remote sensing

MDPI

Review

Reliability and Transferability of Marine Harmful Algal Bloom Detection Using Ocean-Color Satellite Observations: From Optical Signals to Trustworthy Products

Jianhua Zhu [1,2,*,†], Zhen Tian [1,2,†], Zeen Lu [1], Weiwei Li [1], Jingmin Han [1], Xiangtao Zhao [1] and Huimin Zou [1,3]

[1] National Ocean Technology Center, Tianjin 300112, China; tjzhen451076@126.com (Z.T.); luzeen1125@163.com (Z.L.); rikki0909@163.com (W.L.); 18512269860@163.com (J.H.); zhaoxiangtao2020@163.com (X.Z.); hmzou1349@163.com (H.Z.)
[2] Key Laboratory of Ocean Observation Technology, Ministry of Natural Resources, Tianjin 300112, China
[3] Key Laboratory of Coastal Salt Marsh Ecosystems and Resources, Ministry of Natural Resources, Nanjing 210001, China
* Correspondence: besmile@263.net
† These authors contributed equally to this work.

Highlights

What are the main findings?

- Satellite optical observations can support bloom presence and optical phenotypes; more specific biological claims require correspondingly specific independent reference evidence.
- Among 28 purposively prioritized core studies, target-domain applicability was demonstrated in selected cases, but none quantified fixed-configuration source-to-target performance retention under matched tasks and metrics.

What are the implications of the main findings?

- Validation should be matched to the intended biological claim and explicitly test relevant shifts in region, sensor, time, and optical environment.
- Trustworthy HAB products should report uncertainty, applicability limits, and failure conditions together with detection results.

Abstract

Satellite ocean-color remote sensing is used to detect harmful algal blooms (HABs), yet detection capability does not necessarily ensure reliable interpretation or transferable performance. If these limitations are overlooked when developing cross-regional or global automated HAB early-warning systems, biological claims may be overstated and regional changes in false-alarm or missed-detection rates may go unnoticed, potentially compromising warning reliability and related management decisions. We develop an evidence-constrained framework to evaluate both dimensions: reliability, defined by whether the observation–processing–inference chain supports the intended biological claim, and transferability, defined by whether task-relevant performance is retained under changes in time, region, sensor, or optical environment. A structured Web of Science search yielded 1212 unique records, from which 28 purposively prioritized full-text core studies underwent coding for validation design, cross-domain evidence, error, and uncertainty. The synthesis identifies an evidentiary gap between optical detection capability and biological claim specificity, while fixed-configuration cross-domain evidence was scarce and performance retention remained unquantified. Satellite observations can support bloom presence, optical phenotypes, and selected taxon-targeted applications, but specific claims require specific

Check for updates

Academic Editors: Timo Balz and Walter Musakwa

Received: 30 September 2026
Revised: 29 October 2026
Accepted: 4 November 2026
Published: 6 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

independent biological references. Reliability is constrained by reference and labeling practices, preprocessing, observation availability, and incomplete uncertainty characterization, while evidence for transferability is limited by insufficient source–target evaluation. We propose a framework linking claim–evidence alignment, independent validation, domain-shift assessment, and uncertainty reporting. Reliable and transferable HAB remote sensing should move beyond detection outputs toward evidence-supported, uncertainty-aware products for monitoring, early warning, and decision-making.

Keywords: harmful algal blooms; ocean-color remote sensing; satellite-based detection; validation and transferability; uncertainty quantification; reliability assessment

1. Introduction

Marine harmful algal blooms (HABs) threaten coastal ecosystems, aquaculture, fisheries, tourism, and public health, creating a need for timely and sustained monitoring [1]. Field observations provide biologically specific information through microscopy, cell counts, pigment analyses, and toxin measurements, but their spatial and temporal coverage is limited relative to the scale and variability of coastal seas. Ocean-color satellites complement these observations by providing repeated, synoptic measurements of optical signals related to phytoplankton biomass, pigment characteristics, and bloom structure, while supporting bloom delineation and targeted sampling [2–4]. Field measurements therefore provide biological specificity, whereas satellites provide broader spatial and temporal context. Satellite observations should accordingly be interpreted as optical evidence that requires appropriate biological references to support HAB-related conclusions.

The reliability of satellite HAB detection concerns whether satellite-derived optical evidence adequately supports an intended biological claim under specified observation and environmental conditions. Reliability therefore depends on the alignment among the claim, the optical information available from the observation and processing chain, and the reference evidence used to test it. As biological specificity increases, the required reference evidence must become correspondingly more specific. Optical anomalies or elevated biomass may support bloom presence, whereas taxonomic discrimination requires biological evidence linking the signal to a target group, and toxicity inference requires toxin-specific measurements because species presence alone does not establish toxin production [5–7]. Optical proxies and measured toxin concentrations may follow different temporal patterns, while satellite product windows and field sampling schedules can further complicate their comparison [8]. For toxin-related claims, validation should specify the toxin endpoint and justify temporal matching among satellite product windows, toxin-sampling times, and the target state or event window. Reported performance metrics are therefore meaningful only in relation to the stated claim, reference evidence, observation and processing conditions, and task-specific errors. Reliability is thus conditional on the setting in which performance is assessed; whether that performance is retained when those conditions change is a question of transferability.

Transferability concerns whether a detection method retains task-relevant performance beyond the conditions under which it was developed. We refer to the development setting as the source domain and the new application setting as the target domain. Shifts in time, region, water optical type, or sensor can alter the target's optical expression, background confounders, and the information available to the method. As geostationary, high-resolution, and hyperspectral observations expand and machine-learning, segmentation, and hybrid methods are deployed across sensors, regions, and applications [9–13], evaluation under new conditions becomes a practical requirement. Transferability assessment should dis-

tinguish three questions. Target-domain performance establishes applicability under the new conditions. Performance retention requires matched source–target comparisons using comparable task definitions and metrics. When target-domain adaptation or recalibration is introduced, adaptation gain should be evaluated by comparing fixed and adapted configurations on the same target test data. These distinctions separate target-domain applicability, performance retention, and adaptation gain as different forms of evidence.

Existing reviews have synthesized satellite bloom-detection methods, ocean-color phytoplankton products, and uncertainty in coastal water-quality products [7,9,14,15], but these topics are often treated separately. Less attention has been given to whether evidence across the observation–processing–inference–claim chain is sufficient for the intended monitoring use. Rather than organizing the literature by algorithm class or headline accuracy, this review evaluates reported capabilities together with reference evidence, validation design, and application conditions, linking methodological performance to the monitoring conclusions the evidence can defensibly support.

This review combines a structured database search, a broader literature synthesis, and standardized coding of purposively prioritized full-text studies. The search provides an auditable literature base, while the broader synthesis and coded subset support interpretation and comparison of sensors, algorithms, reference evidence, validation design, error, and cross-domain applications. Reliability is evaluated along the observation–processing–inference–claim chain, whereas evidence for transferability is appraised according to whether the reviewed studies provide comparable source–target performance evaluations. We then propose an integrated evaluation framework and minimum reporting requirements to support method selection, field-validation design, and interpretation of satellite HAB products.

2. Review Methodology and Evidence Coding

2.1. Review Design and Information Sources

This focused critical review used three complementary evidence sources: an initial scoping library, a structured database corpus, and a prioritized full-text evidence set. The scoping library, assembled from author bibliographies, Crossref, PubMed, and OpenAlex, informed search terminology, sensor and taxon coverage, and background synthesis. The formal search was conducted in the Web of Science Core Collection on 30 August 2026, covering the database's full available timespan from inception to the search date. This corpus underpinned title-and-abstract screening, candidate identification, and evidence prioritization, and all structured-screening counts refer to it. Scopus was not included, and this coverage limitation is addressed in Section 2.5. Studies with informative evidence on validation, cross-domain application, error, or uncertainty were prioritized for full-text assessment. Retained core studies then underwent standardized coding for study tasks, observation and processing conditions, biological references, validation designs, reported errors, and cross-domain applications. Details of the search, screening, coding, and limitations are provided in Sections 2.2–2.5.

2.2. Structured Search and Record Management

Four complementary searches targeted marine HAB satellite detection and evidence relevant to reliability and transferability. They addressed: (1) satellite detection of HABs and red tides; (2) named bloom-forming taxa and historical synonyms; (3) atmospheric correction, error, and uncertainty; and (4) transferability, generalization, transfer learning, domain adaptation, and external validation. Because these searches could also retrieve freshwater studies, marine or coastal eligibility was assessed during screening. Full search strings, settings, and returned counts are provided in Supplementary Methods S1.

The four searches returned 839, 485, 256, and 59 records, respectively, totaling 1639 before deduplication. Deduplication by Web of Science accession number removed 427 duplicates, leaving 1212 unique records. DOI and normalized-title matching were used to reconcile these records with the scoping library and maintain consistent record identification. These 1212 records formed the corpus for title-and-abstract screening.

The search, screening, and focused evidence-synthesis workflow is summarized in Figure 1.

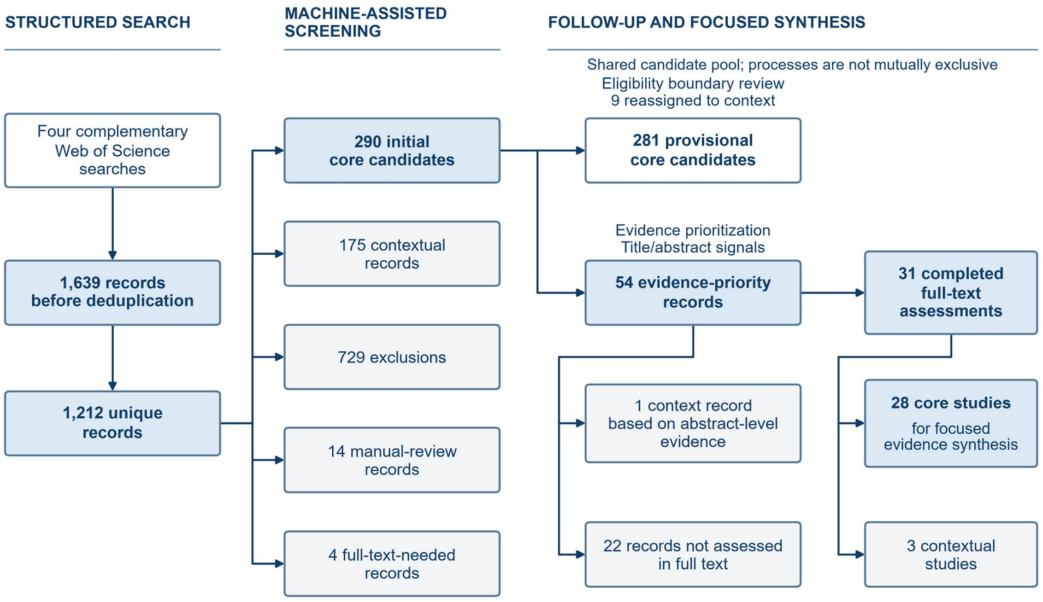


Figure 1. Structured search, screening, and focused evidence-synthesis workflow. The structured search yielded 1212 unique records after deduplication, including 290 initial core candidates. Two non-sequential follow-up processes originated from this candidate pool: eligibility boundary review reassigned nine records to context, leaving 281 provisional core candidates, while evidence prioritization identified 54 records based on title- and abstract-level indications of validation, cross-domain application or comparison, error, or uncertainty. These indicators guided the identification of potentially informative studies rather than serving as mandatory quantitative inclusion thresholds. Within the focused synthesis, 31 records underwent full-text assessment, yielding 28 core studies retained for standardized coding and three contextual studies. One additional record was reassigned to context based on abstract-level evidence, while 22 records were not assessed in full text. The resulting 28-study corpus forms the basis for the descriptive evidence summaries, with the scope limitations discussed in Section 2.5.

2.3. Eligibility and Machine-Assisted Screening

A core candidate had to satisfy four criteria: (1) marine, coastal, estuarine, or brackish transitional waters; (2) satellite ocean-color, multispectral, or hyperspectral optical observations; (3) a HAB, red tide, bloom-forming taxon, or explicitly HAB-relevant optical target; and (4) a detection, mapping, classification, segmentation, discrimination, retrieval, or relevant validation task. Freshwater-only studies, non-satellite optical platforms, synthetic-aperture radar, forecasting without satellite detection assessment, and studies focused on governance, mitigation, or ecological impacts were excluded. Atmospheric-correction, chlorophyll, sensor, and observation-system studies were retained as contextual evidence when they directly informed HAB-related error, failure mechanisms, or uncertainty.

All 1212 unique records underwent machine-assisted title-and-abstract screening. The full record set was then re-screened using the stricter four-criterion codebook, with programmatic checks ensuring that supporting evidence fragments were traceable to the source title or abstract. This produced 290 initial core candidates, 175 contextual records, 729 ex-

clusions, 14 manual-review records, and four requiring full text for a decision. Author review of discrepant and boundary cases reassigned nine of the 290 core candidates to context, leaving 281 provisional core candidates. These represent a candidate set rather than a final full-text-eligible corpus. Further workflow and archival details are provided in Supplementary Methods S2.

2.4. Full-Text Evidence Coding and Synthesis

Among the 290 initial core candidates, 54 records were identified for evidence-priority follow-up based on title- and abstract-level indications of validation, cross-domain application or comparison, error, or uncertainty. These indicators guided the identification of potentially informative studies rather than serving as mandatory quantitative inclusion thresholds. In particular, explicit cross-domain validation and quantitative error statistics were not required for every retained study.

The review adopted a focused, non-exhaustive approach to full-text evidence synthesis. Full-text assessment was completed for 31 records, of which 28 met the core-study scope defined in Section 2.3 and were retained for standardized coding. The other three were classified as contextual studies because their primary tasks fell outside that scope. The resulting 28-study corpus was used to compare validation designs, cross-domain evidence, and error and uncertainty reporting. Its frequency summaries characterize this focused corpus rather than the wider candidate literature, as further explained in Section 2.5. Supplementary Methods S2 provides the recorded prioritization signals and study dispositions.

Studies were coded for region and water optical type, target organism, detection task, sensor/product, preprocessing and quality control, input features and algorithm family, biological references and labels, validation design, reported metrics and errors, failure conditions, and source-location anchors. Numerical values were recorded only when directly reported and traceable to a page, table, or figure. For synthesis, each study received one task-framing class, one primary observation architecture, and a descriptive C1–C5 code representing the most specific documented task/reference endpoint. This empirical coding describes the evidence reported in the reviewed literature and is distinct from the normative L1–L5 claim–evidence hierarchy introduced in Section 3.1.

