudy. The final image stack is therefore tied to the actual research area.

4.3 Water-Land Cloud Result from 500 Scenes

The 500-scene test supported the working hypothesis. Across the image set, the average land cloud rate was about 30 percentage points higher than the lake-surface cloud rate. This does not mean every land pixel was cloudier than every water pixel. It means that the mean difference was large enough to affect scene selection. A global cloud filter can therefore reject a useful lake observation for a reason that is outside the research ROI.

Test item	Value used in paper
Number of images	500 Sentinel-2 scenes
Cloud classes	SCL classes 8, 9, and 10
Water mask rule	JRC water occurrence >= 50
Mean result	mean($C_{land} - C_{water}$) about 0.30
Interpretation	Land was about 30 percentage points cloudier than lake water.

Table 3. Summary of the water-land cloud hypothesis test.

The result gives a data-based reason for local screening. A scene-level cloud value can be pulled upward by cloud over land, even when the lake station window is clear. The workflow therefore does not lower the cloud-quality standard. It changes the spatial unit of the decision from the whole tile to the exact ROI.

5. Tank Experiment for Sensor and Index Design

5.1 Purpose of the Tank Experiment

The workflow above selects suitable Sentinel-2 scenes. A second problem is how to design an index for submerged Cladophora once the scenes are available. Submerged algae are not observed like land vegetation. Light must travel down through the water, interact with the green target and the bottom, and then travel upward to the sensor. Water absorption and scattering change the signal strongly by wavelength. This is why a land vegetation index cannot be assumed to work for benthic Cladophora.

The controlled tank experiment is used as a design step. The goal is not to make a final ecological claim in this paper. The goal is to test which Sentinel-2-like bands remain useful when a green Cladophora-like target is placed at different depths in water. The experiment also defines how spectra should be collected before a later full validation study.

5.2 Tank Geometry and Sensor Choice

The tank setup uses a transparent water tank, a 50 cm by 50 cm green target board, a daylight-like lamp above the water surface, and a receiver above the water looking near nadir. This arrangement better represents satellite remote sensing than a side lamp and side receiver, because Sentinel-2 observes upward water-leaving radiance rather than horizontal light transmission through the tank (Mobley, 1999). The target is placed at several laboratory depths, such as 0.5 m, 1.0 m, and 1.5 m. At each depth, the receiver records the apparent spectral response from the water surface.

A suitable sensor for the laboratory design should first cover the visible and first red-edge range from about 400 to 750 nm. This range is needed for Sentinel-2 B2, B3, B4, and B5, which are the bands most likely to carry a submerged green-target signal. A TriOS RAMSES-type hyperspectral radiometer is a good option for this part of the design because it can measure radiance or irradiance in the UV/VIS range and is commonly used in above-water reflectance geometry (TriOS GmbH, 2026). A three-sensor above-water setup, for example, can measure downwelling irradiance, sky radiance, and water-leaving radiance, which matches the remote-sensing reflectance logic used in aquatic validation work (NASA Earthdata, 2023; Mobley, 1999). In a smaller laboratory version, the same idea can be simplified to a fixed receiver and a stable white-reference procedure, as long as the viewing angle is held constant for every depth.

A second option is a full-range field spectroradiometer. The ASD FieldSpec 4 is useful when the experiment needs both the visible/red-edge response and the SWIR region, because the standard model covers the VNIR and SWIR range with 3 nm VNIR and 10 nm SWIR spectral resolution (Malvern Panalytical, 2026). This matters for the SCI design because the B11 term is a penalty term near 1610 nm. If the laboratory study only tests the green target response with B2 to B5, a visible or VIS-NIR radiometer is enough. If the study tests the complete SCI equation, a full-range spectroradiometer, or a paired sensor system that includes SWIR, is preferred. The key rule is that the same sensor, geometry, and calibration steps must be used for all depths so that changes in the data come from the submerged target depth rather than from instrument setup.

