ASSESSING THE GEOGRAPHIC TRANSFERABILITY OF THE FLOOD-HI INDEX FOR FLOOD EXTENT MAPPING IN SPECTRALLY COMPLEX ENVIRONMENTS

Mateus Domingos[1]*; Guilherme Palermo Coelho*[1]*; Murilo Cesar Lucas*[1]

RESUMO - Este estudo avalia a transferabilidade do Índice Híbrido de Detecção Orientada a Inundações (FLOOD-HI). Desenvolvido com o uso do Google Earth Engine a partir de imagens Sentinel-2, o índice utiliza regressão linear multivariada sobre seis métricas (NDWI, $AWEI_{sh}$, TCW, IMP, NDVI, SAVI) para isolar a água de ruídos urbanos, vegetacionais e atmosféricos. O índice foi testado em dois cenários do Copernicus EMS: inundação com alta turbidez (Moçambique) e sob interferência de nuvens pós-Ciclone IDAI (Moçambique). A validação quantitativa revelou que o FLOOD-HI supera índices convencionais, obtendo os maiores F1-scores e Intersection over Union (IoU) em todos os domínios. O modelo mitigou com sucesso erros de omissão e comissão causados por sedimentos suspensos e aerossóis, confirmando-se como uma ferramenta confiável e transferível para o mapeamento óptico automatizado em ambientes complexos.

Palavras-chave: Sentinel-2; Índices espectrais; Gestão de desastres.

ABSTRACT - This study assesses the transferability of the FLOod Oriented Detection Hybrid Index (FLOOD-HI). Developed within Google Earth Engine from Sentinel-2 imagery, FLOOD-HI applies multivariate linear regression to six spectral metrics (NDWI, $AWEI_{sh}$, TCW, IMP, NDVI, SAVI) to isolate water from urban, vegetation, and atmospheric confounders. The index was tested across two Copernicus EMS scenarios: highly turbid riverine flooding (Mozambique) and cloud-covered conditions post-Cyclone IDAI (Mozambique). Quantitative validation showed that FLOOD-HI consistently outperforms conventional indices, achieving the highest F1-scores and Intersection over Union (IoU) across all domains. By successfully mitigating omission and commission errors caused by suspended sediments and aerosols, FLOOD-HI proved to be a highly reliable and transferable tool for automated optical mapping in complex environments.

Keywords: Sentinel-2; Spectral indices; Disaster management.

INTRODUCTION

Automated water extraction in spectrally complex environments with satellite-based optical remote sensing—particularly via the Sentinel-2 Multispectral Instrument (MSI)—remains a critical challenge (Vongkusolkit et al., 2023). Standard spectral indices, such as the Normalized Difference Water Index (NDWI) (McFEETERS, 1996) and the Automated Water Extraction Index ($AWEI_{sh}$) (Feyisa et al., 2014), frequently struggle to distinguish inundated zones from confounding elements like severe water turbidity, dense vegetation, and atmospheric noise (clouds and aerosols), leading to significant misclassifications and false-positive detections (Pekel et al., 2016; Huang et al., 2025).

To address these inherent optical limitations, the FLOod Oriented Detection Hybrid Index (FLOOD-HI) was recently proposed (Domingos et al., 2026). Although previous studies have established its initial efficacy, a critical requirement for any robust spectral index is its geographic

[1]Faculdade de Tecnologia, Universidade Estadual de Campinas (UNICAMP), 1888 Paschoal Marmo St. - Jardim Nova Italia, Limeira, São Paulo 13484-332, Brazil. Emails: mdomingos32@gmail.com, gpcoelho@unicamp.br, mlucas@unicamp.br

transferability and stability across varied hydrological regimes. Therefore, the primary objective of this work is to assess the operational robustness of FLOOD-HI across diverse geographic domains and extreme disaster scenarios.

In this study, we applied the index to map flood extents across two distinct Areas of Interest (AOIs) derived from official Copernicus Emergency Management Service (EMS) activations. These challenging scenarios encompass a highly turbid riverine flood in Mozambique (Chibabel, January 2026) and the catastrophic, cloud-covered aftermath of Tropical Cyclone IDAI in Mozambique (Quelimane, March 2019). By testing the algorithm across these highly varied optical constraints, we aim to demonstrate its transferability beyond its original calibration zones, providing an accurate and reliable mapping tool for heterogeneous regions.