Validation evidence was coded at two levels: one primary validation display class and a set of nonexclusive validation/reference elements preserving test units, biological references, and external-application features. These relationships are illustrated by an alluvial map in Section 7.5. Its alluvial pathways are descriptive and do not imply causality, temporal sequence, validation strength, or method superiority. Category definitions are provided in Supplementary Methods S3; study-level assignments and concise source anchors are provided in Supplementary Table S1 and Supplementary Data S1.

Cross-domain evidence was characterized by six nonexclusive shift descriptors—time, region, sensor, water optical type, species, and pigment/optical phenotype—and five mutually exclusive evidence states: strict fixed configuration, qualified fixed independent-target evaluation, target-informed or adapted evaluation, held-out non-cross-domain evaluation, and no qualifying cross-domain evidence. A fixed configuration required the same decision layer (threshold, rule, or trained model) to be applied to independent target observations without target-label fitting or target-test selection, with a comparable target-domain metric reported.The descriptors represented among qualifying cross-domain applications are summarized in Section 7.5; detailed decision rules and boundary cases are provided in Supplementary Methods S3.

Error evidence was assigned to four mutually exclusive categories: quantitative class-conditional or confusion-derived evidence, other quantitative error or uncertainty, qualitative failure conditions only, or no relevant error reporting. Their distribution is summarized

in Section 8.2. Composite domain shifts were retained as such rather than interpreted as isolated single-axis effects. Coding rules and decision criteria are provided in Supplementary Methods S3, study-level assignments in Supplementary Table S1, and machine-readable coding fields and concise source anchors in Supplementary Data S1.

2.5. Evidence Interpretation and Review Limitations

Evidence was interpreted according to how directly it supported, contextualized, or constrained the claim under consideration rather than treating different evidence types as equivalent. Numerical results and methodological details were used only when supported by the available full text. Because reported performance depends on task definition, class prevalence, evaluation unit, sensor, and validation design, heterogeneous studies were not pooled or treated as directly comparable.

The structured search was limited to Web of Science, and title-and-abstract screening was not independently duplicated. The standardized synthesis was based on 28 core studies retained through the focused, non-exhaustive evidence-prioritization and full-text assessment process described in Section 2.4, rather than a random or exhaustive sample of the broader candidate literature. Frequencies of validation classes, cross-domain evidence states, descriptive task/reference classes, and error evidence therefore characterize this focused corpus and should not be interpreted as prevalence estimates for the wider candidate pool or the literature. These scope conditions limit statistical generalization of the frequency summaries but do not prevent the coded studies from revealing evidence patterns and methodological gaps within the reviewed corpus or informing future evaluation requirements.

3. Optical Basis and the Hierarchy of HAB Detectability

3.1. What Does an Ocean-Color Satellite Actually Detect?

Ocean-color satellites first measure spectral radiance reaching the sensor rather than directly measuring species identity or toxicity. Radiometric calibration and atmospheric correction are used to derive water-color products such as remote-sensing reflectance (R_{rs}), from which spectral features, pigment-related indices, chlorophyll estimates, and other bio-optical variables can be obtained. HAB interpretation therefore involves a sequence of inference from sensor measurement to corrected optical information and then to biological meaning. Optical signals may support inference of biomass, bloom state, or other biological properties, but increasingly specific claims require increasingly specific independent reference evidence. Figure 2 summarizes this relationship as a normative claim–evidence hierarchy extending from bloom presence to toxicity inference.

Different biological claims require different levels of reference specificity. Optical anomalies, elevated chlorophyll, or high-biomass signals may support bloom presence or surface accumulation, whereas attribution to a HAB-related group requires biological or taxonomic evidence linking the observed optical expression to the target organism. Species-level discrimination requires still more specific reference information. Identifying a potentially toxigenic taxon can inform toxin-related risk assessment, but species dominance alone does not establish toxin occurrence or concentration, and harmful taxa need not be dominant. Toxin production may vary among strains and with physiological or environmental conditions [3,4]. Indirect models can estimate the probability of defined toxin events when developed and evaluated against appropriate toxin-specific reference measurements; these outputs are predictions rather than direct measurements, and their performance remains context-dependent [16]. Claims about toxin occurrence or concentration require corresponding toxin-specific evidence. Attributing toxin production to a particular taxon or strain additionally requires evidence linking that source to the toxin.

Evidence of toxin occurrence alone does not establish ecological harm, which also depends on exposure and the affected organisms; HAB impacts may additionally arise through non-toxin mechanisms [4].

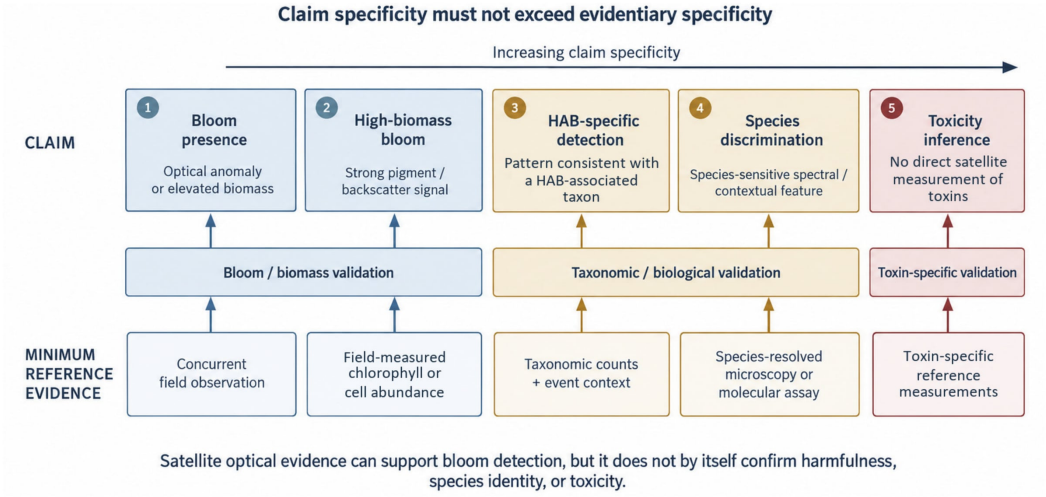


Figure 2. Normative claim–evidence hierarchy for satellite-based marine HAB detection. Claim specificity increases from bloom presence (L1) to toxicity inference (L5), with progressively more specific independent reference evidence required at each level. Bloom/biomass validation supports L1–L2 claims, and taxonomic or biological validation is required for L3–L4. Toxin-specific reference measurements are required to substantiate or validate L5 claims about toxin occurrence, concentration, or the probability of a defined toxin event. Attributing toxin production to a particular taxon or strain additionally requires evidence linking that source to the toxin. The reference evidence shown represents the minimum evidentiary specificity needed to defend each claim rather than a complete or exclusive validation protocol. Satellite optical evidence alone does not establish harmfulness, species identity, or toxicity. Descriptive study-level coding of the evidence reported in the reviewed literature is documented separately in Supplementary Methods S3.

Obtaining frequent, spatially representative toxin reference measurements across broad marine areas is constrained by sampling coverage, differences between the near-surface optical signal and the depth and matrix of collected samples, and the analytical and quality-control requirements of toxin assays [4,17,18]. Satellite-guided targeted sampling, coordinated fixed and autonomous observations, and toxin-specific laboratory or validated portable assays can provide complementary information, but optical and non-toxin biological measurements cannot substitute for toxin-specific reference evidence [2–4,17,18]. In Portuguese coastal waters, satellite chlorophyll observations were examined alongside biotoxin concentrations measured in bivalve tissues [8]. Such measurements provide toxin evidence for the sampled tissue matrix, rather than simultaneous toxin concentrations in seawater. The central principle is therefore that claim specificity should not exceed evidentiary specificity. The tasks and reference evidence documented in the reviewed studies are coded separately in Supplementary Methods S3 rather than treated as part of this normative hierarchy.

3.2. Spectral Expressions and Their Conditional Meaning

Phytoplankton alter water color through pigment absorption, cellular and detrital scattering, chlorophyll fluorescence, and other optically active constituents. Foundational ocean-color studies established relationships among absorption, scattering, water-leaving radiance, and pigment-sensitive spectral structure [19,20]. These processes modify reflectance magnitude and shape: pigment absorption generally suppresses blue reflectance, whereas high-biomass conditions can enhance green, red, or near-infrared signals through

the combined effects of absorption, scattering, and fluorescence. Solar-stimulated chlorophyll fluorescence and red-region reflectance peaks provide additional information when suitable bands are available [21–23]. Blue–green ratios, red-edge indices, fluorescence line height, chlorophyll anomalies, backscattering–chlorophyll relationships, and multiband spectral shapes translate these mechanisms into detectable bloom features.

The biological meaning of these features, however, depends on environmental and observational context. Their expression varies with phytoplankton physiology, background optical constituents, retrieval quality, and bloom structure. Photoacclimation can alter cellular pigment content and composition, while pigment packaging reduces chlorophyll-specific absorption through intracellular self-shading, with the strength of this packaging effect depending on cell size and intracellular pigment concentration [24]. Cellular pigment content also varied during the decline of a natural Karenia brevis bloom [25]. Fluorescence depends on chlorophyll concentration, physiological state, light history, and fluorescence efficiency, while suspended sediment and imperfect atmospheric correction can modify red and near-infrared signals. Chlorophyll anomalies indicate departures from regional background rather than species identity [26], and floating-algae indices preferentially detect surface accumulations rather than dispersed or subsurface blooms [27]. Thus, a physically interpretable spectral feature is not necessarily biologically specific across environments.

3.3. Species, Optical Phenotype, and Detectability

Taxon- or species-level discrimination becomes plausible only when a bloom exhibits a sufficiently distinguishable spectral or spatial expression relative to its background water under the conditions of application. Such an optical phenotype is not a context-independent spectral fingerprint, but emerges from interactions among pigment composition, cellular and physiological properties, aggregation, background constituents, sensor configuration, and processing. Even within the same taxon, physiological or growth-state variation can modify its optical expression [24,25]; laboratory absorption-based discrimination has also shown sensitivity to the light-acclimation conditions represented in the analysis [28]. Accordingly, spectral separability should be evaluated across physiological states and optical backgrounds representative of the intended application. Detectability and taxonomic attribution must therefore be distinguished: the former asks whether a bloom signal can be separated from its background, whereas the latter additionally requires independent biological evidence linking that signal to a particular taxon or species. Reviews of satellite phytoplankton functional-type products likewise show that optical-group discrimination depends on the algorithmic target, spectral information, and the ecological and optical conditions represented during validation [14]. Karenia brevis studies, for example, have used chlorophyll anomalies, fluorescence, and backscattering–chlorophyll relationships, while showing that their utility remains region- and condition-dependent [29,30].

Regional algorithms for Cochlodinium polykrikoides, Prorocentrum donghaiense, Noctiluca scintillans, and other bloom-forming taxa similarly exploit characteristic spectral or spatial expressions within defined optical backgrounds. A GOCI red-to-blue-ratio model achieved $R^2 = 0.92$ ($n = 13$) during one East Sea event, yet also showed a mean normalized bias of −44% and a mean normalized gross error of 63.5% [31]. Thus, high correlation can capture relative variation without implying low systematic bias or high absolute accuracy. Species- or taxon-level detectability is therefore conditional on whether the optical phenotype remains separable from confounders under the sensor, environmental, and validation conditions of application.

3.4. Optical Confounders and Observation Mismatch

Because optical separability depends on the surrounding environment, coastal HAB detection operates in a feature space shared by multiple water constituents, environmental structures, and biological targets. Suspended particles, colored dissolved organic matter (CDOM), river plumes, resuspension, shallow-bottom reflectance, and sun glint can all alter spectral or spatial patterns, while non-harmful phytoplankton blooms and floating macroalgae may closely resemble HABs. Spectral or spatial similarity alone is therefore insufficient to establish HAB specificity; reliable discrimination requires that the target remain separable from relevant confounders.

These confounders can generate both false positives and false negatives. Sediment or correction residuals may inflate chlorophyll anomalies, regional thresholds may fail when optical backgrounds change, and spatial models may learn local background or annotation patterns that do not transfer. A different failure occurs when the target is biologically present but lacks a satellite-observable surface expression. During a toxic Pseudo-nitzschia event in southern California, MODIS characterized near-surface chlorophyll-a, whereas autonomous gliders and molecular samplers revealed subsurface population structure before a visible surface bloom developed [17]. The absence of a detectable surface signal is therefore not equivalent to biological absence. Validation should represent the optical, spatial, vertical, and temporal conditions expected in deployment so that tests expose the conditions producing false positives, false negatives, and observation mismatch. These limitations are further shaped by sensor sampling and preprocessing, discussed in Section 4.

4. Ocean-Color Observations and Preprocessing

4.1. Sensor Capability Is a Multi-Dimensional Trade-Off

The optical confounding and observation mismatch discussed in Section 3 are partly shaped by the spatial, temporal, and spectral sampling of the observing system. No single sensor simultaneously maximizes spatial resolution, temporal frequency, spectral detail, signal-to-noise ratio, swath, and long-term continuity; satellite HAB monitoring therefore requires trade-offs among these dimensions. From CZCS onward, ocean-color missions have progressively improved radiometric and spectral capabilities [15,32,33]. Polar-orbiting sensors such as MODIS, VIIRS, and OLCI provide repeated regional coverage and long archives, whereas GOCI-class geostationary sensors resolve intraday bloom variability [34], including documented diurnal evolution in the East China Sea [10]. Higher-spatial-resolution imagers better resolve narrow coastal waters and bloom boundaries, while hyperspectral observations provide richer information on pigments and optical phenotypes, albeit with greater demands on calibration, atmospheric correction, signal-to-noise control, and field validation [11,35,36].

Near the Guadiana estuary, Sentinel-2 MSI depicted small bloom patches less clearly represented in Sentinel-3 OLCI imagery, with field samples identifying *Lingulodinium polyedra* [6]. GF-1 WFV broad-band indices have also mapped narrow *Noctiluca scintillans* red-tide streaks [37,38]. A pseudo-hue-angle index developed using GF-1 data was subsequently applied to HY-1D CZI, Landsat-8 OLI, and Sentinel-2 MSI observations [38]. These studies illustrate the value of fine-scale spatial information, while spectral configuration, atmospheric correction, sunglint, and usable observation dates constrain its application [6,37,38].

