

Systematic Review

Beyond Accuracy: Transferability Limits, Validation Inflation, and Uncertainty Gaps in Satellite-Based Water Quality Monitoring—A Systematic Quantitative Synthesis and Operational Framework

Saeid Pourmorad ^{1,*}, Valerie Graw ², Andreas Rienow ² and Luca Antonio Dimuccio ¹

¹ University of Coimbra, Centre of Studies in Geography and Spatial Planning (CEGOT), Department of Geography and Tourism, Faculty of Arts and Humanities, 3004-530 Coimbra, Portugal; luca@ci.uc.pt

² Institute of Geography, Ruhr-University Bochum (RUB), D-44780 Bochum, Germany; valerie.graw@ruhr-uni-bochum.de (V.G.); andreas.rienow@ruhr-uni-bochum.de (A.R.)

* Correspondence: omid2red@gmail.com

Highlights

What are the main findings?

- Satellite-based water quality monitoring is heavily influenced by sensor performance, validation design, and model transferability, with substantial variability in reported accuracies across different studies.
- A lack of standardised uncertainty quantification and inconsistent validation practices continues to hinder the reliability and scalability of satellite-derived water quality models.

What are the implications of the main findings?

- These findings underscore the need for a unified framework that integrates robust validation protocols, multi-sensor harmonisation, and uncertainty-aware modelling to ensure accurate, transferable, and decision-grade environmental monitoring.
- To address challenges in transferability and operational deployment, the research highlights the importance of physics-informed models and standardised uncertainty reporting in developing scalable, reliable water quality monitoring systems.

Abstract

Satellite remote sensing has become essential for water quality assessment across inland and coastal environments, with rapid improvements in recent years. Significant advances have been made in detecting optically active parameters (such as chlorophyll-a, suspended matter, and turbidity), showing consistently strong performance across multiple studies. Specifically, the median validation performance (R^2) derived from the quantitative synthesis indicates $R^2 = 0.82$ for chlorophyll-a (interquartile range—IQR: 0.75–0.90), $R^2 = 0.80$ for total suspended matter (IQR: 0.78–0.85), and $R^2 = 0.88$ for turbidity (IQR: 0.85–0.90). Conversely, the retrieval of optically inactive parameters (such as nutrients like total phosphorus and total nitrogen) remains more context dependent. It exhibits moderate, more variable results, with median $R^2 = 0.68$ (IQR: 0.64–0.74) for total phosphorus and $R^2 = 0.75$ (IQR: 0.70–0.80) for total nitrogen. These findings clearly illustrate the varying success of retrievals of optically active and inactive parameters and underscore the inherent difficulties of indirect estimation methods. However, high reported accuracy has yet to translate into transferable, uncertainty-informed, and operational monitoring systems. This gap stems from structural issues in validation design, physics integration, uncertainty management,



Academic Editors: Xiaodong Li and Frédéric Frappart

Received: 25 February 2026

Revised: 31 March 2026

Accepted: 1 April 2026

Published: 7 April 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

and multi-sensor compatibility rather than data limitations alone. We present a PRISMA-guided, distribution-aware quantitative synthesis of 152 peer-reviewed studies (1980–2025), based on a systematic search protocol, to evaluate satellite-based retrievals of both optically active and inactive parameters. Instead of simply averaging performance, we analyse the empirical distributions of validation metrics, considering the validation protocol, sensor type, parameter category, degree of physics integration, and uncertainty quantification. The synthesis demonstrates that validation strategy often influences reported results more than the algorithm class itself, with accuracy inflated under non-independent cross-validation methods and notable variability between studies concealed by mean-based reports. Across four decades, four persistent structural challenges remain: limited transferability across sites and sensors beyond calibration areas; weak or implicit physical integration in many data-driven models; lack of or inconsistency in uncertainty quantification; and fragmented multi-sensor harmonisation that restricts operational scalability. To address these issues, we introduce two evidence-based coding frameworks: a physics-integration taxonomy (P0–P4) and an uncertainty-quantification hierarchy (U0–U4). Applying these frameworks shows that most studies remain focused on low-to-moderate levels of physics integration and primarily consider uncertainty at the prediction stage, with limited attention to upstream sources throughout the observation and inference process. Building on this structured synthesis, we propose a transferable, physics-informed, and uncertainty-aware conceptual framework that links model architecture, validation robustness, and probabilistic uncertainty to well-founded design principles. By shifting satellite water quality modelling from isolated algorithm demonstrations towards integrated, evidence-based system design, this study promotes scalable, decision-grade environmental monitoring amid the accelerating impacts of climate change.