METHODOLOGY

The methodological framework employed in this study integrates cloud-based geospatial processing within Google Earth Engine (GEE) to apply and evaluate the FLOOD-HI, a statistically optimized flood detection index. A comprehensive description of the initial spectral analysis, mathematical formulation, and calibration procedures is available in our companion paper currently under review (Domingos et al., 2026).

As illustrated in Figure 1, the complete workflow progresses through six sequential phases. It is crucial to emphasize that Phases 1 through 5 represent the foundational development and internal validation of the index, which are extensively detailed in our previous work. The present study focuses exclusively on Phase 6, introducing an independent, multi-regional transferability assessment under extreme international disaster scenarios.

To rigorously assess this operational transferability, the target areas for Phase 6 were strategically selected to represent distinct, extreme optical constraints frequently encountered in disaster mapping. The Chibabel domain (January 2026) was chosen to evaluate the model's robustness against severe water turbidity and massive suspended sediment loads, which typically cause omission errors in standard single-band ratios. Conversely, the Quelimane domain (March 2019, Tropical Cyclone IDAI) was selected to test the algorithm's resilience under massive atmospheric interference—such as dense clouds and scattered aerosols—which classically induce severe commission errors (false positives). Furthermore, the seven-year gap between these two events introduces a crucial temporal dimension to the validation, allowing us to simultaneously assess the temporal stability of FLOOD-HI and ensure its operational reliability is independent of specific chronological windows. Together, these contrasting scenarios provide a comprehensive spatio-temporal stress-test for the hybrid index using official Copernicus EMS Rapid Mapping activations as the baseline.

The core procedures of the workflow are summarized below:

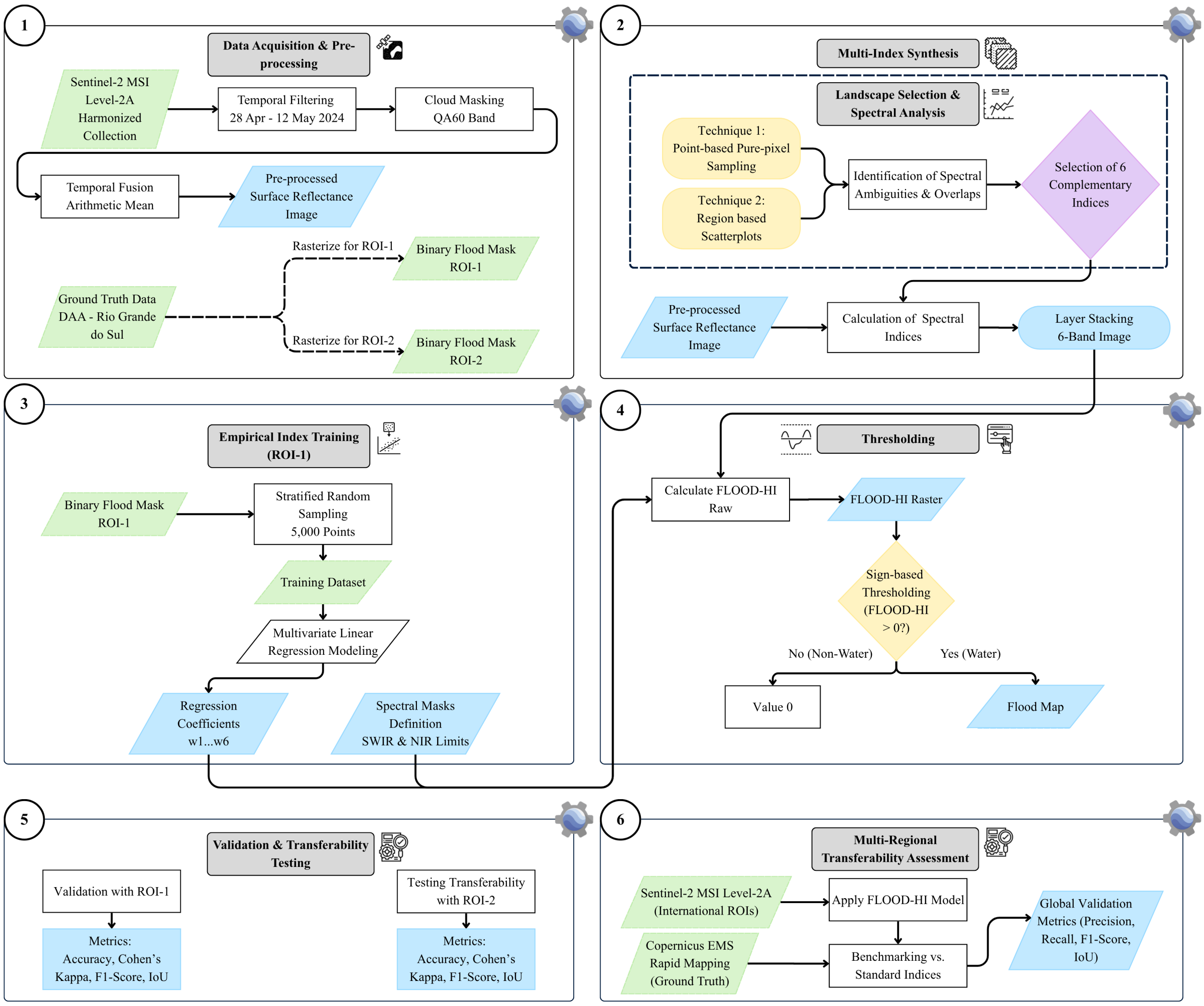

Figure 1: Methodological flowchart. Phases 1 to 5 encompass the original formulation and internal validation of the FLOOD-HI (Domingos et al., 2026). Phase 6 represents the core contribution of this study: testing the index's spatio-temporal transferability across diverse global environments using Copernicus EMS reference data.

- **Phases 1 to 5: Foundational Framework (Previous Work):**
  - *Data Acquisition and Synthesis:* Utilization of Harmonized Sentinel-2 MSI Level-2A imagery (processed at a 10 m spatial resolution) and selection of six complementary spectral indices (NDWI, $AWEI_{sh}$, TCW, IMP, NDVI, and SAVI) to target water signals while suppressing vegetation and urban confounders.
  - *Calibration and Thresholding:* The final formula was determined via Multivariate Linear Regression trained on stratified random points within a calibration domain in Rio Grande do Sul, Brazil (ROI-1). This resulted in a weighted linear combination with a reliable sign-based thresholding approach (values > 0 classified as water).
  - *Internal Validation:* Initial accuracy assessment of the raw model against local ground

truth data. This baseline segmentation capacity was tested in an independent local domain, also in Rio Grande do Sul (ROI-2), before exporting the model to international scenarios.

- **Phase 6: Multi-Regional Transferability Assessment (Current Study):**
  - *Global Application:* To rigorously test the spatial transferability of the algorithm beyond its original calibration zone, the established FLOOD-HI model was applied to two highly distinct international Areas of Interest (AOIs): a highly turbid riverine flood in Mozambique (Chibabel) and the cloud-covered aftermath of a tropical cyclone in Mozambique (Quelimane).
  - *Benchmarking and Accuracy Assessment:* The binary flood maps generated by FLOOD-HI for these extreme events were benchmarked against traditional single-band ratios (NDWI, MNDWI, $AWEI_{sh}$, TCW). The quantitative validation was performed using official Copernicus EMS Rapid Mapping polygons as ground truth, utilizing metrics such as Accuracy, Precision, Recall, F_1-Score, and Intersection over Union (IoU) to evaluate phenomenological error mitigation.

RESULTS AND DISCUSSION

FLOOD-HI Formulation

The multivariate linear regression calibration, performed in the ROI-1 domain, yielded the following formulation for the FLOOD-HI:

$$\begin{aligned} \mathrm{FLOOD-HI} = & 3.72 \cdot \mathrm{NDWI} - 2.67 \cdot \mathrm{IMP} + 2.69 \cdot \mathrm{AWEI_{sh}} \\ & - 1.77 \cdot \mathrm{TCW} - 2.67 \cdot \mathrm{NDVI} + 3.30 \cdot \mathrm{SAVI} \end{aligned} \quad (1)$$

The coefficient signs reflect the physical logic of the hybrid index: positive weights on NDWI and $AWEI_{sh}$ amplify water signals and suppress shadows, while negative weights on IMP, NDVI, and TCW act as suppressors for urbanization, green vegetation, and bright soil artifacts, respectively. Furthermore, the strong positive weight on SAVI acts as a crucial counterbalance to NDVI; it fine-tunes the vegetation suppression by adjusting for soil background reflectance, thereby preventing the erroneous masking of highly turbid, sediment-rich floodwaters. This structure ensures that a pixel is only classified as inundated when water signals overpower the confounding background signals.