Sensor selection should therefore be driven by the monitoring task and the biological specificity of the intended claim rather than by maximizing any single resolution dimension. Different applications require different combinations of spatial, temporal, and spectral information, from regional high-biomass monitoring to small-patch delineation and taxon-targeted discrimination. Higher spatial, temporal, or spectral resolution does not by itself make a HAB product more reliable; the relevant information must be appropriate to the

intended claim and supported by corresponding reference evidence. These trade-offs also make cross-sensor application non-trivial, because nominally similar bands can differ in spectral response, signal-to-noise characteristics, spatial mixing, and processing history, effectively creating a new observation domain rather than a simple sensor substitution. The implications of these sensor trade-offs for HAB monitoring are summarized in Table 1.

Table 1. Observation-system trade-offs and evaluation implications for marine HAB detection.

Observation System	Principal Observational Advantage	Monitoring Roles Best Supported	Dominant Limitation	Key Evaluation Implication
Polar-orbiting ocean color (e.g., MODIS, VIIRS, OLCI)	Broad repeated coverage and long observational archives	Regional bloom presence, high-biomass events, long-term variability, and selected regional HAB indicators	Coastal pixel mixing, cloud cover, and limited taxon-sensitive bands	Long archives do not by themselves ensure temporal or regional consistency; processor version, regional background, and product continuity should be documented
Geostationary ocean color (GOCI class)	High-frequency intraday sampling of bloom evolution and transport	Event occurrence, intraday dynamics, and evolution of high-biomass surface blooms	Regional coverage, cloud/glint, and moderate spatial resolution	Dense temporal sampling strengthens process-level evidence but does not replace independent target-domain evaluation in a new region
High-spatial-resolution multispectral imaging (e.g., Sentinel-2, Landsat, GF, HY-CZI)	Fine delineation of narrow coastal waters, bloom patches, and boundaries	Small blooms, aquaculture-adjacent monitoring, patch segmentation, and boundary delineation	Lower revisit frequency and fewer dedicated ocean-color bands	Models may exploit coastline, morphology, or regional background; application across coasts or water types requires renewed evaluation
Hyperspectral ocean color (e.g., PACE OCI and related imaging spectroscopy)	Dense spectral sampling of pigments and subtle optical phenotypes	Optical-phenotype discrimination, pigment or functional-group analysis, and testable taxon hypotheses	Coastal atmospheric correction, spatial mixing, signal-to-noise ratio (SNR), and demanding field validation	Additional bands support more specific inference only when the biological specificity of the reference evidence matches the intended claim
Complementary in situ and autonomous observations	Biological specificity, vertical structure, and continuous local environmental context	Species confirmation, toxin measurement, label definition, and satellite-product validation	Sparse spatial coverage and spatial–temporal matchup mismatch	Provide independent reference evidence for satellite claims and characterize target-domain conditions and observation mismatch

4.2. Atmospheric Correction Is Part of the Detection Model

Ocean-color satellites measure top-of-atmosphere radiance, so interpretation of the aquatic signal depends critically on atmospheric correction. Classical open-ocean schemes

often assume negligible near-infrared water-leaving radiance [39,40], but suspended particles and high-biomass blooms can violate this assumption in optically complex coastal waters. Coastal processing has therefore adopted turbid-water similarity approaches, Near-Infrared to Short-Wave Infrared (NIR–SWIR) combinations, Case-2 processors, and learning-based methods [41–44]. Residual aerosol mismatch, adjacency effects, and bright pixels can still bias retrieved reflectance, and because many HAB indices rely on small spectral differences among neighboring bands, such biases can alter whether a pixel crosses the bloom-detection decision boundary. In a GF-1 scene, artificial perturbations to input remote-sensing reflectance changed the detection accuracy of a pseudo-hue-angle red-tide method [38].

Atmospheric correction should therefore be treated as part of the HAB detection chain rather than as generic preprocessing. Atmospheric and surface contributions can account for approximately 90–99% of the top-of-atmosphere signal over water, making even modest residual errors large relative to the water-leaving component [7]. Downstream chlorophyll-product discrepancies can also reflect coastal optical complexity. For a nearshore transect near Charlotte Harbor on 19 October 2004, MODIS OC3M band-ratio chlorophyll estimates were reported to be 3–15 times the corresponding in situ values; the authors discussed possible CDOM interference and shallow-bottom reflectance without quantifying atmospheric-correction effects [45]. A GOCI red-tide study used aerosol estimates from nearby non-bloom waters before retrieving chlorophyll, which was evaluated against field measurements [31]. In other coastal GOCI and HY-1C studies, quantitative validation focused on retrieved remote-sensing reflectance [46,47]. Correction uncertainty can thus propagate through the reflectance–feature–decision chain, so validating an intermediate reflectance product does not by itself establish the reliability of the final HAB product.

Evaluation should address the HAB output generated by the stated atmospheric-correction, retrieval, and decision workflow as an integrated chain. Where multiple correction processors are plausible, sensitivity of the final HAB product to processor choice is more informative than performance of a single intermediate product, while intermediate reflectance quality should still be documented. Specific reporting requirements are provided in Section 8.3.

4.3. Clouds, Glint, Adjacency, and Observation Opportunity

Quality control determines not only which pixels are retained but also which environmental and event conditions a satellite product represents. Cloud and sun-glint masking reduce valid observations, while adjacency contamination is strongest near bright land, shallow features, and other high-reflectance coastal targets. Stricter masks can suppress false detections but may also remove bloom edges or bright bloom pixels, whereas permissive masks increase coverage at the cost of greater contamination. In a GOCI diurnal study, standard processing excluded some cloud-free coastal-water observations. The alternative workflow combined normalized Rayleigh-corrected reflectance with revised cloud masking and detection thresholds [10]. Quality control therefore involves a trade-off between error suppression and observation opportunity.

Observation gaps can depend on environmental and processing conditions [6,10,46]. If training and validation archives are dominated by clear-sky, low-glint, well-corrected scenes, reported performance mainly characterizes those observable conditions rather than the full range of bloom conditions. Evaluation should therefore report the proportion of observations retained, the pattern of missing data, and whether known reference events lacked usable satellite coverage. High conditional accuracy may otherwise coexist with a low probability of producing a usable product when a bloom occurs. Operational reliability thus depends on both performance under valid observations and the opportunity to obtain

those observations during relevant events. Multi-sensor observations can increase such opportunities but introduce cross-sensor consistency and harmonization requirements, considered in Section 4.4.

4.4. Cross-Sensor Harmonization

Combining sensors can increase observation opportunities; sufficient comparability is required when merging products or applying common detection features across sensors. Nominally similar bands or products are not necessarily equivalent because sensors differ in spectral response, radiometric calibration, signal-to-noise characteristics, spatial support, acquisition timing, and preprocessing. Harmonization may therefore involve spectral-response adjustment, spatial and temporal matching, inter-sensor bias correction, and consistent quality control. A recent spatial–spectral collaborative method cross-calibrated HY-1D COCTS radiances against Aqua MODIS radiance products, illustrating a feature-based route toward inter-sensor radiometric consistency [48]. A Mediterranean multi-sensor processing system combined band shifting and inter-sensor bias correction before merging observations, then evaluated reflectance and derived ocean-color products against field measurements [49]. Related studies have addressed product consistency and calibration across sensors and historical records [50–52]. Harmonization should thus reduce observation-system differences to a documented level of comparability rather than assume sensor interchangeability.

Multi-sensor use, cross-sensor product consistency, and model transferability address different questions. Multi-sensor use broadens coverage; cross-sensor consistency asks whether harmonized products are quantitatively comparable; model transferability asks whether a threshold, rule, or trained model retains task-relevant performance in an independent target-sensor domain. Target-domain evaluation can establish applicability under new observation conditions, but performance retention requires matched source–target comparisons using comparable tasks and metrics. Harmonization can support the comparability required when merging products or applying common detection features across sensors, but does not itself demonstrate HAB model transferability.

5. HAB Detection Algorithms

Figure 3 compares the characteristic capabilities, key conditions, and most informative evaluations across the five method families.

5.1. Empirical and Semi-Empirical Indices

Empirical and semi-empirical methods detect blooms using a small number of interpretable features such as band ratios or differences, regional anomalies, fluorescence line height, and color-space transformations. Their strengths are transparency, low computational cost, and compatibility with long satellite archives and repeated regional monitoring. Chlorophyll anomalies, fluorescence-based features, and red-tide indices have been applied to Karenia brevis, high-biomass blooms, and red tides in optically complex coastal waters [29,37,38,45,53–55]. In regional K. brevis monitoring, chlorophyll anomalies have been defined relative to preceding background conditions and interpreted alongside field or ancillary observations [29,53]. GF-1 approaches illustrate another design route, from radiance-based broad-band indices to pseudo-hue-angle transformations of green, red, and near-infrared reflectance [37,38]. These approaches remain useful when the target optical phenotype is relatively stable, sensor configuration is well defined, and background conditions are reasonably constrained.

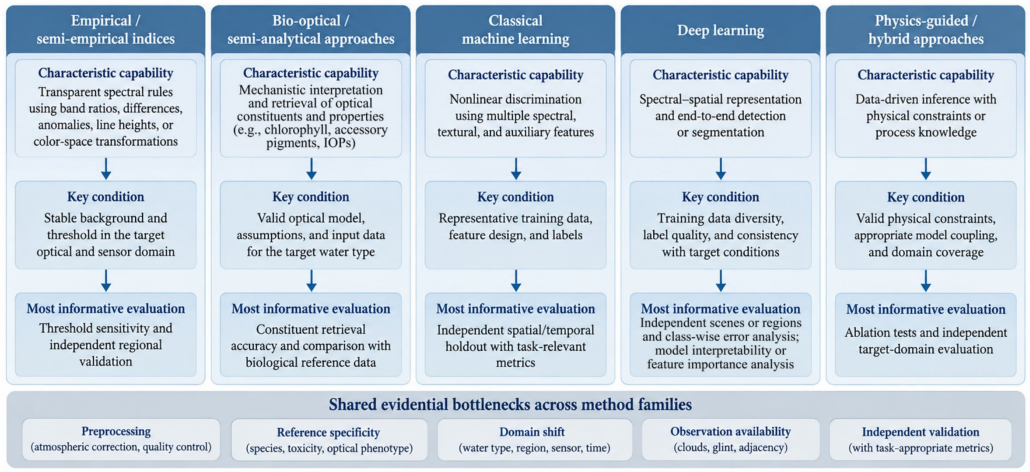


Figure 3. Method families, characteristic capabilities, and evidential requirements in marine HAB detection. Each column summarizes the principal capability contributed by a method family, the condition on which that capability depends, and the evaluation most directly suited to testing that condition. The lower band highlights evidence bottlenecks shared across method families. Column order is organizational and does not indicate chronology, model capacity, or expected performance.

Their main limitation is that fixed thresholds or low-dimensional rules implicitly assume stability in background water, season, and sensor response. The same threshold may therefore map to a different non-bloom distribution when transferred to another water type, region, or sensor. Spectral-index comparisons have also examined red-tide occurrence indicators and turbidity-related confounding [56]. Evaluation should therefore consider threshold sensitivity, false positives in turbid waters, detectable biomass ranges, and documented failure conditions. Independent target-domain references can establish whether the rule remains useful under new conditions; claims of performance retention require comparable source–target evaluation.

5.2. Bio-Optical and Semi-Analytical Approaches

Bio-optical HAB approaches build on chlorophyll algorithms and semi-analytical retrievals of inherent optical properties, which link observed reflectance to optical constituents without directly identifying harmful taxa [57–59]. Semi-analytical retrievals can help distinguish constituent contributions and generate testable hypotheses about bloom-related optical properties. For K. brevis, a SeaWiFS approach combined elevated chlorophyll with particulate backscattering below a reference chlorophyll–backscattering relationship, with field cell counts providing biological reference evidence [30]. In the southern Benguela, red-to-near-infrared line-height indicators were linked to field-observed phytoplankton communities to support regional risk mapping for aquaculture [60]. Mechanistic interpretability, however, does not guarantee biological specificity. Retrievals remain dependent on atmospheric correction, optical-model assumptions, and the target water type, and elevated chlorophyll or a fitted optical property is not itself confirmation of a HAB, species identity, or toxicity. Reliability therefore requires validation of both the optical retrieval and the biological interpretation against reference evidence appropriate to the intended claim.

5.3. Conventional Machine Learning

Support vector machines, random forests, boosting methods, regularized least-squares classifiers, and shallow neural networks can combine multiple spectral and environmental features to represent nonlinear decision boundaries. They are particularly useful for regional HAB problems with modest sample sizes and engineered features. Feature-removal tests can assess how engineered inputs contribute to classification performance. Their

reliability nevertheless depends on representative sampling, class balance, and coverage of the intended application domain.

The critical evaluation issue is whether development and test samples are genuinely separated in space, time, or event identity. Randomly splitting neighboring pixels can place shared background, acquisition, and annotation information in both sets and inflate apparent generalization. A regularized least-squares classifier with a radial basis function (RBF) kernel, using combined MODIS remote-sensing reflectance and normalized water-leaving radiance features, reported 94.37% overall accuracy for Indian-coast bloom imagery [61]. Because the labels were event-derived and testing used pixel-level splits within historical imagery, the result primarily demonstrates within-archive separability rather than independent regional species-classification performance. Scene-, event-, year-, or region-level holdouts provide stronger evidence when they better match the intended deployment setting.

5.4. Deep Learning and Semantic Segmentation

Convolutional and encoder–decoder networks are well suited to HAB mapping because blooms often form spatially coherent objects with boundaries, textures, and multiscale structure. U-Net-type architectures and residual learning provide foundations for these applications [62,63]. Representative HAB studies illustrate several methodological developments: a VIIRS model combined spectral information with spatial coherence for K. brevis detection [12]; RTDNet used residual and receptive-field modules for high-resolution red-tide delineation [64]; and a GOCI network incorporated ConvLSTM to exploit multitemporal information within a bloom event [65]. Spectral-enhancement networks have also incorporated derived optical channels into red-tide segmentation [66]. The VIIRS study reported an image-pair detection F1 of 0.89 on 53 retained image pairs, whereas RTDNet reported pixel-level F1 values near 0.90 on two test sub-images [12,64].

These results demonstrate useful discrimination within their respective evaluation settings, but the scores reflect different reference data and evaluation units. In the VIIRS application, sediment-related obscuration of bloom signals contributed to missed detections [12]. A multi-sensor Res-UNet study trained separate models for six satellite sensors and reported a mean pixel-level F1 of 0.85 across their test evaluations; its separate 81.09% agreement with disaster bulletins was based on event-level spatial and temporal matching [13]. Their evaluations used held-out images, later observations within the same event, or separate test sets for sensor-specific models. Feature-contribution tests and intermediate-feature visualizations can help examine model behaviour when their scope is explicit [64,67]. Input-removal analyses should state whether models were retrained; prediction explanations should be checked for stability across relevant scenes and optical backgrounds. Deep-learning reliability should be assessed through independent evaluation, class-wise errors, and reference quality rather than headline accuracy alone.