Keywords: satellite remote sensing; water quality retrieval; model transferability; independent validation; uncertainty quantification; physics-informed machine learning; multi-sensor integration; environmental monitoring systems

1. Introduction

Escalating freshwater demand driven by population growth, industrial expansion, and climate-induced hydrological change has imposed unprecedented pressure on inland and coastal aquatic systems worldwide. These pressures are further intensified by rapid and spatially heterogeneous warming of lake surface waters, which modifies stratification regimes, alters biogeochemical cycling, and reduces ecosystem resilience, thereby increasing susceptibility to water quality degradation and harmful algal bloom events [1–3]. Under these converging stressors, water quality (WQ) monitoring has evolved from a primarily observational scientific activity into a critical infrastructure for public health protection, regulatory compliance, and adaptive water resources governance. Although in situ sampling remains the analytical reference standard supported by established laboratory protocols and quality-assurance frameworks [4], cost, logistics, and limited spatiotemporal coverage inherently constrain field-based monitoring networks. These constraints limit their ability to resolve the high heterogeneity and rapid dynamics of aquatic systems at scales relevant to operational decision-making. Satellite remote sensing has therefore emerged as a cornerstone of contemporary WQ monitoring, enabling synoptic, repeatable, and cost-effective observations across spatial and temporal domains inaccessible to conventional sampling alone [5,6]. Recent studies have further highlighted rapid advances in satellite-based

monitoring frameworks, including the integration of machine learning, multi-platform observations, and data-driven modelling approaches for water quality assessment [7–10].

Early applications using Landsat-class sensors demonstrated the feasibility of retrieving optically active constituents such as chlorophyll-a and suspended particulate matter in both inland and coastal waters [11,12]. The exploitation of long-term satellite archives subsequently enabled multi-decadal assessments of water clarity and trophic-state dynamics across thousands of lakes and reservoirs [13,14]. Continued advances in sensor design have expanded monitoring capabilities across both spatial and temporal dimensions, supporting event-scale analyses of turbidity pulses and bloom dynamics while extending surveillance to smaller and more morphologically complex inland systems [15–18]. Complementing satellite observations, UAV-borne and proximal sensing systems increasingly provide ultra-high-resolution data that strengthen calibration and validation, particularly in optically complex or atmospherically challenging environments [19–22]. Recent developments further demonstrate the growing role of multi-sensor satellite constellations, Sentinel-2-based retrieval strategies, and hybrid modelling approaches in improving water quality monitoring performance across diverse aquatic environments [23–26].

From a process-based perspective, water quality parameters retrieved via remote sensing can be broadly categorised into optically active parameters, which directly influence water-leaving reflectance (e.g., chlorophyll-a, suspended particulate matter, turbidity), and optically inactive parameters, which are not directly observable but can be inferred through indirect pathways such as biogeochemical coupling, proxy relationships, or auxiliary environmental data. While optically active parameters have reached a relatively mature stage of retrieval with increasingly consistent performance across sensors and environments, recent studies have expanded the scope of satellite-based monitoring to include optically inactive parameters such as total nitrogen (TN), total phosphorus (TP), and dissolved oxygen (DO), often through machine learning and hybrid modelling frameworks [7,8,27,28]. Nevertheless, the retrieval of these optically inactive parameters remains fundamentally dependent on context-specific proxy relationships and multi-source data integration, leading to lower transferability and greater structural uncertainty than for optically active constituents. In this study, the selection of water quality parameters is explicitly justified based on three criteria: (i) optical observability and spectral relevance, ensuring representation of parameters that directly or indirectly influence reflectance; (ii) biogeochemical significance, capturing key processes governing aquatic ecosystem dynamics (e.g., primary productivity, sediment transport, nutrient cycling); and (iii) prevalence in the satellite remote sensing literature, ensuring sufficient empirical evidence for systematic synthesis. This principled selection framework enables a coherent comparison between optically active and inactive parameter retrievals while preserving relevance to both scientific understanding and operational monitoring.