Assessment of Transferability: Multi-regional Flood Mapping

To evaluate the geographic transferability and robustness of the proposed methodology, the analysis was streamlined to focus on two highly representative Regions of Interest (ROIs), encompassing distinct hydrological regimes and environmental challenges. To ensure a rigorous quantitative assessment, the ground truth for flooded areas in all calculated metrics (Accuracy, Precision, Recall, F_1-score, IoU, and Kappa) was strictly derived from official crisis maps provided by the Copernicus Emergency Management Service (EMS) Rapid Mapping activations (European Parliament and Council of the European Union, 2014; Voigt et al., 2016).

The first evaluation was conducted in the Chibabel domain (Mozambique, Copernicus EMS activation EMSR857_AOI01), capturing a severe riverine flood event triggered by heavy rainfall in January 2026 that affected multiple regions across Mozambique and South Africa. Riverine floods are characteristically challenging to map due to the massive transport of suspended sediments, which dramatically alters the spectral signature of the water, increasing its reflectance in the visible and near-infrared (NIR) bands (Doxaran et al., 2002).

As shown in Table 1, FLOOD-HI demonstrated exceptional performance in this heterogeneous environment, achieving the highest overall Accuracy (0.88), Precision (0.92), F_1-score (0.93), and IoU (0.87). While MNDWI yielded the highest Kappa coefficient (0.63), FLOOD-HI registered a lower Kappa (0.24). This apparent ambiguity is rooted in cross-sensor validation dynamics rather than true classification superiority. The official Copernicus EMS ground truth, derived from SAR imagery and establishing the maximum active flood extent (Figure 2a), maps the extensive inundation seamlessly through cloud cover. In the corresponding Sentinel-2 optical image, clouds obscure portions of the inundated area. MNDWI notoriously misclassifies these thick clouds as water (a systemic failure visually demonstrated later in Figure 3b for the Quelimane domain); because these atmospheric artifacts spatially coincide with the SAR-detected flood underneath, MNDWI's false positives accidentally align with the SAR reference, artificially inflating its agreement metrics (Kappa). In contrast, FLOOD-HI successfully suppresses atmospheric noise, correctly excluding clouds. While this strict physical logic results in apparent omission errors when compared to the unobstructed SAR ground truth—subsequently penalizing its Kappa—it proves that FLOOD-HI maps genuine water signatures rather than capitalizing on coincidental misclassifications. Finally, traditional indices like NDWI collapsed completely, registering a severe drop in Precision (0.16) and F_1-score (0.28).

Table 1: Validation metrics for the Chibabel domain (Mozambique, Jan 2026 - Riverine Flood). FLOOD-HI shows the best overall performance in F_1-score and IoU.

Index	Accuracy	Precision	Recall	F_1-score	IoU	Kappa
FLOOD-HI	**0.88**	**0.92**	0.95	**0.93**	**0.87**	0.24
NDWI	0.39	0.16	**0.96**	0.28	0.16	0.09
MNDWI	0.84	0.84	0.94	0.88	0.79	**0.63**
$AWEI_{sh}$	0.81	0.79	0.94	0.86	0.75	0.58
TCW	0.79	0.75	0.95	0.84	0.72	0.55

The quantitative superiority of FLOOD-HI is visually corroborated in Figure 2. Figure 2a establishes the baseline, displaying the maximum flood extent identified by the official Copernicus EMS reference. Standard indices often fail to identify water bodies with high sediment loads, as the physical properties of turbid water mimic bare soil or sparse vegetation in standard spectral band ratios. Figure 2b illustrates this critical limitation: NDWI completely omits extensive flooded areas characterized by high turbidity (visible in brown tones), leading to severe underestimations of the disaster's true extent in those specific zones. Conversely, Figure 2c demonstrates that the hybrid formulation of FLOOD-HI successfully captures these highly sedimented floodwaters, classifying the entire inundated section without confusing the turbid water with the surrounding terrestrial landscape.