5.5. Physics-Guided and Hybrid Methods

Physics-guided and hybrid methods can incorporate optical or process knowledge through input features, network architecture, or explicit physical constraints [68]. In a Sentinel-2 red-tide application, SIG-Net combined a spectral-index branch with a spectral encoder through gated cross-attention [67]. On the scene-held-out test set, gated fusion achieved a higher class-mean F1 than ungated fusion (83.67% versus 81.33%) [67]. Separate ablations examined the contribution of each spectral index [67]. Optical indices have also been incorporated into a Noctiluca network alongside learned spatial features [69]. These applications incorporated optical knowledge through features and architecture, without imposing a radiative-transfer constraint or a physics-based loss. Radiative-transfer constraints

and correction-aware losses are candidate extensions requiring separate application-specific validation. Their contributions should be assessed through ablation and matched comparisons of calibration, label demand, error structure, and target-domain performance.

Despite differences in representation and model complexity, all method families ultimately face two evidential questions: whether their outputs support the intended task under the evaluated conditions, and whether performance is retained when the sensor, region, period, or water type changes. The following section therefore shifts from algorithm families to validation design and comparative performance, where these questions can be tested directly.

6. Validation and Comparative Performance

6.1. Reference Evidence Supports the Scientific Claim

Reference measurements support different biological claims. Chlorophyll-a primarily indicates biomass, microscopy and species counts support taxonomic attribution, molecular methods strengthen species identification, and toxicity inference requires direct toxin measurements. Expert image interpretation mainly supports spatial delineation, while disaster bulletins and event databases provide event-level evidence but may lack pixel-scale correspondence. Operational early-warning systems therefore combine optical, imaging, molecular, and toxin measurements to address different monitoring questions [4,18]. Satellite validation should use reference evidence whose biological specificity matches the claim being evaluated.

The evidentiary value of validation also depends on reference independence and satellite–field matchup design. Information used for label construction, threshold setting, or model selection is not fully independent when reused for testing. Even independent references are affected by scale mismatch between local field samples and larger satellite pixels. Bloom patchiness, vertical migration, advection, and sampling delay can further weaken correspondence. Reporting matchup windows, included and excluded samples, exclusion criteria, and aggregation rules helps distinguish algorithmic error from uncertainty introduced by the reference-matching process [70,71].

6.2. Classification and Mapping Performance Metrics

Overall accuracy can be misleading when non-bloom observations dominate an evaluation set. Classification performance should therefore be reported with class-sensitive measures such as precision, recall/sensitivity, specificity, and F1. For segmentation and mapping, intersection over union (IoU), bloom-area error, and boundary displacement evaluate mask overlap, area differences, and boundary-position differences, respectively. Event-level detection requires a definition of a positive event, a stated evaluation unit, and a rule for matching detections to reference events.

In the satellite–field matchup evaluation of Carvalho et al. [70], reference-positive cases were defined by in situ *K. brevis* counts ≥15,000 cells L^{-1}, while samples below this threshold were treated as reference-negative. Sensitivity was calculated as TP/(TP + FN) among reference-positive matchups, and specificity as TN/(TN + FP) among reference-negative matchups. In pooled summer–autumn matchups from the central west Florida shelf, the initial empirical, bio-optical, and automated chlorophyll-anomaly configurations yielded sensitivity/specificity pairs of 86%/53% (n = 371), 77%/65% (n = 371), and 26%/84% (n = 328), respectively [70]. These results describe aggregate detection and false-positive profiles over the pooled period. The operational method used a smaller valid-matchup subset, so the reported three-method comparison was not fully paired. A related analysis of the same regional matchup archive reported 79% overall accuracy, 80% sensitivity, and 78% specificity for its first hybrid evaluation configuration using 145 reference-positive and

226 reference-negative matchups, with corresponding false-negative and false-positive rates of 20% and 22% [71]. Evaluation should therefore report task-relevant error profiles rather than a single summary metric.

6.3. Validation Designs and the Claims They Support

Validation design should be described along four complementary dimensions. Development–test separation specifies whether pixels, patches, scenes, events, years, regions, or sensors were independent of model development. Reference specificity identifies whether evaluation used optical proxies, biomass measurements, taxonomic observations, toxin measurements, or event records. Deployment-domain separation records explicit shifts in time, region, sensor, water optical type, species, or pigment/optical phenotype. Operational relevance considers observation availability, class prevalence, timeliness, and error costs. These dimensions answer different questions and should not be collapsed into a single hierarchy of validation strength.

Different designs support different inferences. Held-out scenes test scene-level generalization, field matchups assess biological correspondence, event records assess event-level agreement, and chronological separation supports temporal generalization. Cross-domain transfer additionally requires an explicit target-domain shift and evaluation of a fixed configuration in that domain; performance retention further requires comparable source- and target-domain tasks and metrics. Labels such as "independent" or "external validation" are therefore insufficient unless the underlying validation dimensions are reported explicitly.

6.4. What the Comparative Evidence Supports

The studies in Table 2 substantiate different components of HAB monitoring performance. Hu et al. evaluated a regional fluorescence–chlorophyll relationship, whereas Carvalho et al. evaluated binary bloom decisions against field cell counts [45,70]. The latter study found different detection and false-positive profiles among initial configurations, with the operational method evaluated on a smaller valid-matchup subset [70]. These findings distinguish evidence that a satellite fluorescence signal is associated with chlorophyll-a from evidence that an algorithm identifies blooms defined by a field cell-count threshold. The two evaluations address different questions and cannot be substituted for one another.

Spatial evaluations provide complementary evidence, but their interpretation depends on how the reference maps were constructed. Image-derived reference masks, sometimes informed by field cell counts, were used for image-pair detection, testing on held-out scenes, test-subimage evaluations, and sensor-specific pixel assessments [12,13,64,67]. Du et al. additionally compared detections with independently sourced event records under explicit spatial and temporal matching rules [13]. That comparison assessed correspondence with reported events rather than pixel-level taxonomic truth. Together, these designs show that separation from model development and biological specificity of the reference are distinct properties: broader test coverage does not by itself provide a more specific biological reference. Table 2 should therefore be interpreted in light of each result's evaluation unit, reference data, and evidential boundaries, rather than as a cross-study ranking.

Table 2. Illustrative quantitative evidence relevant to marine HAB detection across heterogeneous study and evaluation designs. Values are reproduced as reported in the source studies and are not directly comparable across rows. The table is intended as an evidence map rather than a performance ranking: each example is interpreted in light of its data basis, method or evaluation design, and principal evidential boundary. Water-optical context follows the descriptions reported in the original studies. Case-1/Case-2 labels are used only where explicitly assigned by the authors and may characterize a study region rather than individual matchups. Otherwise, reported optical conditions are described without assigning a Case class. These descriptions do not imply water-type-stratified performance evaluation. Rrs, remote-sensing reflectance; FLH, fluorescence line height; PCR, polymerase chain reaction; CNN, convolutional neural network; RMSE, root-mean-square error; NIR, near-infrared.

Study/Target	Sensor/Method	Reference Data and Validation	Directly Reported Result	Interpretation and Boundary
Hu et al.; *K. brevis* [45]	MODIS/Aqua FLH; regional FLH–chlorophyll relationship; OC3M band-ratio comparison	Satellite–field chlorophyll comparisons within ±2 h along a nearshore transect near Charlotte Harbor (19 October 2004); reference chlorophyll derived from calibrated underway fluorescence; FLH $n = 77$, OC3M $n = 75$.	FLH–in situ chlorophyll: r = 0.92 ($n = 77$; reference Chl 0.4–4 mg m^{-3}). OC3M–in situ chlorophyll: r = 0.71 ($n = 75$); OC3M estimates were reported to be 3–15 times in situ values along the nearshore transect.	The FLH–chlorophyll correlation describes the relationship fitted to the October transect; independent predictive accuracy was not assessed in this comparison. OC3M overestimation was discussed in relation to possible CDOM interference and shallow-bottom reflectance; atmospheric-correction contributions were not isolated.
Ahn and Shanmugam; *C. polykrikoides* red tide [55]	SeaWiFS; red-tide index (RI), red-tide chlorophyll algorithm (RCA), and referenced local chlorophyll algorithm (LCA)	Multi-cruise in situ radiometric, chlorophyll-a, and cell-count observations; regional SeaWiFS application. The cited LCA relationship involved 118 records, whose detailed provenance and satellite–field matchup status were not clarified in this article.	$R^2 = 0.85$ ($n = 118$), reported for a cited LCA empirical relationship between the blue–green reflectance ratio and chlorophyll concentration.	The authors characterized the Northeast Asian coastal study environment as Case-2, with turbid and relatively clearer subregions. The cited LCA statistic does not establish independent satellite–field validation of RI/RCA.
Carvalho et al.; *K. brevis* [70]	MODIS-Aqua; initial empirical, bio-optical, and automated chlorophyll-anomaly (operational) methods	In situ K. brevis cell counts (positive: ≥15,000 cells L^{-1}); pooled summer–autumn satellite–field matchups on the central west Florida shelf, subject to optical-quality screening; $n = 371$ for empirical/bio-optical and $n = 328$ for the operational method.	Sensitivity/specificity: empirical 86%/53%; bio-optical 77%/65%; operational 26%/84% (initial configurations).	Pooled multi-year matchups show contrasting detection and false-positive profiles among the initial configurations. The operational method used a smaller valid subset, so the three-method comparison was not fully paired. No separate test set was reported for subsequent empirical/bio-optical threshold tuning. Riverine CDOM was discussed as a potential source of optical interference.

Table 2. *Cont*.

Study/Target	Sensor/Method	Reference Data and Validation	Directly Reported Result	Interpretation and Boundary
Babu et al.; six bloom classes [61]	MODIS-Aqua; regularized least-squares classifier	Event-derived labels from multiple Indian coastal bloom sites; random pixel split within the same image archive. Optical-water type and sample-level CDOM or suspended-matter conditions were not reported for the evaluated pixels.	Overall accuracy = 94.37% using R_{rs} + normalized water-leaving radiance.	Demonstrates strong within-archive separability. Random pixel splitting within the same imagery and the absence of contemporaneous pixel-level species matchups limit evidence for scene- or region-level generalization.
Noh et al.; *C. polykrikoides* [31]	GOCI; red-to-blue-ratio chlorophyll retrieval	PCR/species confirmation and 13 satellite–field chlorophyll-a matchups from one event.	R^2 = 0.92; mean normalized bias = −44%; mean normalized gross error = 63.5%.	Shows that strong correlation can coexist with substantial absolute error. The evaluation was restricted to one low-sediment East Sea event, with spatial mixing, advection, high-NIR radiance, and atmospheric correction identified as limitations.
Liu et al.; *N. scintillans* plus spectral tests of other taxa [38]	GF-1, HY-1D, Landsat-8, and Sentinel-2; pseudo-hue-angle index	Field spectra plus development/validation samples; sample-level cross-sensor application.	GF-1 source validation: mean overall accuracy = 92%, F1 = 0.92, precision = 0.92.	The GF-1 source evaluation included clear and turbid waters, with residual spectral confusion between blooms and turbid waters. The index was applied across several sensors, but no controlled sensor-held-out source–target comparison quantified performance retention.
Yao et al.; *K. brevis* [12]	VIIRS; CNN semantic segmentation with spatial coherence	Field cell-count archive (23,232 records) and semi-objectively delineated reference maps; 53 retained image pairs for detection evaluation.	Image-pair detection: F1 = 89%, accuracy = 81%, precision = 83%. Separate bloom-area comparison: relative RMSE = 31.5%, R^2 = 0.92.	The model produced bloom masks evaluated through image-pair detection and separate bloom-area comparisons within the West Florida Shelf domain, where estuaries and land-adjacent pixels were masked. Sediment resuspension could obscure bloom signals and lead to missed detections. Bloom labeling combined field counts $>10^5$ cells L^{-1} with RGCI and image features. Held-out images did not constitute a new deployment domain.

Table 2. *Cont*.

Study/Target	Sensor/Method	Reference Data and Validation	Directly Reported Result	Interpretation and Boundary
Du et al.; large-scale HAB mapping [13]	MODIS, VIIRS, and OLCI; Res-UNet	Sensor-specific image-derived labels; 36 test images per sensor, separate from training and validation; 28 high-spatial-resolution images for same-day qualitative spatial comparison; separate Chinese marine-disaster bulletin comparison.	Mean sensor-specific pixel-level F1 = 0.85; event-level bulletin agreement (reported as accuracy) = 81.09%.	The main article illustrated nearshore and highly turbid challenge scenes across Chinese marginal seas but did not report optical-water classifications for individual reference images. Multi-sensor mapping and comparisons with marine disaster bulletins provided broad application evidence, but a controlled source-to-target performance decrement was not isolated.

7. Factors Shaping Reliability and Evidence for Transferability

7.1. Reliability Is a Chain Property

Reliability of a satellite HAB product is a property of the complete observation-to-claim chain rather than of the algorithm alone (Sections 3–6). Target signal, water and atmospheric background, sensor information, preprocessing, training data, and validation design all influence the final interpretation [6,38]. Product outputs must also be evaluated against reference evidence of appropriate biological specificity and accompanied by task-relevant error characterization [4,70]. Algorithmic stability alone does not establish reliable interpretation when observation information, preprocessing, or reference evidence is inadequate.

This chain perspective also explains why the same method can perform differently across environments. In a VIIRS application, sediment resuspension could obscure *K. brevis* bloom signals and lead to missed detections [12]. Separately, changes in label definition can alter what a model learns and how its outputs are evaluated (Sections 6.1 and 7.4). Performance should therefore be interpreted as conditional on the observation, environmental, and validation setting rather than as an intrinsic property of the algorithm. Figure 4 integrates this chain perspective with the empirical logic required to evaluate transferability.