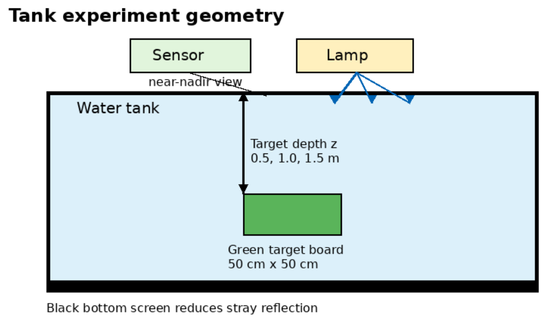

Figure 5. Tank experiment geometry. The green target is placed at 0.5 m, 1.0 m, and 1.5 m depths, with the lamp and receiver above the water.

Element	Design value	Reason
Target	50 cm x 50 cm green board	Cladophora-like signal source.
Depths	0.5, 1.0, 1.5 m	Tests water-column attenuation.
Light	Above-water daylight-like lamp	Simulates downwelling light.
Receiver	Above water, near nadir	Matches satellite-style view.
Sensor	TriOS RAMSES or ASD FieldSpec 4	Records visible/red-edge response; ASD also supports SWIR.
Main bands	B2, B3, B4, B5	Visible and first red-edge response.
Optional band	B11	SWIR penalty for non-water confusion.

Table 4. Main tank experiment design elements.

The tank experiment should be run as a small factorial design. For each depth, the operator records a water-only reading, a green-target reading, a dark reading, and a white-reference reading. The target-present and water-only readings should be collected without changing the receiver height or view angle. This makes the water-only reading a local background for the same optical path. The recommended minimum design is three depths, three repeated readings per depth, and one fixed lamp angle. A second block can repeat the same depth sequence under a different lamp angle, but the angle block should not be mixed with the first block. This design keeps the main depth effect easy to read and avoids overinterpreting the angle effect.

The green object should also be treated as a controlled target, not as a biological sample. Its size, color, position, and orientation should remain fixed during the experiment. The bottom of the tank should be dark or covered with a low-reflectance material so that the recorded signal is dominated by the submerged target and water column. Before each run, the water should be allowed to settle so that bubbles and surface waves do not become a larger source of noise than the target depth.

5.3 Data Collection Plan

Each run should begin with a dark-current reading and a white-reference reading. The water-only background should be recorded before the green target is placed in the tank. The target should then be measured at each depth under the same lamp intensity and viewing geometry. If illumination angle is tested, it should be changed only after a full depth sequence is finished. This avoids mixing depth effects with angle effects.

The preferred measurement order is depth first, then angle. For example, the same target is measured at 0.5 m, 1.0 m, and 1.5 m with the lamp and receiver fixed. Only after that sequence is finished should the lamp angle be changed. This keeps the depth effect easy to separate from illumination geometry. At least three repeat readings should be stored for each depth. The median value can be used in the index-design table, while the spread among repeats can be used later as a simple uncertainty estimate. A run sheet should also record water depth, target depth, lamp angle, receiver height, room-light condition, and file name. This small amount of metadata is important because the spectral values are only useful if the optical geometry can be repeated.

The output of this stage is a simple band-by-depth table rather than a final map. For each Sentinel-2-like band, the table stores the calibrated response of the green target at each tested depth. The same table structure can later be filled with measured data from the selected sensor. This keeps the experiment close to the later satellite workflow: data are measured, calibrated, stored in a reproducible table, and then used to evaluate candidate formulas.

For each band, apparent reflectance can be computed from the target reading, the dark reading, and the white-reference reading:

$$R_b = \frac{L_{target,b} - L_{dark,b}}{L_{white,b} - L_{dark,b}} \tag{14}$$

Here, R_b is apparent reflectance for band b, $L_{target,b}$ is the target radiance or sensor count, $L_{dark,b}$ is the dark reading, and $L_{white,b}$ is the white-reference reading. The same calibration sequence should be repeated for every target depth. If water-only background subtraction is used, the corrected response is:

$$R_{corr,b} = R_{target,b} - R_{water,b} \tag{15}$$

where $R_{water,b}$ is the measured water-only response for band b. This correction helps separate the green target from water-column brightness.