Despite substantial advances in observational capacity, the principal constraints on satellite-based WQ monitoring increasingly arise not from data availability but from how models are designed, validated, and generalised across sites, sensors, and spatiotemporal contexts. Reported model performance often reflects site-specific calibration strategies, validation designs, and sensor-dependent preprocessing choices rather than the model's intrinsic capability, introducing systematic biases that inflate reported accuracy and limit cross-study comparability and operational transferability. Furthermore, uncertainty is frequently treated implicitly or omitted altogether, constraining interpretability and decision relevance. Recent literature has emphasised these persistent challenges, particularly regarding model generalisation, uncertainty quantification, and the robustness of machine-learning-based approaches across heterogeneous aquatic environments [8,28,29]. In this sense, satellite-derived WQ retrieval provides a high-impact testbed for broader

environmental modelling challenges, where validation design, physical consistency, and uncertainty representation are as critical as algorithmic performance. A central dimension of this challenge lies in the heterogeneity of contemporary sensing platforms and their operational trade-offs. Monitoring frameworks now integrate long-archive moderate-resolution missions (e.g., Landsat), high-resolution constellations (e.g., Sentinel-2/3), climate-scale sensors (e.g., MODIS/VIIRS), emerging missions such as SDGSAT-1 and GCOM-C/SGLI, UAV hyperspectral systems, and ground-based IoT spectrometers. Each platform contributes distinct strengths while introducing specific limitations related to spatial resolution, revisit frequency, atmospheric correction, and calibration consistency. These characteristics fundamentally condition model transferability and operational scalability.

Crucially, the heterogeneity summarised in Table 1 propagates directly into modelling uncertainty and limits cross-study comparability. Differences in spatial resolution, revisit frequency, radiometric stability, and atmospheric-correction dependencies shape match-up datasets and determine whether models can generalise beyond calibration domains. In parallel with the evolution of sensing, modelling paradigms for satellite-based WQ retrieval have undergone a fundamental transition. Early empirical and semi-analytical approaches grounded in spectral indices and radiative-transfer formulations [30–32] have progressively been complemented by machine-learning (ML) and deep-learning (DL) approaches capable of capturing nonlinear relationships in optically complex waters [33,34]. While these approaches often demonstrate strong in-domain performance [35,36], their generalisation remains constrained, particularly under non-independent validation schemes that may inflate reported accuracy. To enable a structured, cross-paradigm evaluation, dominant modelling approaches are synthesised with respect to transferability, interpretability, uncertainty quantification, and operational readiness.

Table 1. Critical comparison of major remote sensing platforms for water quality monitoring, highlighting their complementary strengths, limitations, and implications for transferability and operational deployment.

Sensor Platform	Primary Strengths	Key Limitations	Implications for Transferability & Operations	Representative References
Landsat Series (TM, ETM+, OLI)	Long-term, radiometrically stable archive (since 1984); enables historical trend analysis and interannual variability assessment.	Moderate spatial resolution (~30 m) limits applicability in narrow rivers, small ponds, and heterogeneous shorelines.	High temporal transferability and cross-era consistency; limited spatial transferability in small or fragmented systems; well-suited as a baseline sensor in multi-sensor fusion frameworks.	[37,38]
Sentinel-2/Sentinel-3 (MSI/OLCI)	High spatial (10–60 m) and temporal (2–5 days) resolution; robust for coastal and inland waters; strong sensitivity to chl-a and turbidity.	Sentinel-3's coarse resolution (~300 m) restricts use in small lakes; Sentinel-2 lacks thermal bands, constraining surface temperature retrievals.	Strong spatial transferability for inland and coastal waters; complementary pairing (MSI-OLCI) enhances cross-scale monitoring but requires harmonisation and AC consistency.	[39–42]
MODIS/VIIRS	Daily global coverage; indispensable for long-term trend analysis, anomaly detection, and large-scale climatic assessments.	Very coarse spatial resolution (~250–1000 m), restricting applicability to large lakes, reservoirs, and open coastal waters.	High temporal robustness but limited spatial transferability; best used for synoptic context or as temporal constraints in multi-sensor fusion rather than standalone retrieval.	[43–45]
UAV/Hyperspectral Imaging (HSI)	Ultra-high spatial and spectral resolution; flexible deployment for targeted, on-demand monitoring and algorithm calibration.	Limited spatial coverage; high operational cost; strong sensitivity to atmospheric and illumination conditions.	Excellent local generalisation and interpretability; poor scalability; primarily suited for calibration, validation, and transfer learning support rather than regional monitoring.	[21,22,36,46]

Table 1. Cont.