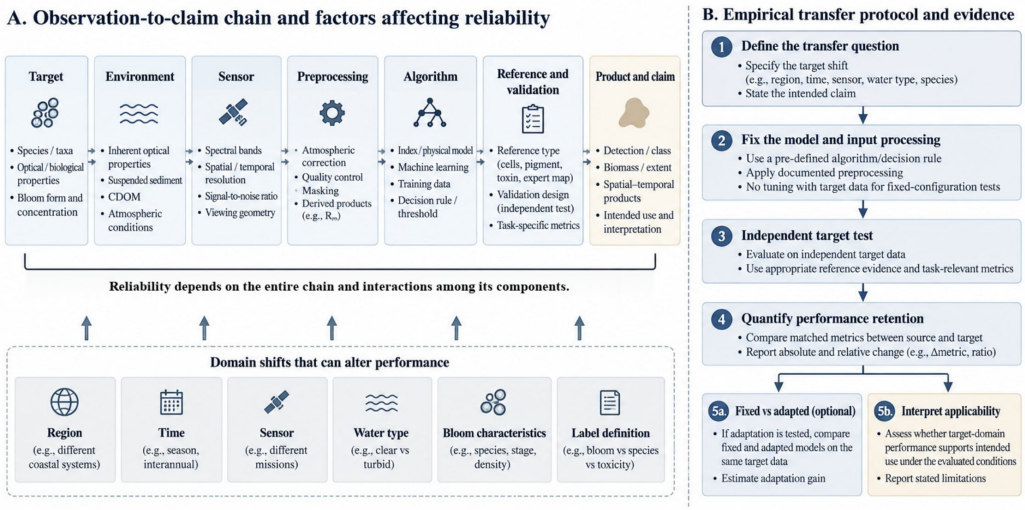


Figure 4. Reliability and empirical transfer logic for satellite HAB products. (**A**) Reliability depends on the full observation-to-claim chain, from the target signal through environment, sensor, preprocessing, algorithm, and reference/validation to the final product and intended claim. Domain shifts in region, time, sensor, water type, bloom characteristics, or label definition can alter performance even when the same method is used. (**B**) Empirical transfer assessment begins with a clearly defined target-domain shift, followed by a prespecified model/decision configuration and documented preprocessing, independent target-domain evaluation, and matched source–target comparison to quantify performance retention. When adaptation is evaluated, fixed and adapted configurations should be compared on the same target data to estimate adaptation gain. Target-domain performance supports applicability within the evaluated conditions and should be interpreted together with its stated limitations.

7.2. Species and Water Optical Type

Taxonomic composition and physiological state influence bloom optical properties [14,30], while vertical distribution affects the near-surface signal available to satellites [17]. Water optical conditions shape the background against which these signals are observed. CDOM and suspended particles alter water-leaving spectra and perceived color, while shallow-bottom reflectance can further modify the signal in optically shallow waters [26]. Performance changes across regions therefore should not be attributed to

geography alone [38]. Regional transfer tests should identify accompanying changes in species composition and water optical background and, where possible, evaluate their contributions separately.

7.3. Sensor Configuration and Preprocessing

Cross-sensor application introduces spectral, radiometric, spatial, and processing differences, so a new sensor often constitutes a new observation domain rather than a simple substitution of input data. Methods that rely on fluorescence or other narrow-band information cannot transfer unchanged to sensors lacking corresponding bands, and even nominally similar bands may yield different feature distributions after sensor-specific processing. Sensor sensitivity and dynamic range also contribute to the uncertainty budget [72].

Apparent cross-sensor performance may nevertheless arise from reliance on spatial morphology or broadband color rather than sensor-stable spectral information. Feature ablation, sensor-stratified validation, and independent target-domain testing under a pre-specified processing and decision chain can help determine whether retained performance reflects transferable target information or shared background structure.

7.4. Training Data and Labels

Training datasets encode not only target examples but also source-domain choices in event selection, negative-class definition, annotation, spatial unit, and sampling. "Non-HAB" may denote clear water in one study but turbid water, high-chlorophyll non-harmful blooms, or other confounders in another. Models trained with easy negatives may therefore achieve high internal scores while underestimating deployment false-positive risk. Cross-domain datasets should include representative hard negatives, uncertain labels, and event-, region-, and sensor-stratified metadata so that performance changes can be distinguished from shifts in the biological target, background distribution, or label definition.

7.5. Transferability Evidence in the Coded Corpus

Cross-domain evidence in the 28 full-text core studies fell into five mutually exclusive states: one strict fixed-configuration case, one qualified fixed independent-target evaluation, four target-informed or adapted evaluations, 13 held-out evaluations that were not cross-domain, and nine studies with no qualifying cross-domain evidence. The strict case applied the same pseudo-hue-angle decision rule to later target data under a composite time–region–sensor shift and reported target-sensor performance [38]. A second study tested fixed pigment thresholds on independent images, providing qualified evidence of target-domain applicability for pigment/optical-phenotype classes [73].

Four additional studies modified thresholds, wavelengths, parameters, or other configuration elements using target-domain information [37,74–76]. These studies document applications using target-informed or adapted configurations, including a qualitative external application; they do not quantify retention of the original configuration. Other field comparisons, held-out images, event records, and prospective temporal tests provide reliability or generalization evidence but are not equivalent to transferability tests without an explicit source–target shift and fixed-configuration comparison [13,77–79]. Figure 5a maps validation and cross-domain evidence states across the 28 studies, while Figure 5c summarizes the nonexclusive domain-shift descriptors among qualifying cross-domain applications.

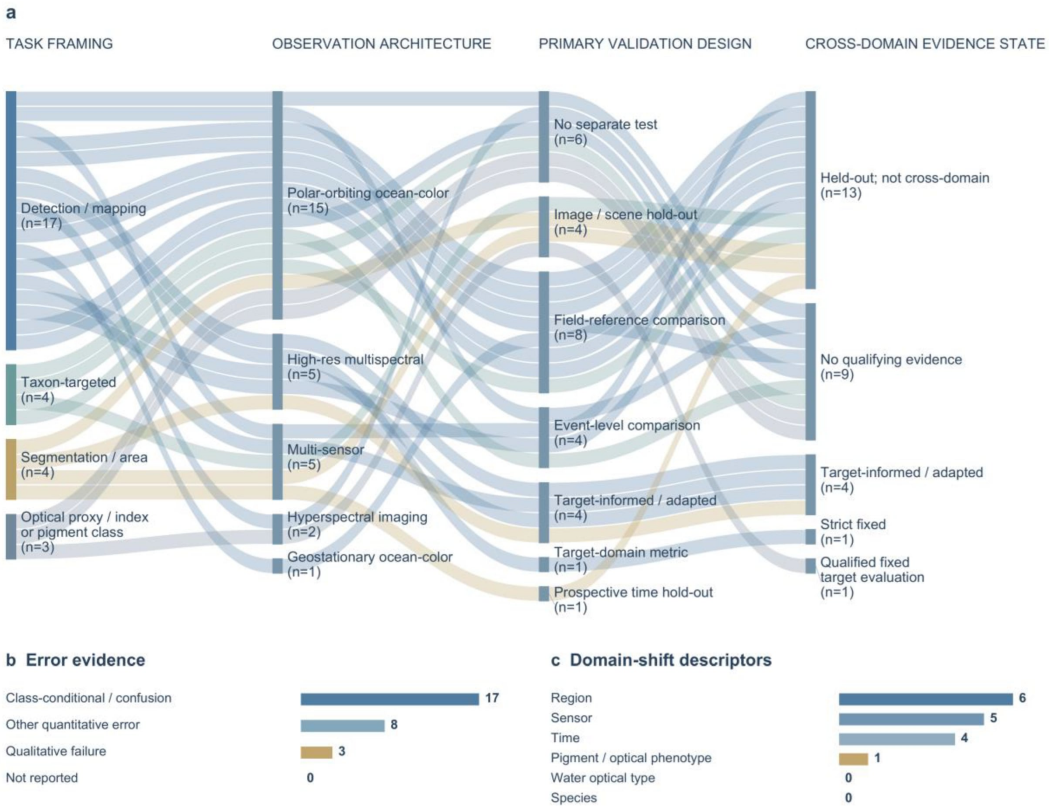


Figure 5. Evidence pathways and transfer-related reporting across the 28 retained full-text core studies. (**a**) Alluvial map linking task framing, observation architecture, primary validation design, and cross-domain evidence state. Each study contributes once to each axis. Ribbon color denotes task framing, and ribbon width represents the number of studies. (**b**) Mutually exclusive categories of reported error evidence. (**c**) Nonexclusive domain-shift descriptors among six cross-domain applications (one strict fixed, one qualified fixed, and four target-informed or adapted); co-occurring descriptors are not isolated shift effects. Task framing and observation architecture are normalized descriptive categories derived from the reported study designs, whereas validation and cross-domain evidence follow the review's documented coding framework and adjudication rules. The figure is descriptive and is intended as an evidence map rather than a performance ranking. Alluvial pathways do not imply causality, chronological progression, validation strength, or method superiority.

Within the 28 coded core studies, no study quantified paired source-to-target performance retention under matched tasks and metrics, met the review criteria for species transfer, or isolated the effect of a single shifted axis. The available evidence therefore supports particular target-domain applications but leaves performance retention unresolved. Table 3 summarizes the corresponding evidence and inferential boundaries.

To address this evidence gap, Figure 4B outlines a source–target evaluation procedure. Researchers should first prespecify the domains and shift descriptors, task and reference definitions, evaluation units and metrics, fixed decision layer, and sensor-specific input-processing rules. The fixed configuration should then be evaluated on independent source and target test data separated from model development and configuration selection, with comparable results used to assess performance retention. When adaptation involves target-domain fitting or selection, including the use of unlabeled target data, those data should be kept separate from the final target test set. Fixed and adapted configurations should then be compared on the same independent target test set to estimate adaptation gain. Finally, target-domain performance should be interpreted in relation to the evaluated conditions and applicability limits.

Table 3. Transfer-related evidence in selected studies from the coded corpus and the inference each design supports.

Study	Target Setting and Validation Basis	Cross-Domain Evidence State	Evidence Directly Supports	Interpretation Boundary
Carvalho et al. [70]	Pooled multi-year satellite–field matchups; no separate test set was reported for subsequent empirical/bio-optical threshold tuning	No qualifying cross-domain evidence	Contrasting detection and false-positive profiles among initial configurations; the three-method comparison was not fully paired	No independently defined target domain or matched source–target comparison
Westberry et al. [77]	Independent SeaWiFS–field comparison with quantified false positives	Held-out evaluation, not cross-domain	Field-tested *Trichodesmium* discrimination and class-specific error characterization	Independent reference data do not by themselves constitute a source-to-target transfer test
Liu et al. [38]	Same pseudo-hue-angle decision rule applied to later target datasets with per-sensor metrics	Strict fixed configuration	Fixed-rule applicability under a composite time–region–sensor shift	Source-to-target performance loss was not quantified; shifted axes co-varied
Hyperspectral accessory-pigment framework [73]	Fixed pigment thresholds tested on independent images	Qualified fixed independent-target evaluation	Independent-target applicability for pigment/optical-phenotype classes	Performance was pooled across constituent domains; no source baseline or species-transfer test
Multi-/hyperspectral self-supervised framework [79]	Prospective later-period tests within separately trained instrument/data-stream configurations	Held-out evaluation, not cross-domain	Temporal generalization within each prespecified configuration	No unchanged configuration crossed an independently defined sensor or regional domain; no paired loss estimate
GF-1 broadband detector [37]	Later regional application after threshold adjustment; no target-region accuracy metric	Target-informed or adapted evaluation	Qualitative applicability after regional adaptation	Target-informed threshold changes and absence of a target metric preclude fixed-configuration or quantified-retention claims
Binary-quantization vision transformer [74]	Spatially separated target regions; training duration selected after inspection of target-test performance	Target-informed or adapted evaluation	Performance on geographically separated target imagery after target-informed selection	Configuration was not fixed before target evaluation; sensor families were represented during training

Table 3. *Cont*.

Study	Target Setting and Validation Basis	Cross-Domain Evidence State	Evidence Directly Supports	Interpretation Boundary
Normalized red-tide index [75]	Target-domain field comparison after wavelength modification and regional fitting	Target-informed or adapted evaluation	Target-domain applicability after adaptation	Wavelength modification and target-data fitting prevent interpretation as fixed-configuration transfer
Spectral–spatial prior-guided detector [76]	Multi-platform target evaluation with within-image fitting and target-image parameter analysis	Target-informed or adapted evaluation	External-target performance after target-specific fitting	The decision configuration was not fixed across target images; no paired transfer-loss estimate

Notes: "Fixed" indicates that the decision layer—threshold, rule, or trained model—was carried to independent target observations without target-label fitting or target-test selection. "Adapted" indicates that target-domain information was used to modify one or more configuration elements. Held-out evaluation without an explicit source–target domain shift is treated as generalization evidence rather than direct transfer evidence. The complete record-level coding for all 28 retained full-text core studies is provided in Supplementary Table S1.

8. Reported Error, Uncertainty, and Propagation Evidence

Building on the determinants of reliability and transferability discussed in Section 7, this section examines how error and uncertainty may propagate along the satellite HAB observation–processing–inference chain and what evidence the reviewed literature actually provides for these pathways. The aim is not to assume that all sources of uncertainty can already be quantified independently, but to identify where uncertainty enters the chain, which downstream products it may affect, and what minimum uncertainty information should accompany the final product.

8.1. An End-to-End Uncertainty Chain

We organize uncertainty along an observation-to-claim chain comprising data quantities and products, processing operations, and conditional interpretation:

top-of-atmosphere radiance → radiometric calibration and geolocation → atmospheric correction → remote-sensing reflectance (R_{rs}) → features or retrieved properties → classification/segmentation → bloom boundary and area → event or biological-claim interpretation.

Error, uncertainty, and information loss can enter at each stage. Calibration, geolocation, and atmospheric correction determine the spectral information available downstream; feature construction and bio-optical retrieval can propagate or transform upstream errors; and classification, segmentation, aggregation, and masking add uncertainty in model decisions, bloom geometry, and observation availability. Inferential uncertainty also arises when optical proxies are not uniquely linked to the intended biological property. Non-random losses from clouds, sun glint, adjacency effects, or processing failure require particular attention because they affect both product accuracy and when or where monitoring is possible. Figure 6 therefore presents these components as potential uncertainty pathways and an audit framework, not as evidence of monotonic error accumulation or fully quantified propagation across all stages.

8.2. Reported Errors and Evidence Relevant to Propagation

Reviewed studies document errors at multiple stages of the observation-to-claim chain, but most do not isolate or quantitatively propagate them across processing stages. A GOCI chlorophyll retrieval showed that high correlation can coexist with substantial bias and gross error [31]. MODIS OC3M overestimation along a nearshore transect was discussed in relation to possible CDOM interference and shallow-bottom reflectance, without isolating atmospheric-correction effects [45]. In a VIIRS application, sediment-related obscuration contributed to missed detections [12]. Multi-year *K. brevis* matchups showed configuration-dependent detection and false-positive profiles, with the three-method comparison not fully paired [70]. In one GF-1 scene, artificial reflectance perturbations changed the classification results of the pseudo-hue-angle red-tide method [38]. Overall, these studies support plausible propagation pathways more strongly than quantitatively resolved relationships among successive stages.