For conference reporting, the tank data should be presented as design evidence rather than as a complete ecological validation. The useful information is the experimental path: which depths were tested, which sensor was used, how the target and water-only readings were calibrated, and which bands were retained for candidate formulas. Field validation with Lake Ontario observations should remain a separate step.

5.4 Example Band Table and Two Candidate Indices

The simulated response table below is used only to show the planned data format and the expected band behavior under different target depths. It is not treated as a final index validation result. In an actual run, the same table would be populated from sensor measurements after dark-current and white-reference correction. The example values illustrate the type of depth-dependent response that the experiment is designed to capture, especially the change in visible and red-edge bands as the green target is lowered in the tank.

Band (nm)	0.5 m	1.0 m	1.5 m
B2 490	2.7950	1.2018	0.5168
B3 560	5.3101	2.8197	1.4973
B4 665	0.6850	0.1564	0.0357
B5 705	0.5588	0.0520	0.0048
B11 1610	0.0000	0.0000	0.0000

Table 5. Example apparent reflectance values from the tank simulation, reported by target depth.

Based on the example band table, only two index candidates are retained in this conference paper. The first is a simple green-red contrast. It uses B3 as the main green response and B4 as the red absorption band:

$$CI_{GR} = \frac{B_3 - B_4}{B_3 + B_4} \tag{16}$$

The second is a more complete candidate Submerged Cladophora Index. It keeps the same green-red core, adds weak B5 red-edge support for shallow water, and uses B11 as a penalty term for land, floating material, or other non-submerged confusion:

$$SCI = \frac{B_3 - B_4}{B_3 + B_4} + 0.27\left(\frac{B_5 - B_4}{B_5 + B_4}\right) - 0.50\left(\frac{B_{11}}{B_3 + B_{11}}\right) \tag{17}$$

In Equations (16) and (17), B3, B4, B5, and B11 are Sentinel-2 reflectance values after the same preprocessing. B3 is the green band, B4 is the red band, B5 is the first red-edge band, and B11 is a shortwave-infrared band. These formulas are design candidates only. Future research will continue this topic by collecting measured tank spectra, matching the retained Sentinel-2 scenes with field observations, and refining the coefficients and thresholds before the index is used for formal mapping.

6. Conclusions

This paper presents a local cloud-aware Sentinel-2 workflow for nearshore Lake Ontario monitoring. The core problem is that a small lake ROI can be clear even when the full Sentinel-2 scene fails a broader cloud filter. The workflow addresses this problem by downloading quicklook and footprint files first, mapping the station ROI into quicklook space, estimating local cloud cover, and downloading full spectral bands only for retained scenes.

The geometric test shows why homography is needed. Compared with direct longitude-latitude rectangle scaling, homography increased IoU from 0.2279 to 0.8352 at OIR and from 0.4722 to 0.8706 at OOL. The 500-scene water-land comparison further supports the local-screening approach, because land cloud rate was about 30 percentage points higher than lake-water cloud rate in this dataset.

The tank experiment is reported as an experimental design step rather than a final index validation. The current paper keeps the simulated band-response table and proposes two possible index directions: a simple green-red contrast and a broader SCI form with B5 support and a B11 penalty. Future work will continue this topic by completing measured tank experiments, testing the two candidate indices against field-matched Sentinel-2 scenes, and refining the final Cladophora index before operational mapping.

Software Availability

The implementation is available as an open-source repository (f-jgp, 2026). It includes GUI tools for scene search, ROI registration, local cloud scoring, band download, quality statistics, water-land cloud testing, and a full pipeline path for automated batch processing.

References

Auer, M.T., McDonald, C.P., Kuczynski, A., Huang, C., Xue, P., 2021. Management of the phosphorus-Cladophora dynamic at a site on Lake Ontario using a multi-module bioavailable P model. *Water*, 13(3), 375. doi.org/10.3390/w13030375.

Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of Copernicus Sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. *Remote Sensing*, 11(4), 433. doi.org/10.3390/rs11040433.