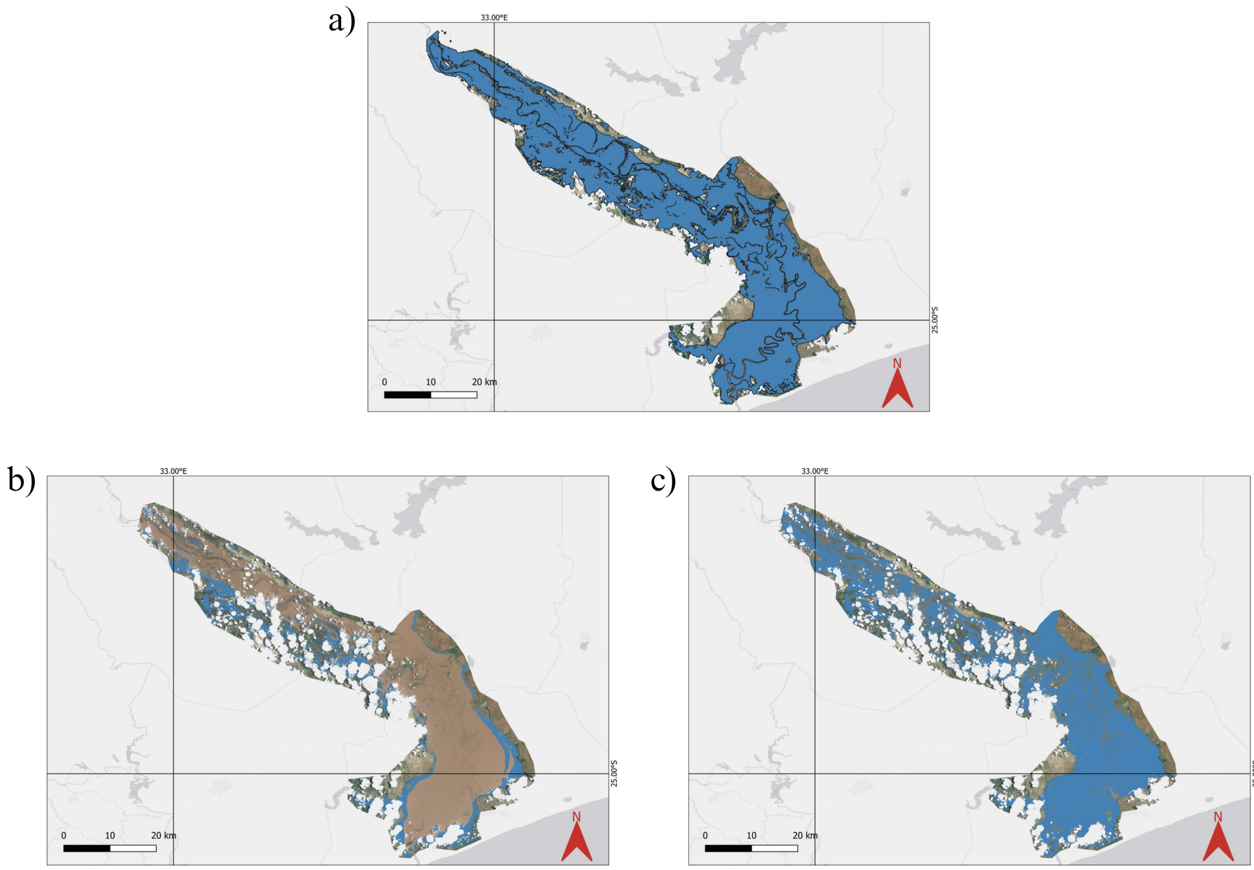

Figure 2: Comparison of flood extent extraction in highly turbid conditions (Chibabel domain). (a) Official Copernicus EMS reference map detailing the maximum active flood extent in dark blue. (b) The NDWI map fails to identify vast areas of sediment-laden floodwaters (brown areas), leading to significant omission errors. (c) The proposed FLOOD-HI successfully detects the entire inundated extent, demonstrating robustness regardless of water turbidity levels.

The second evaluation examines the Quelimane domain, capturing the catastrophic aftermath of Tropical Cyclone IDAI in March 2019 (Copernicus EMS activation EMSR348_AOI01). Making landfall with sustained winds of up to 167 km/h, the cyclone generated heavy rainfall and storm surges that devastated the Sofala region. This scenario represents an extreme, highly adverse mapping environment. Optical remote sensing during cyclonic events is typically severely hindered by massive atmospheric interference (dense clouds and aerosols) and extreme variations in water turbidity (Guo et al., 2021).