Among the 28 full-text core studies, 17 reported quantitative class-conditional or confusion-derived evidence, eight reported other quantitative error or uncertainty, and three reported qualitative failure conditions only; none lacked relevant error information. These mutually exclusive categories are summarized in Figure 5b. They characterize reported error evidence rather than a complete uncertainty budget: sensitivity, specificity, and confusion matrices describe classification error but do not establish probability calibration or quantify propagation of upstream uncertainty.

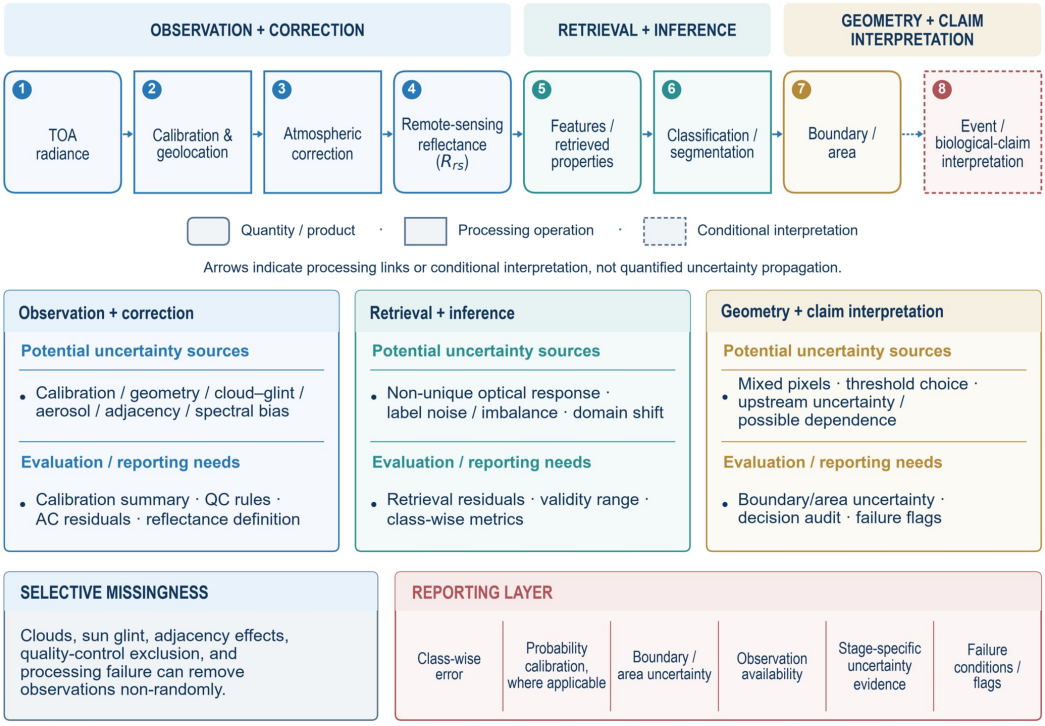


Figure 6. Potential uncertainty pathways across the ocean-color HAB observation-to-claim chain. Numbered elements distinguish data quantities and products (1, 4, and 5), processing operations (2, 3, and 6), product geometry (7), and conditional event or biological-claim interpretation (8). Solid arrows show schematic processing and product-generation links; the dashed link to element 8 indicates a conditional interpretive step, for which the strength of the biological claim depends on appropriate reference evidence rather than mapped geometry alone. Stage summaries identify major sources of uncertainty and corresponding evaluation or reporting needs across observation/correction, retrieval/inference, and geometry/claim-interpretation stages. These links locate potential uncertainty influences and dependencies; they do not imply monotonic error accumulation, a universal propagation coefficient, or quantitatively resolved propagation at every link. Selective missingness is shown separately because clouds, sun glint, adjacency effects, quality-control exclusion, and processing failure can remove observations non-randomly. The reporting layer emphasizes class-wise error, probability calibration where applicable, boundary/area uncertainty, observation availability, stage-specific uncertainty evidence, and failure conditions. Accordingly, a single average performance metric is insufficient to characterize full product-level uncertainty.

Error priorities also depend on the monitoring task. Missed blooms, false alerts, biased area estimates, and incorrect biological claims have different consequences. Quantification should therefore be organized around the error endpoints relevant to the intended monitoring task. For a prespecified decision rule, perturb inputs at a stated entry point and rerun the relevant downstream chain for each scenario; compare alternative processors or retuned models as separate configurations. Use physically or empirically supported perturbation bounds where available, and label hypothetical perturbations as sensitivity scenarios. Use data separate from the final test set to select configurations or retrain models. Distinguish changes in product outputs from errors evaluated against appropriate independent references, using consistent evaluation units and definitions. When masks differ, evaluate common valid support where feasible and report each configuration's coverage and results on its own evaluable support. Scenario ranges are not confidence or prediction intervals without an appropriate sampling or probabilistic basis. Such comparisons do not automatically identify independent error contributions or justify summing separate effects without accounting for dependence and interactions.

For future assessments, Monte Carlo sampling could propagate justified joint input distributions through specified nonlinear processing stages to characterize task-defined outputs [80]. Inputs should distinguish inherited uncertainty from additional stage residuals without counting upstream effects twice; processor differences alone do not define error distributions. For thresholded HAB products, simulated classifications and derived areas require output-specific summaries; simulated detection frequencies are conditional on the assumed input model and decision rule, not automatically calibrated probabilities of HAB occurrence. Bayesian networks could represent conditional dependencies when their structure and conditional distributions are justified, but network links alone do not establish causality [81]. These are proposed frameworks, not analyses performed here. Model-based intervals remain conditional on their assumptions; empirical interval coverage and the calibration of predicted probabilities should be evaluated, as applicable, using suitable independent reference data where available, with validation gaps and limitations from unobserved areas reported explicitly.

8.3. Minimum Uncertainty Reporting

We recommend that future satellite HAB studies report:

- The atmospheric-correction processor and version, quality-control and masking rules, and reflectance convention;
- The number and proportion of scenes or pixels excluded by quality control and the resulting spatial and temporal availability of valid observations, including spatial coverage maps of valid-observation counts or fractions for defined monitoring periods, with low- or zero-coverage areas identified and any fraction's observation-opportunity denominator defined before the exclusions being assessed;
- Biological definitions and labeling criteria for positive, negative, and uncertain cases;
- Spatial and temporal matchup criteria between satellite observations and reference measurements;
- Training, validation, and test partitions at the event, scene, region, year, and sensor levels;
- Class prevalence and task-appropriate metrics such as precision, recall/sensitivity, specificity, F1, or IoU;
- Area, boundary, or other spatial-error measures for mapping products;
- Uncertainty intervals or variability across repeated splits or experiments rather than a single performance score;
- Probability or confidence calibration when probabilistic outputs are used;
- The consequences or costs of false positives and false negatives for the intended application;
- Comparable source- and target-domain results for transferability claims and, when adaptation is used, fixed-versus-adapted performance on the same target test data.

Confidence should characterize the reliability of the final product rather than only model output probability. A well-calibrated classifier cannot remove unquantified atmospheric-correction bias, selective observation loss, or ambiguity in biological labels. Uncertainty reporting should therefore address the major error sources relevant to the intended product and identify stage-specific evidence and unresolved components when complete end-to-end propagation cannot be quantified, rather than substituting a single model-performance metric for product-level uncertainty.

9. Emerging Opportunities

Sections 3 and 6 distinguish biological claim specificity from optical agreement and test independence. Within the 28 coded studies, matched source–target performance retention remained unquantified, while most reviewed error studies did not resolve cross-stage propagation (Sections 7.5 and 8.2). These findings motivate comparative tests of the added value of sensing and modelling choices for defined biological claims, domains, and monitoring tasks.

9.1. Hyperspectral Detection

Hyperspectral ocean-color observations provide denser spectral information than multispectral sensors and may improve discrimination of pigments, optical phenotypes, and taxonomic differences. Their main potential lies at the L3–L4 claim levels in Figure 2, where HAB detection moves toward taxon- or species-level discrimination. This added information strengthens inference only if diagnostic features remain detectable after atmospheric correction, spatial mixing, and coastal optical interference, and if biological references match the intended claim.

Freshwater DESIS studies and early PACE cyanobacterial applications offer methodological examples for hyperspectral feature design and validation [82,83], but their performance cannot be assumed to transfer to marine HABs. Marine applications therefore require renewed evaluation across target taxa, representative confounders, and contrasting optical waters, together with reporting of detection limits, correction residuals, spectral separability, and target-domain performance. The value of hyperspectral observations lies not simply in more bands, but in whether richer spectra support more specific biological claims.

9.2. Geostationary and Multi-Sensor Synergy

Geostationary, polar-orbiting, and high-spatial-resolution sensors provide complementary capabilities for HAB monitoring: frequent intraday sampling, broad regional continuity, and finer delineation of small coastal patches and bloom boundaries. Multi-sensor integration should therefore aim to increase usable observation opportunities and compensate for the temporal, spatial, and spectral limitations of individual sensors rather than merely create seamless composites.

Reliable multi-sensor integration should preserve acquisition time, native resolution, resampling history, processing provenance, and source-specific uncertainty, while linking satellite observations to continuous or coincident in-water measurements where possible [84]. To determine its operational value, future evaluations should compare valid-observation coverage and independently referenced event-detection performance against appropriate single-sensor baselines, reporting these outcomes separately. Where a common decision rule or trained model is applied across sensors, its applicability should be evaluated on independent target-sensor data without target-test tuning. Matched source–target performance comparisons are additionally needed to quantify performance retention.

9.3. Transferable and Data-Efficient Learning

Transfer learning, self-supervised learning, and related data-efficient approaches may reduce labeling demands and improve data efficiency in satellite HAB detection. Inland-water studies have explored transfer learning and cross-lake evaluation [85,86], while self-supervised learning has been applied to marine HAB monitoring [79]. These examples motivate testing whether learned representations remain useful under changing observation conditions.

Larger datasets or more complex models, however, do not ensure transferability. Domain shifts in water optical properties, sensors, target assemblages, and label definitions can limit generalization. Future frameworks should therefore combine data-efficient representation learning with explicit target-domain evaluation and comparable source–target assessment, distinguishing target-domain applicability, performance retention, and, where adaptation is used, adaptation gain.

Data-efficiency studies should compare learning curves under matched annotation budgets on the same independent target test set. Training labels and, where used, biological references for context assignment should be accounted for separately by domain; unlabeled and pretraining data use should also be documented.

9.4. Physics-Informed and Uncertainty-Aware Detection

Future satellite HAB detection can benefit from integrating data-driven models with ocean-optical knowledge, ecological processes, and physical constraints. Physics-informed approaches can incorporate bio-optical relationships, environmental suitability, spatiotemporal continuity, and transport information to provide more realistic priors, reduce process-inconsistent predictions, and improve interpretability. Their role is not to replace data-driven models but to constrain inference toward physically and biologically plausible solutions. Such constraints, however, do not guarantee generalizability: their validity may vary across regions, optical conditions, and target organisms and should therefore be evaluated with independent observations and target-domain validation.

Uncertainty-aware detection is equally important for reliable HAB products. Beyond categorical predictions or bloom areas, future systems should characterize observation quality, retrieval reliability, model confidence, and applicability limits. Probabilistic outputs, prediction intervals, confidence calibration, and explicit domain-of-applicability information can support this goal where appropriate. Trustworthy HAB monitoring should therefore couple physically informed inference with uncertainty characterization and continued validation, so that detection results are accompanied by evidence about when and where they can be relied upon.

9.5. From Detection to Forecasting and Operational Use

Satellite HAB monitoring should extend beyond post-event detection toward earlier risk assessment and dynamic forecasting that integrates satellite observations, environmental conditions, and ecological processes [87]. This transition, however, should not be equated with simply adopting more complex machine-learning models.

Reliable forecasting requires explicit prediction targets, forecast horizons, and information boundaries. Pre-event risk prediction, in-event transport and expansion forecasting, and post-event impact assessment are distinct tasks that require different inputs, validation designs, and uncertainty representations.

Operational systems should connect these forecasts with sustained monitoring, early warning, risk assessment, and response support. As summarized in Table 4, useful forecast products should report not only expected bloom location or magnitude but also confidence, applicability conditions, and potential failure modes. Their operational value ultimately depends on whether forecasts are interpretable, independently evaluated, and accompanied by explicit reliability information.

Table 4. Emerging opportunities and evidence requirements for improving satellite-based marine HAB detection and interpretation.

Challenge Identified in This Review	Emerging Opportunity	Expected Contribution	Required Evidence/Validation
Limited taxonomic specificity of multispectral observations	Hyperspectral ocean-color observations	Improve discrimination of pigment composition and optical phenotypes; potentially support higher-specificity biological claims	Compare hyperspectral and multispectral inputs using matched taxon-specific references and evaluation support across representative optical waters; assess spectral separability and detection limits.
Limited observation frequency and spatial coverage	Geostationary and multi-sensor observations	Increase observation availability and improve characterization of bloom development and dynamics	Against single-sensor baselines, evaluate valid-observation coverage and event-detection performance separately; document acquisition and processing consistency, source-specific uncertainty, and independent target-domain performance.
Insufficient matched source–target evidence under regional, sensor, and optical-domain shifts	Transferable and data-efficient learning	Reduce dependence on region-specific labels and improve representation learning across domains	Fixed-configuration evaluation on independent target domains with comparable source–target metrics; fixed-versus-adapted comparison on the same independent target test set when adaptation is used; learning curves under matched annotation budgets.
Black-box predictions and inconsistent biological interpretation	Physics-informed and uncertainty-aware detection	Improve interpretability, constrain unrealistic predictions, and characterize task-specific uncertainty	Compare constrained and unconstrained configurations on matched independent tests; assess task errors, constraint consistency, probability calibration where applicable, and applicability limits.
Need to evaluate detection products against forecasting and operational decision requirements	Forecasting and operational systems	Move from event detection toward early warning, risk assessment, and management support	Evaluate forecasts at stated lead times using only information available at issue time; compare reasonable baselines on independent events or chronological holdouts, and assess decision relevance.

10. Conclusions

Satellite ocean-color remote sensing has substantially advanced the observation of harmful algal blooms and now spans empirical spectral methods, multi-source observations, machine learning, and high-dimensional spectral information. However, as detection capability has expanded, the central challenge has shifted from whether optical anomalies can be identified to whether satellite observations support the intended biological claims and whether task-relevant performance is retained when observation and environmental conditions change.