Sensor Platform	Primary Strengths	Key Limitations	Implications for Transferability & Operations	Representative References
SDGSAT-1 (MII)	High signal-to-noise ratio ($\approx 4\text{--}7 \times$ Sentinel-2 MSI); 10 m spatial resolution; improved performance in turbid inland and coastal waters.	Broader spectral bands reduce sensitivity to specific pigments (e.g., chl-a); calibration and atmospheric correction remain evolving.	Promising for regional transferability in turbid regimes; requires further cross-sensor benchmarking and AC standardisation for operational adoption.	[47]
GCOM-C/SGLI	Long-term global reflectance and WQP (chl-a, TSM) products capture interannual and seasonal variability.	Blue-band biases; underestimation at high chl-a/TSM concentrations; performance sensitive to AC scheme selection.	Suitable for large-scale comparative analyses; limited reliability in optically complex inland waters without regional recalibration.	[48]
Ground-based IoT Hyperspectral Systems	Real-time, high-frequency monitoring via fixed or semi-autonomous spectrometers; critical for validation and uncertainty calibration.	Fixed spatial footprint; high instrumentation cost; maintenance requirements.	Essential for uncertainty quantification and model calibration; complements satellite/UAV data but is not a standalone spatial monitoring solution.	[49]

Despite substantial methodological progress, the comparative synthesis presented in Table 2 reveals that the evolution of modelling paradigms has not translated into balanced advances across the key dimensions required for operational deployment. Instead, systematic asymmetries persist across transferability (T), interpretability and physical consistency (I), uncertainty quantification (UQ), and operational readiness (O), which collectively constrain real-world applicability. First, limited transferability remains a dominant constraint, as reflected by the predominantly partial or weak performance ($\Delta/\times$) across most modelling paradigms in Table 2. Even advanced ML and DL approaches, while powerful at capturing nonlinear relationships, exhibit strong dependence on training-domain characteristics and frequently degrade in cross-region, cross-sensor, or cross-season applications [6,50]. Second, insufficient physical interpretability and integration are evident in the contrast between physics-based and data-driven models in Table 2. While physics-based approaches achieve high interpretability ($\checkmark$), they suffer from scalability and operational limitations, whereas ML/DL models prioritise predictive performance at the expense of mechanistic transparency ($\times$). This structural imbalance limits both scientific interpretability and robustness under changing environmental conditions. Third, uncertainty quantification remains fragmented and inconsistently implemented, as indicated by the limited explicit treatment of UQ across most model classes in Table 2. Even when uncertainty is addressed (e.g., probabilistic or ensemble approaches), it is often limited to prediction-level variability. It fails to capture upstream sources of uncertainty, thereby constraining its utility for decision-making [51,52]. Fourth, operational readiness and multi-sensor integration remain underdeveloped, with Table 2 highlighting that only a subset of approaches (primarily empirical or simplified models) achieves high deployability ($\checkmark$), often at the cost of reduced generalizability or physical consistency. Conversely, more advanced or hybrid approaches, while theoretically promising, remain difficult to standardise and scale across heterogeneous observational systems [5].

Beyond these modelling-related constraints, an additional and fundamentally interacting challenge lies in the optical heterogeneity of aquatic systems. Variations in constituent composition, bottom reflectance, and water–atmosphere interactions give rise to distinct optical regimes that directly condition model performance [32]. Optical Water Type (OWT) frameworks provide a principled basis for stratifying this variability [53], yet as implicitly reflected in the uneven performance patterns across Table 2, their integration into model design, validation strategies, and uncertainty characterisation remains limited and non-systematic. Against this evidence-based backdrop (Table 2), the present study advances

three primary contributions: (1) a PRISMA-guided, distribution-aware quantitative synthesis of 152 peer-reviewed studies that moves beyond mean-based evaluation toward empirical performance distributions; (2) the systematic operationalisation of transferability, validation design, physics integration, and uncertainty through structured coding frameworks; and (3) the development of a transferable, physics-informed, and uncertainty-aware conceptual framework that explicitly links sensing configurations, modelling strategies, and validation protocols into a coherent and operationally consistent design paradigm. By reframing satellite water quality monitoring from isolated algorithmic performance toward an integrated, evidence-based system perspective grounded in the comparative insights of Table 2, this study establishes a reproducible foundation for scalable, uncertainty-aware, and decision-grade environmental observation [6,54].

Table 2. Analytical comparison of major methodological paradigms for satellite-based water quality monitoring, evaluating their respective contributions and limitations with respect to four key research gaps: transferability (T), interpretability and physical consistency (I), uncertainty quantification (UQ), and operational readiness (O). For each paradigm, ✓ indicates aspects that are explicitly addressed in the literature, △ denotes partial or context-dependent treatment, and ✗ indicates dimensions that remain largely unaddressed.