As evidenced in Table 2, these extreme conditions caused a steep drop in overall metrics, particularly Recall, for all tested algorithms. Crucially, this generalized degradation explicitly highlights the inherent limitations of optical-only indices during extreme weather events when compared to Synthetic Aperture Radar (SAR) products. While optical sensors fail to penetrate dense cloud cover, SAR signals remain unobstructed, making radar-based mapping the inherently superior standard for immediate crisis response in cyclonic scenarios. Furthermore, this performance drop in optical assessments is often exacerbated by temporal discrepancies between the delayed optical satellite pass and the earlier SAR acquisitions typically used to generate the Copernicus ground truth in cloudy conditions (Tarpanelli et al., 2022). Despite these severe inherent optical constraints, FLOOD-HI maintained the highest relative performance, registering the best F_1-score (0.29), IoU (0.17), and Recall (0.18), while tying with MNDWI for the highest Precision (0.72).

Table 2: Validation metrics for the Quelimane domain (Mozambique, Mar 2019 - Tropical Cyclone IDAI).

Index	**Accuracy**	**Precision**	**Recall**	**F_1-score**	**IoU**	**Kappa**
FLOOD-HI	0.80	**0.72**	**0.18**	**0.29**	**0.17**	**0.22**
NDWI	0.91	0.12	0.07	0.09	0.05	0.05
MNDWI	0.71	**0.72**	0.08	0.15	0.08	0.09
AWEI	0.81	0.65	0.12	0.20	0.11	0.15
TCW	**0.92**	0.36	**0.18**	0.24	0.13	0.20

Figure 3 provides a detailed visual breakdown of the specific phenomenological failures of standard indices and the corresponding corrections achieved by the proposed method. Figure 3a establishes the baseline Copernicus EMS reference. The subsequent zoomed panels illustrate three distinct failure modes of traditional formulations:

First, regarding atmospheric sensitivity (Panels b, c, d): MNDWI (b) and AWEI$_{sh}$ (c) exhibited severe spectral confusion, erroneously classifying thick clouds and scattered aerosols as flooded areas (red arrows). FLOOD-HI (d) successfully suppressed these atmospheric artifacts, preserving high precision. Second, regarding low-turbidity omission (Panels e, f): TCW (e) failed to detect several flooded patches characterized by lower turbidity levels. FLOOD-HI (f) correctly identified these inundations. Third, regarding high-turbidity omission (Panels g, h): Conversely, NDWI (g) was entirely blind to highly turbid, sediment-rich floodwaters, a common limitation when relying solely on standard band ratios that confuse mud with bare soil. Once again, FLOOD-HI (h) accurately captured these complex water signatures.

Ultimately, this multi-panel comparison demonstrates that while single indices fail under specific spectral conditions (either succumbing to atmospheric noise or missing varying levels of turbidity), the hybrid structure of FLOOD-HI simultaneously mitigates these diverse confounding factors.

CONCLUSION

This study successfully demonstrated the geographic and spatio-temporal transferability of the FLOOD-HI index for automated flood extent mapping. By rigorously testing the algorithm