This review developed an evidence-constrained framework to examine reliability across the satellite HAB observation–processing–inference–claim chain and transferability across

changing application domains. The analysis highlights a fundamental evidence boundary between optical signals, algorithmic outputs, and biological conclusions. Bloom presence or high-biomass optical expressions may be supported by remote-sensing evidence, whereas higher-specificity claims such as taxon attribution, species-level discrimination, or toxicity inference require correspondingly specific independent biological references. Within the 28 coded core studies, we identified one strict fixed-configuration application and one qualified fixed independent-target evaluation, but no study quantified source-to-target performance retention under matched tasks and metrics. Current limitations therefore arise not only from algorithm performance, but also from validation design, reference and labeling practices, preprocessing and observation constraints, limited cross-domain evaluation, and incomplete uncertainty characterization.

Future satellite HAB monitoring should move beyond single performance metrics toward joint assessment of reliability and transferability. Hyperspectral observations, multi-sensor synergy, data-efficient learning, physics-informed approaches, and uncertainty-aware detection provide important opportunities to improve monitoring capability, but their value depends on validation frameworks that match the specificity of the intended claims, use appropriate independent reference evidence, and explicitly evaluate performance under relevant target-domain shifts. Cross-domain applicability should not be interpreted as performance retention unless comparable source- and target-domain evidence is available.

Ultimately, HAB remote-sensing systems should move beyond providing detection results alone toward delivering evidence-supported products together with their applicability limits, uncertainty information, and documented failure conditions. By integrating observation capability, ecological and optical knowledge, model inference, independent validation, uncertainty characterization, and operational requirements, satellite remote sensing can evolve from an event-detection tool into a more trustworthy monitoring system for marine ecological assessment, early warning, and decision support.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18213813/s1. Supplementary Methods S1: Structured Web of Science search strategy; Supplementary Methods S2: Machine-assisted screening and evidence prioritization; Supplementary Methods S3: Full-text evidence coding and transfer adjudication; Supplementary Table S1: Study-level validation, cross-domain evidence, error reporting, and inferential boundaries for the 28 retained full-text core studies; Supplementary Data S1: Simplified submission dataset.

Author Contributions: Conceptualization, J.Z. and Z.T.; methodology, Z.T. and J.Z.; validation, W.L., J.H. and X.Z.; formal analysis, Z.T. and Z.L.; investigation, Z.L., W.L. and J.H.; resources, J.Z. and H.Z.; data curation, Z.L., W.L., J.H. and X.Z.; writing—original draft preparation, Z.T. and J.Z.; writing—review and editing, J.Z., Z.T., W.L., J.H., X.Z. and H.Z.; visualization, Z.T. and Z.L.; supervision, J.Z.; project administration, J.Z. and H.Z.; funding acquisition, J.Z. and H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Laboratory of Coastal Salt Marsh Ecosystems and Resources, Ministry of Natural Resources, grant number KLCSMERMNR202511.

Data Availability Statement: Final study-level coding and concise source anchors are provided in Supplementary Table S1 and Supplementary Data S1. Structured search strategies, screening procedures, and coding rules are described in Supplementary Methods S1–S3. Detailed internal audit records and intermediate working files are retained by the authors and are not included in the submitted Supplementary Files. Bibliographic records retrieved from the Web of Science Core Collection are subject to database licensing restrictions and are therefore not redistributed in full.

Acknowledgments: During the preparation of this manuscript, the authors used DeepSeek-V4.1-Flash for machine-assisted literature screening, language refinement, manuscript organization, figure and

caption refinement, and consistency and mechanical checks. The authors reviewed the AI-assisted materials used in the study and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Anderson, D.M.; Fensin, E.; Gobler, C.J.; Hoeglund, A.E.; Hubbard, K.A.; Kulis, D.M.; Landsberg, J.H.; Lefebvre, K.A.; Provoost, P.; Richlen, M.L.; et al. Marine harmful algal blooms (HABs) in the United States: History, current status and future trends. *Harmful Algae* **2021**, *102*, 101975. [CrossRef] [PubMed]
2. Stumpf, R.P. Applications of Satellite Ocean Color Sensors for Monitoring and Predicting Harmful Algal Blooms. *Hum. Ecol. Risk Assess. Int. J.* **2001**, *7*, 1363–1368. [CrossRef]
3. Babin, M.; Cullen, J.; Roesler, C.; Donaghay, P.; Doucette, G.; Kahru, M.; Lewis, M.R.; Scholin, C.A.; Sieracki, M.E.; Sosik, H.M. New Approaches and Technologies for Observing Harmful Algal Blooms. *Oceanography* **2005**, *18*, 210–227. [CrossRef]
4. Stauffer, B.A.; Bowers, H.A.; Buckley, E.; Davis, T.W.; Johengen, T.H.; Kudela, R.; McManus, M.A.; Purcell, H.; Smith, G.J.; Vander Woude, A.; et al. Considerations in Harmful Algal Bloom Research and Monitoring: Perspectives From a Consensus-Building Workshop and Technology Testing. *Front. Mar. Sci.* **2019**, *6*, 399. [CrossRef]
5. Gokul, E.A.; Raitsos, D.E.; Gittings, J.A.; Alkawri, A.; Hoteit, I. Remotely sensing harmful algal blooms in the Red Sea. *PLoS ONE* **2019**, *14*, e0215463. [CrossRef] [PubMed]
6. Caballero, I.; Fernández, R.; Escalante, O.M.; Mamán, L.; Navarro, G. New capabilities of Sentinel-2A/B satellites combined with in situ data for monitoring small harmful algal blooms in complex coastal waters. *Sci. Rep.* **2020**, *10*, 8743. [CrossRef] [PubMed]
7. Zheng, G.; DiGiacomo, P.M. Uncertainties and applications of satellite-derived coastal water quality products. *Prog. Oceanogr.* **2017**, *159*, 45–72. [CrossRef]
8. Monteiro, H.; Goela, P.; Pinto, R.; Cristina, S. Harmful Algal Blooms on the Portuguese coast: Cross-checking events with remote sensing ocean colour data for coastal management. *Reg. Stud. Mar. Sci.* **2024**, *77*, 103723. [CrossRef]
9. Blondeau-Patissier, D.; Gower, J.F.R.; Dekker, A.G.; Phinn, S.R.; Brando, V.E. A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Prog. Oceanogr.* **2014**, *123*, 123–144. [CrossRef]
10. Lou, X.; Hu, C. Diurnal changes of a harmful algal bloom in the East China Sea: Observations from GOCI. *Remote Sens. Environ.* **2014**, *140*, 562–572. [CrossRef]
11. Werdell, P.J.; Behrenfeld, M.J.; Bontempi, P.S.; Boss, E.; Cairns, B.; Davis, G.T.; Franz, B.A.; Gliese, U.B.; Gorman, E.T.; Hasekamp, O.; et al. The Plankton, Aerosol, Cloud, Ocean Ecosystem Mission: Status, Science, Advances. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 1775–1794. [CrossRef]
12. Yao, Y.; Hu, C.; Cannizzaro, J.P.; Barnes, B.B.; English, D.C.; Xie, Y.; Hubbard, K.; Wang, M. Detection of *Karenia brevis* red tides on the West Florida Shelf using VIIRS observations: Accounting for spatial coherence with artificial intelligence. *Remote Sens. Environ.* **2023**, *298*, 113833. [CrossRef]
13. Du, Y.; Wang, M.; Li, Z.B.; Li, J.; Hu, J. Abrupt intensification and spatiotemporal dynamics of large-scale harmful algal blooms in China's marginal seas mapped using multi-sensor satellite imagery and deep learning. *ISPRS J. Photogramm. Remote Sens.* **2026**, *237*, 183–198. [CrossRef]
14. Mouw, C.B.; Hardman-Mountford, N.J.; Alvain, S.; Bracher, A.; Brewin, R.J.W.; Bricaud, A.; Ciotti, A.M.; Devred, E.; Fujiwara, A.; Hirata, T.; et al. A Consumer's Guide to Satellite Remote Sensing of Multiple Phytoplankton Groups in the Global Ocean. *Front. Mar. Sci.* **2017**, *4*, 41. [CrossRef]
15. Groom, S.; Sathyendranath, S.; Ban, Y.; Bernard, S.; Brewin, R.; Brotas, V.; Brockmann, C.; Chauhan, P.; Choi, J.K.; Chuprin, A.; et al. Satellite Ocean Colour: Current Status and Future Perspective. *Front. Mar. Sci.* **2019**, *6*, 485. [CrossRef] [PubMed]
16. Anderson, C.R.; Kudela, R.M.; Kahru, M.; Chao, Y.; Rosenfeld, L.K.; Bahr, F.L.; Anderson, D.M.; Norris, T.A. Initial skill assessment of the California Harmful Algae Risk Mapping (C-HARM) system. *Harmful Algae* **2016**, *59*, 1–18. [CrossRef] [PubMed]
17. Seegers, B.N.; Birch, J.M.; Marin, R.; Scholin, C.A.; Caron, D.A.; Seubert, E.L.; Howard, M.D.A.; Robertson, G.L.; Jones, B.H. Subsurface seeding of surface harmful algal blooms observed through the integration of autonomous gliders, moored environmental sample processors, and satellite remote sensing in southern California. *Limnol. Oceanogr.* **2015**, *60*, 754–764. [CrossRef]
18. Ruiz-Villarreal, M.; Sourisseau, M.; Anderson, P.; Cusack, C.; Neira, P.; Silke, J.; Rodriguez, F.; Ben-Gigirey, B.; Whyte, C.; Giraudeau-Potel, S.; et al. Novel Methodologies for Providing In Situ Data to HAB Early Warning Systems in the European Atlantic Area: The PRIMROSE Experience. *Front. Mar. Sci.* **2022**, *9*, 791329. [CrossRef]
19. Morel, A.; Prieur, L. Analysis of variations in ocean color. *Limnol. Oceanogr.* **1977**, *22*, 709–722. [CrossRef]

20. Gordon, H.R.; Brown, O.B.; Evans, R.H.; Brown, J.W.; Smith, R.C.; Baker, K.S.; Clark, D.K. A semianalytic radiance model of ocean color. *J. Geophys. Res. Atmos.* **1988**, *93*, 10909–10924. [CrossRef]
21. Neville, R.A.; Gower, J.F.R. Passive remote sensing of phytoplankton via chlorophyll α fluorescence. *J. Geophys. Res.* **1977**, *82*, 3487–3493. [CrossRef]
22. Gitelson, A. The peak near 700 nm on radiance spectra of algae and water: Relationships of its magnitude and position with chlorophyll concentration. *Int. J. Remote Sens.* **1992**, *13*, 3367–3373. [CrossRef]
23. Gower, J.F.R.; Doerffer, R.; Borstad, G.A. Interpretation of the 685nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS. *Int. J. Remote Sens.* **1999**, *20*, 1771–1786. [CrossRef]
24. Johnsen, G.; Nelson, N.B.; Jovine, R.V.M.; Prézelin, B.B. Chromoprotein- and pigment-dependent modeling of spectral light absorption in two dinoflagellates, *Prorocentrum minimum* and *Heterocapsa pygmaea*. *Mar. Ecol. Prog. Ser.* **1994**, *114*, 245–258. [CrossRef]
25. Millie, D.F.; Schofield, O.M.; Kirkpatrick, G.J.; Johnsen, G.; Tester, P.A.; Vinyard, B.T. Detection of harmful algal blooms using photopigments and absorption signatures: A case study of the Florida red tide dinoflagellate, *Gymnodinium breve*. *Limnol. Oceanogr.* **1997**, *42*, 1240–1251. [CrossRef]
26. Dierssen, H.M.; Kudela, R.M.; Ryan, J.P.; Zimmerman, R.C. Red and black tides: Quantitative analysis of water-leaving radiance and perceived color for phytoplankton, colored dissolved organic matter, and suspended sediments. *Limnol. Oceanogr.* **2006**, *51*, 2646–2659. [CrossRef]
27. Hu, C. A novel ocean color index to detect floating algae in the global oceans. *Remote Sens. Environ.* **2009**, *113*, 2118–2129. [CrossRef]
28. Johnsen, G.; Samset, O.; Granskog, L.; Sakshaug, E. In vivo absorption characteristics in 10 classes of bloom-forming phytoplankton: Taxonomic characteristics and responses to photoadaptation by means of discriminant and HPLC analysis. *Mar. Ecol. Prog. Ser.* **1994**, *105*, 149–157. [CrossRef]
29. Tomlinson, M.C.; Stumpf, R.P.; Ransibrahmanakul, V.; Truby, E.W.; Kirkpatrick, G.J.; Pederson, B.A.; Vargo, G.A.; Heil, C.A. Evaluation of the use of SeaWiFS imagery for detecting *Karenia brevis* harmful algal blooms in the eastern Gulf of Mexico. *Remote Sens. Environ.* **2004**, *91*, 293–303. [CrossRef]
30. Cannizzaro, J.P.; Carder, K.L.; Chen, F.R.; Heil, C.A.; Vargo, G.A. A novel technique for detection of the toxic dinoflagellate, *Karenia brevis*, in the Gulf of Mexico from remotely sensed ocean color data. *Cont. Shelf Res.* **2008**, *28*, 137–158. [CrossRef]
31. Noh, J.H.; Kim, W.; Son, S.H.; Ahn, J.H.; Park, Y.J. Remote quantification of *Cochlodinium polykrikoides* blooms occurring in the East Sea using geostationary ocean color imager (GOCI). *Harmful Algae* **2018**, *73*, 129–137. [CrossRef] [PubMed]
32. Hovis, W.A.; Clark, D.K.; Anderson, F.; Austin, R.W.; Wilson, W.H.; Baker, E.T.; Ball, D.; Gordon, H.R.; Mueller, J.L.; El-Sayed, S.Z.; et al. Nimbus-7 Coastal Zone Color Scanner: System Description and Initial Imagery. *Science* **1980**, *210*, 60–63. [CrossRef] [PubMed]
33. Esaias, W.E.; Abbott, M.R.; Barton, I.; Brown, O.B.; Campbell, J.W.; Carder, K.L.; Clark, D.K.; Evans, R.H.; Hoge, F.E.; Gordon, H.R.; et al. An overview of MODIS capabilities for ocean science observations. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1250–1265. [CrossRef]
34. Choi, J.K.; Park, Y.J.; Ahn, J.H.; Lim, H.S.; Eom, J.; Ryu, J.H. GOCI, the world's first geostationary ocean color observation satellite, for the monitoring of temporal variability in coastal water turbidity. *J. Geophys. Res. Ocean.* **2012**, *117*, C09004. [CrossRef]
35. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
36. Donlon, C.; Berruti, B.; Buongiorno, A.; Ferreira, M.H.; Féménias, P.; Frerick, J.; Goryl, P.; Klein, U.; Laur, H.; Mavrocordatos, C.; et al. The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. *Remote Sens. Environ.* **2012**, *120*, 37–57. [CrossRef]
37. Liu, R.J.; Zhang, J.; Cui, B.G.; Ma, Y.; Song, P.J.; An, J.B. Red Tide Detection Based on High Spatial Resolution Broad Band Satellite Data: A Case Study of GF-1. *J. Coast. Res.* **2019**, *90*, 120–128. [CrossRef]
38. Liu, R.; Xiao, Y.; Ma, Y.; Cui, T.; An, J. Red tide detection based on high spatial resolution broad band optical satellite data. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 131–147. [CrossRef]
39. Gordon, H.R.; Wang, M. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: A preliminary algorithm. *Appl. Opt.* **1994**, *33*, 443–452. [CrossRef] [PubMed]
40. Gordon, H.R. Atmospheric correction of ocean color imagery in the Earth Observing System era. *J. Geophys. Res. Atmos.* **1997**, *102*, 17081–17106. [CrossRef]
41. Ruddick, K.G.; Ovidio, F.; Rijkeboer, M. Atmospheric correction of SeaWiFS imagery for turbid coastal and inland waters. *Appl. Opt.* **2000**, *39*, 897–912. [CrossRef] [PubMed]
42. Wang, M.; Shi, W. The NIR-SWIR combined atmospheric correction approach for MODIS ocean color data processing. *Opt. Express* **2007**, *15*, 15722–15733. [CrossRef] [PubMed]