Bradski, G., 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 25(11), 120-125.

Butler, H., Daly, M., Doyle, A., Gillies, S., Hagen, S., Schaub, T., 2016. The GeoJSON Format. Internet Engineering Task Force, RFC 7946. doi.org/10.17487/RFC7946.

Congalton, R.G., Green, K., 2019. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 3rd ed. CRC Press, Boca Raton. doi.org/10.1201/9780429052729.

Copernicus Data Space Ecosystem, 2026a. Quotas and limitations. Copernicus Documentation. documentation.dataspace.copernicus.eu/Quotas.html (15 June 2026).

Copernicus Data Space Ecosystem, 2026b. STAC product catalogue. Copernicus Documentation. documentation.dataspace.copernicus.eu/APIs/STAC.html (15 June 2026).

Copernicus Data Space Ecosystem, 2026c. Sentinel-2 L2A data and scene classification classes. Copernicus Documentation. documentation.dataspace.copernicus.eu/APIs/SentinelHub/Data/S2L2A.html (15 June 2026).

ESA, 2015. Sentinel-2 User Handbook. European Space Agency. sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook (15 June 2026).

f-jgp, 2026. Mapping Nearshore Cladophora in Lake Ontario. GitHub software repository. github.com/f-jgp/Mapping-Nearshore-Cladophora-in-Lake-Ontario (15 June 2026).

Hartley, R., Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, Cambridge. doi.org/10.1017/CBO9780511811685.

Howell, E.T., 2018. Cladophora (green algae) and dreissenid mussels over a nutrient loading gradient on the north shore of Lake Ontario. *Journal of Great Lakes Research*, 44(1), 86-104. doi.org/10.1016/j.jglr.2017.10.005.

Kuczynski, A., Auer, M.T., Brooks, C.N., Grimm, A.G., 2016. The Cladophora resurgence in Lake Ontario: characterization and implications for management. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(6), 999-1013. doi.org/10.1139/cjfas-2015-0261.

Luo, Y., Yu, J., Ren, P., Gao, W., 2023. A review of homography estimation: advances and challenges. *Electronics*, 12(24), 4977. doi.org/10.3390/electronics12244977.

Malvern Panalytical, 2026. FieldSpec 4 Standard-Res Spectroradiometer. Malvern Panalytical product documentation. malvernpanalytical.com/en/products/product-range/asd-range/fieldspec-range/fieldspec-4-standard-res-spectroradiometer (15 June 2026).

Mobley, C.D., 1999. Estimation of the remote-sensing reflectance from above-surface measurements. *Applied Optics*, 38(36), 7442-7455. doi.org/10.1364/AO.38.007442.

NASA Earthdata, 2023. SHIFT: Ramses TriOS radiometer above-water measurements. NASA Earthdata dataset documentation. doi.org/10.3334/ORNLDAAC/2234.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540, 418-422. doi.org/10.1038/nature20584.

STAC Contributors, 2024. SpatioTemporal Asset Catalog specification. github.com/radiantearth/stac-spec (15 June 2026).

TriOS GmbH, 2026. RAMSES hyperspectral radiometers. TriOS product documentation. trios.de/en/ramses/ (15 June 2026).

Wright, N., Duncan, J.M.A., Callow, J.N., Thompson, S.E., George, R.J., 2024. CloudS2Mask: a novel deep learning approach for improved cloud and cloud shadow masking in Sentinel-2 imagery. *Remote Sensing of Environment*, 306, 114122. doi.org/10.1016/j.rse.2024.114122.

Wulder, M.A., Hilker, T., White, J.C., Coops, N.C., Masek, J.G., Pflugmacher, D., Crevier, Y., 2016. Virtual constellations for global terrestrial monitoring. *Remote Sensing of Environment*, 170, 62-76. doi.org/10.1016/j.rse.2015.09.001.

Zhu, Z., Wang, S., Woodcock, C.E., 2015. 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*, 159, 269-277. doi.org/10.1016/j.rse.2014.12.014.