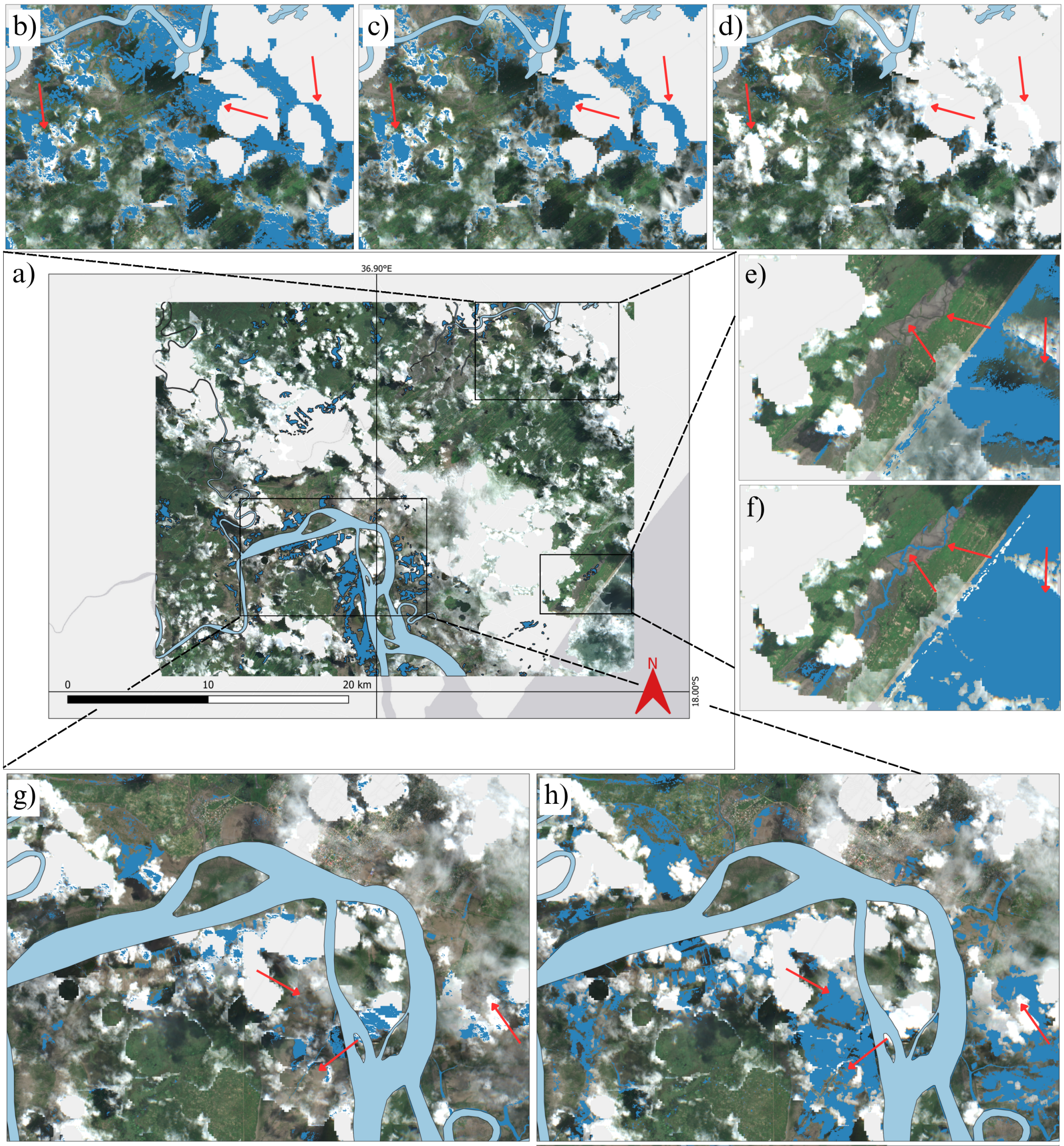

Figure 3: Detailed visual assessment of failure modes and hybrid index corrections in the Quelimane domain (Tropical Cyclone IDAI). (a) Copernicus EMS reference map displaying hydrography (light blue) and active flood extent (dark blue). (b, c) Zoomed region showing MNDWI and $AWEI_{sh}$ erroneously classifying clouds and aerosols as inundation (red arrows). (d) FLOOD-HI correctly suppressing these atmospheric artifacts in the same area. (e) A second region highlighting TCW's failure to detect low-turbidity floodwaters. (f) FLOOD-HI accurately detecting these specific inundated patches. (g) A third region where NDWI completely fails to identify high-turbidity, sediment-rich floods. (h) FLOOD-HI successfully capturing the highly turbid inundations, demonstrating its stability across varying spectral conditions.

across highly distinct and extreme disaster scenarios—a highly turbid riverine flood (Chibabel) and a cloud-covered cyclonic event (Quelimane)—this research confirms that FLOOD-HI's statistically calibrated, multivariate formulation remains robust and highly operational far beyond its original calibration domains.

The quantitative and visual assessments establish that the proposed hybrid structure consistently mitigates the most critical phenomenological failures of traditional single-band ratios. In sediment-rich environments, FLOOD-HI effectively eliminated severe omission errors common to indices like NDWI by appropriately counterbalancing vegetation and soil reflectance signals. Conversely, in extreme atmospheric conditions, the index exhibited superior resilience by systematically suppressing commission errors (false positives) triggered by dense clouds and aerosols, a systemic vulnerability observed in MNDWI and $AWEI_{sh}$. Consequently, despite the inherent cross-sensor challenges when benchmarking optical data against SAR-derived ground truth, FLOOD-HI achieved the highest F_1-scores and Intersection over Union (IoU) across all evaluated scenarios.