43. Doerffer, R.; Schiller, H. The MERIS Case 2 water algorithm. *Int*. *J*. *Remote Sens*. **2007**, *28*, 517–535. [CrossRef]
44. Shanmugam, P. CAAS: An atmospheric correction algorithm for the remote sensing of complex waters. *Ann*. *Geophys*. **2012**, *30*, 203–220. [CrossRef]
45. Hu, C.; Muller-Karger, F.E.; Taylor, C.; Carder, K.L.; Kelble, C.; Johns, E.; Heil, C.A. Red tide detection and tracing using MODIS fluorescence data: A regional example in SW Florida coastal waters. *Remote Sens*. *Environ*. **2005**, *97*, 311–321. [CrossRef]
46. Men, J.; Yao, T.; Yang, C.; Tian, L. A Neural Network-Based Atmospheric Correction Algorithm for GOCI Imagery Over Coastal Waters. *IEEE Trans*. *Geosci*. *Remote Sens*. **2024**, *62*, 1–16. [CrossRef]
47. Zhao, X.; Ma, Y.; Xiao, Y.; Liu, J.; Ding, J.; Ye, X.; Liu, R. Atmospheric correction algorithm based on deep learning with spatial-spectral feature constraints for broadband optical satellites: Examples from the HY-1C Coastal Zone Imager. *ISPRS J*. *Photogramm*. *Remote Sens*. **2023**, *205*, 147–162. [CrossRef]
48. Guo, X.; Ren, P.; Xiao, Y.; Zhao, X.; Song, Q.; Ye, X.; Kim, K.; Xie, Q.; Fan, M.; Ma, Y.; et al. A Spatial–Spectral Collaborative Method for Radiometric Cross-Calibration of Ocean Color Satellite Sensors. *IEEE Trans*. *Geosci*. *Remote Sens*. **2026**, *64*, 5530713. [CrossRef]
49. Volpe, G.; Colella, S.; Brando, V.E.; Forneris, V.; La Padula, F.; Di Cicco, A.; Sammartino, M.; Bracaglia, M.; Artuso, F.; Santoleri, R. Mediterranean ocean colour Level 3 operational multi-sensor processing. *Ocean Sci*. **2019**, *15*, 127–146. [CrossRef]
50. Morel, A.; Huot, Y.; Gentili, B.; Werdell, P.J.; Hooker, S.B.; Franz, B.A. Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. *Remote Sens*. *Environ*. **2007**, *111*, 69–88. [CrossRef]
51. Volpe, G.; Colella, S.; Forneris, V.; Tronconi, C.; Santoleri, R. The Mediterranean Ocean Colour Observing System—System development and product validation. *Ocean Sci*. **2012**, *8*, 869–883. [CrossRef]
52. Antoine, D.; Morel, A.; Gordon, H.R.; Banzon, V.F.; Evans, R.H. Bridging ocean color observations of the 1980s and 2000s in search of long-term trends. *J*. *Geophys*. *Res*. *Ocean*. **2005**, *110*, C06009. [CrossRef]
53. Stumpf, R.P.; Culver, M.E.; Tester, P.A.; Tomlinson, M.; Kirkpatrick, G.J.; Pederson, B.A.; Truby, E.; Ransibrahmanakul, V.; Soracco, M. Monitoring *Karenia brevis* blooms in the Gulf of Mexico using satellite ocean color imagery and other data. *Harmful Algae* **2003**, *2*, 147–160. [CrossRef]
54. Hu, C.; Feng, L. Modified MODIS fluorescence line height data product to improve image interpretation for red tide monitoring in the eastern Gulf of Mexico. *J*. *Appl*. *Remote Sens*. **2016**, *11*, 012003. [CrossRef]
55. Ahn, Y.H.; Shanmugam, P. Detecting the red tide algal blooms from satellite ocean color observations in optically complex Northeast-Asia Coastal waters. *Remote Sens*. *Environ*. **2006**, *103*, 419–437. [CrossRef]
56. Xie, M.; Qin, M.; Li, Y.; Liu, Z.; Gou, T. Experimental analysis on the optimal spectral index for the risk assessment of red tide occurrence. *J*. *Oceanol*. *Limnol*. **2025**, *43*, 735–745. [CrossRef]
57. O'Reilly, J.E.; Maritorena, S.; Mitchell, B.G.; Siegel, D.A.; Carder, K.L.; Garver, S.A.; Kahru, M.; McClain, C. Ocean color chlorophyll algorithms for SeaWiFS. *J*. *Geophys*. *Res*. *Ocean*. **1998**, *103*, 24937–24953. [CrossRef]
58. Lee, Z.; Carder, K.L.; Arnone, R.A. Deriving inherent optical properties from water color: A multiband quasi-analytical algorithm for optically deep waters. *Appl*. *Opt*. **2002**, *41*, 5755–5772. [CrossRef] [PubMed]
59. Maritorena, S.; Siegel, D.A.; Peterson, A.R. Optimization of a semianalytical ocean color model for global-scale applications. *Appl*. *Opt*. **2002**, *41*, 2705–2714. [CrossRef] [PubMed]
60. Smith, M.E.; Bernard, S. Satellite Ocean Color Based Harmful Algal Bloom Indicators for Aquaculture Decision Support in the Southern Benguela. *Front*. *Mar*. *Sci*. **2020**, *7*, 61. [CrossRef]
61. Babu, M.J.; Geetha, P.; Soman, K.P. MODIS-Aqua Data Based Detection and Classification of Algal Blooms along the Coast of India Using RLS Classifier. *Procedia Comput*. *Sci*. **2016**, *93*, 424–430. [CrossRef]
62. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect*. *Notes Comput*. *Sci*. **2015**, *9351*, 234–241. [CrossRef]
63. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
64. Cui, B.; Fang, X.; Lu, Y.; Huang, L.; Liu, R. RTDNet: Red tide detection network for high-resolution satellite images. *J*. *Image Graph*. **2023**, *28*, 3911–3921. [CrossRef]
65. Ding, T.; Xu, Z.; Wang, Y.; Hou, Q.; Liu, X.; Ma, F. Red Tide Detection Method Based on a Time Series Fusion Network Model: A Case Study of GOCI Data in the East China Sea. *Sensors* **2025**, *25*, 3455. [CrossRef] [PubMed]
66. Sun, K.; Jiang, G.; Wang, N.; Yu, D.; Teng, J.; Gao, S.; Huang, J.; Zhao, Z.; Song, Y.; Xin, L.; et al. A semantic segmentation network for red tide detection based on enhanced spectral information using HY-1C/D CZI satellite data. *Mar*. *Pollut*. *Bull*. **2025**, *215*, 117813. [CrossRef] [PubMed]
67. Zhou, L.; Li, H.; Chen, X.; Li, Z. SIG-Net: A Spectral-Index-Guided Network for Red Tide Extraction from Sentinel-2 Multispectral Imagery. *Remote Sens*. **2026**, *18*, 1928. [CrossRef]

68. Karpatne, A.; Atluri, G.; Faghmous, J.H.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.; Kumar, V. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2318–2331. [CrossRef]
69. Liu, R.; Cui, B.; Dong, W.; Fang, X.; Xiao, Y.; Zhao, X.; Cui, T.; Ma, Y.; Wang, Q. A refined deep-learning-based algorithm for harmful-algal-bloom remote-sensing recognition using *Noctiluca scintillans* algal bloom as an example. *J. Hazard. Mater.* **2024**, *467*, 133721. [CrossRef] [PubMed]
70. Carvalho, G.A.; Minnett, P.J.; Banzon, V.F.; Baringer, W.; Heil, C.A. Long-term evaluation of three satellite ocean color algorithms for identifying harmful algal blooms (*Karenia brevis*) along the west coast of Florida: A matchup assessment. *Remote Sens. Environ.* **2011**, *115*, 1–18. [CrossRef] [PubMed]
71. Carvalho, G.A.; Minnett, P.J.; Fleming, L.E.; Banzon, V.F.; Baringer, W. Satellite remote sensing of harmful algal blooms: A new multi-algorithm method for detecting the Florida Red Tide (*Karenia brevis*). *Harmful Algae* **2010**, *9*, 440–448. [CrossRef] [PubMed]
72. Hu, C.; Feng, L.; Lee, Z.; Davis, C.O.; Mannino, A.; McClain, C.R.; Franz, B.A. Dynamic range and sensitivity requirements of satellite ocean color sensors: Learning from the past. *Appl. Opt.* **2012**, *51*, 6045–6062. [CrossRef] [PubMed]
73. Zoffoli, M.L.; Pochic, V.; Bresciani, M.; Braga, F.; Lacour, T.; Retho, M.; Manach, S.; Röttgers, R.; Detoni, A.M.S.; Harmel, T.; et al. Toward elucidating accessory pigments in intense phytoplankton blooms using hyperspectral satellite remote sensing in support of harmful algal bloom (HAB) monitoring. *Remote Sens. Environ.* **2026**, *341*, 115439. [CrossRef]
74. Xie, Y.F.; Hou, X.; Ren, J.C.; Zhang, X.C.; Ma, C.C.; Zheng, J.B. Binary Quantization Vision Transformer for Effective Segmentation of Red Tide in Multispectral Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4202814. [CrossRef]
75. Lee, M.S.; Arrigo, K.; Smith, A.; Woodson, C.B.; Lee, J.; Micheli, F. Monitoring Harmful Algal Blooms in the Southern California Current Using Satellite Ocean Color and In Situ Data. *J. Mar. Sci. Eng.* **2025**, *13*, 2044. [CrossRef]
76. Wu, Q.; Kang, X.; Wang, Z.; Duan, P.; Li, S. Spectral and Spatial Prior-Guided Red Tide Detection for Optical Satellite Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 28192–28205. [CrossRef]
77. Westberry, T.K.; Siegel, D.A.; Subramaniam, A. An improved bio-optical model for the remote sensing of *Trichodesmium* spp. blooms. *J. Geophys. Res. Ocean.* **2005**, *110*, C06012. [CrossRef]
78. Fan, Q.C.; Liu, Y.; Liu, Y.M.; Yang, X.M.; Wang, Z.H. A Remote Sensing Monitoring System for Marine Red Tides Based on Targeted Negative Sample Selection Strategies. *J. Mar. Sci. Eng.* **2026**, *14*, 556. [CrossRef]
79. LaHaye, N.J.; Luis, K.M.; Gierach, M.M. Fusing Multi- and Hyperspectral Satellite Data for Harmful Algal Bloom Monitoring With Self-Supervised and Hierarchical Deep Learning. *Earth Space Sci.* **2026**, *13*, e2025EA004881. [CrossRef]
80. *JCGM 101:2008*; Evaluation of Measurement Data—Supplement 1 to the "Guide to the Expression of Uncertainty in Measurement"—Propagation of Distributions Using a Monte Carlo Method. Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures: Sèvres, France, 2008. [CrossRef]
81. Heckerman, D. *A Tutorial on Learning with Bayesian Networks; Technical Report MSR-TR-95-06*; revised November 1996; Microsoft Research: Redmond, WA, USA, 1995. Available online: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr-95-06.pdf (accessed on 22 October 2026).
82. Legleiter, C.J.; King, T.V.; Carpenter, K.D.; Hall, N.C.; Mumford, A.C.; Slonecker, T.; Graham, J.L.; Stengel, V.G.; Simon, N.; Rosen, B.H. Spectral mixture analysis for surveillance of harmful algal blooms (SMASH): A field-, laboratory-, and satellite-based approach to identifying cyanobacteria genera from remotely sensed data. *Remote Sens. Environ.* **2022**, *279*, 113089. [CrossRef]
83. Kumar, A.; Maniyar, C.B.; Tesfayi, N.; Grunert, B.K.; Fiorentino, I.R.; Herweck, K.; Hyland, E.; Liu, B.; Bartelme, B.; Mishra, D.R. Advancing CyanoHAB monitoring with hyperspectral data from NASA PACE: First results and validation. *Int. J. Appl. Earth Obs. Geoinf.* **2026**, *146*, 105032. [CrossRef]
84. Shi, J.; Qin, H.; Cheng, J.; Xu, J.; Liu, G.; Li, Z.; Song, K. Real-time detection of harmful algal blooms using GOCI II and buoy-based dissolved oxygen variations. *Environ. Res. Lett.* **2025**, *20*, 074007. [CrossRef]
85. Shi, X.; Gu, L.; Li, X.; Jiang, T.; Gao, T. Automated spectral transfer learning strategy for semi-supervised regression on Chlorophyll-a retrievals with Sentinel-2 imagery. *Int. J. Digit. Earth* **2024**, *17*, 2313856. [CrossRef]
86. Çölkesen, I.; Saygı, M.; Öztürk, M.Y.; Altuntaş, O.Y. U-shaped deep learning networks for algal bloom detection using Sentinel-2 imagery: Exploring model performance and transferability. *J. Environ. Manag.* **2025**, *381*, 125152. [CrossRef] [PubMed]
87. Caballero, C.B.; Martins, V.S.; Paulino, R.S.; Butler, E.; Sparks, E.; Lima, T.M.; Novo, E.M.L.M. The need for advancing algal bloom forecasting using remote sensing and modeling: Progress and future directions. *Ecol. Indic.* **2025**, *172*, 113244. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.