Ultimately, these findings carry significant implications for global disaster management. By leveraging the cloud-computing architecture of Google Earth Engine and open-access Sentinel-2 imagery, FLOOD-HI offers a fast, accessible, and reliable operational tool for rapid crisis response. While SAR technology remains the absolute standard for mapping during active, cloud-heavy storms, FLOOD-HI maximizes the utility of optical remote sensing, providing emergency responders with a highly accurate and autonomous mapping solution capable of navigating the complex spectral realities of heterogeneous landscapes.

ACKNOWLEDGMENTS

This study was supported by the Fundo de Apoio ao Ensino, à Pesquisa e à Extensão (FAEPEX) of the State University of Campinas (UNICAMP) (Grant No. 3111/23); and by the São Paulo Research Foundation (FAPESP) (Grant No. 2025/16161-0). We acknowledge FAEPEX for awarding a master's scholarship to the first author.

REFERENCES

Mateus Domingos, Guilherme Palermo Coelho, Edson Wendland, and Murilo Cesar Lucas. Overcoming spectral confusion: A fast, transferable index for flood extent mapping in urban areas. *SSRN Preprint*, 2026. doi: 10.2139/ssrn.6234146. Under review at Computers & Geosciences.

David Doxaran, Jean-Marie Froidefond, Samantha Lavender, and Patrice Castaing. Spectral signature of highly turbid waters. *Remote Sensing of Environment*, 81:149–161, 7 2002. ISSN 00344257. doi: 10.1016/S0034-4257(01)00341-8.

European Parliament and Council of the European Union. Regulation (EU) No 377/2014 of the European Parliament and of the Council of 3 April 2014 establishing the Copernicus Programme and repealing Regulation (EU) No 911/2010, 2014. URL `https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32014R0377`. Official Journal of the European Union L 122/44. Accessed: 2026-04-15.

Gudina L. Feyisa, Henrik Meilby, Rasmus Fensholt, and Simon R. Proud. Automated water extraction index: A new technique for surface water mapping using landsat imagery. *Remote Sensing of Environment*, 140:23–35, 1 2014. ISSN 00344257. doi: 10.1016/j.rse.2013.08.029.

Jinyun Guo, Yujie Luan, Zhen Li, Xin Liu, Chengming Li, and Xiaotao Chang. Mozambique flood (2019) caused by tropical cyclone idai monitored from sentinel-1 and sentinel-2 images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:8761–8772, 2021. ISSN 1939-1404. doi: 10.1109/JSTARS.2021.3107279.

Chunyan Huang, Mingyang Wang, Zichao Zhu, and Yanling Li. High-resolution remote sensing imagery water body extraction using a u-net with cross-layer multi-scale attention fusion. *Sensors*, 25(18):5655, 2025.

S. K. McFEETERS. The use of the normalized difference water index (ndwi) in the delineation of open water features. *International Journal of Remote Sensing*, 17:1425–1432, 5 1996. ISSN 0143-1161. doi: 10.1080/01431169608948714.

Jean-François Pekel, Andrew Cottam, Noel Gorelick, and Alan S. Belward. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540:418–422, 12 2016. ISSN 0028-0836. doi: 10.1038/nature20584.

Angelica Tarpanelli, Alessandro C. Mondini, and Stefania Camici. Effectiveness of sentinel-1 and sentinel-2 for flood detection assessment in europe. *Natural Hazards and Earth System Sciences*, 22:2473–2489, 8 2022. ISSN 1684-9981. doi: 10.5194/nhess-22-2473-2022.

Stefan Voigt, Fabio Giulio-Tonolo, Josh Lyons, Jan Kučera, Brenda Jones, Tobias Schneiderhan, Gabriel Platzeck, Kazuya Kaku, Manzul Kumar Hazarika, Lorant Czaran, Suju Li, Wendi Pedersen, Godstime Kadiri James, Catherine Proy, Denis Macharia Muthike, Jerome Bequignon, and Debarati Guha-Sapir. Global trends in satellite-based emergency mapping. *Science*, 353:247–252, 7 2016. ISSN 0036-8075. doi: 10.1126/science.aad8728.

Jirapa Vongkusolkit, Bo Peng, Meiliu Wu, Qunying Huang, and Christian G. Andresen. Near real-time flood mapping with weakly supervised machine learning. *Remote Sensing*, 15:3263, 6 2023. ISSN 2072-4292. doi: 10.3390/rs15133263.