Methodological Paradigm	Core Contribution	Primary Limitations	Relevance to Key Gaps (T/I/UQ/O)	Representative References
Empirical Models	Simple and computationally efficient; foundational for early remote sensing of water quality and baseline algorithm development.	Strong site- and sensor-dependence; limited scalability and transferability without frequent recalibration.	T: ✗ (site-specific) I: △ (partially interpretable) UQ: ✗ O: ✓ (easy to deploy)	[37,55,56]
Physics-Based Models	Mechanistically grounded in radiative transfer theory, it provides physically interpretable and scientifically transparent retrievals.	Require extensive ancillary inputs; computationally demanding; difficult to scale and operationalise across regions.	T: △ (physics-consistent but data-limited) I: ✓ UQ: △ (model-based sensitivity) O: ✗	[57,58]
Machine Learning (ML)	Effectively capture nonlinear relationships; robust to noise; relatively efficient training with moderate data volumes.	Generalisation is often limited beyond calibration domains; uncertainty is typically implicit or unreported.	T: △ I: ✗ UQ: ✗ O: △	[59–63]
Deep Learning (DL)	Automated feature learning; strong performance in optically complex waters; enables super-resolution and spatiotemporal fusion.	Requires large, labelled datasets; limited interpretability; high computational cost; transferability rarely demonstrated.	T: △ I: ✗ UQ: △ (via ensembles or Bayesian DL) O: △	[36,64–68]
Hybrid & Physics-Informed Approaches	Combine physical constraints with data-driven learning; improve interpretability and robustness; promising for cross-domain generalisation.	Methodologically complex; lacks standardised architectures and benchmarks; limited large-scale validation.	T: ✓ I: ✓ UQ: △ O: △	[69,70]
Probabilistic & Uncertainty-Aware Models	Explicitly quantify predictive uncertainty; support risk-aware and decision-grade applications.	Increased model complexity, computational overhead, and limited adoption in operational pipelines.	T: △ I: △ UQ: ✓ O: △	[39,71,72]
Hybrid & Novel Indices	Expand anomaly detection and multi-source integration; useful for event-focused and exploratory analyses.	Generalisation across anomaly types and environmental contexts remains limited.	T: ✗ I: △ UQ: ✗ O: △	[73–75]

Despite substantial methodological progress, a consistent set of system-level barriers continues to impede the transition from research-oriented demonstrations to transferable and decision-grade monitoring systems. Foremost among these is limited transferability: models calibrated for specific water bodies, optical regimes, or sensor configurations frequently exhibit marked performance degradation when applied across regions, seasons, or observational platforms [6,8,29]. Closely related is the challenge of interpretability and physical consistency, particularly for purely data-driven models whose internal represen-

tations may obscure mechanistic relationships between reflectance and constituents. A third barrier concerns insufficiently standardised uncertainty quantification, with many studies reporting deterministic metrics without predictive intervals or calibration diagnostics [7,38,39]. Finally, an operationalisation gap persists, encompassing fragmented multi-sensor integration, limited interoperability between satellite, UAV, and in situ observations, and the scarcity of standardised near-real-time processing pipelines [5]. An additional, underexplored dimension of this challenge is the pronounced optical diversity between inland and coastal waters. Variability in constituent composition, bottom reflectance, and water–atmosphere interactions gives rise to distinct optical regimes that fundamentally condition the performance of algorithms [32]. Optical Water Type (OWT) frameworks offer a principled mechanism for stratifying this diversity [40], yet their integration into model design, validation, and uncertainty reporting remains inconsistent. Against this backdrop, the present study conducts a comprehensive systematic review and quantitative evidence synthesis of satellite-based water quality monitoring studies published between 1980 and 2025, encompassing 152 peer-reviewed articles. Treating WQ retrieval as an instance of a broader environmental modelling problem, the analysis interrogates transferability, validation robustness, physics integration, uncertainty characterisation, and optical-regime diversity as explicit empirical variables rather than assumed properties. Building directly on the evidence-based synthesis, the study articulates a transferable, physics-informed, and uncertainty-aware conceptual framework for operational WQ monitoring, accompanied by an open-source implementation roadmap (Appendix A). Collectively, this work advances satellite-derived water quality monitoring beyond algorithm-centric evaluation toward a structured, reproducible, and operationally defensible environmental modelling paradigm [6,41].

2. Methodology

This systematic review was designed and conducted in full accordance with the PRISMA 2020 reporting guidelines. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) is a structured reporting framework developed to improve the transparency, completeness, and reproducibility of systematic reviews by explicitly guiding the identification, screening, eligibility assessment, and final inclusion of studies. In the present review, PRISMA was used not merely as a reporting checklist but as an organising methodological framework to ensure traceable evidence selection and consistent documentation across all stages of the review process. A review protocol was developed a priori to define the search strategy, eligibility criteria, quality appraisal framework, and standardised data-extraction schema before the initiation of screening. Although the protocol was not formally registered in PROSPERO or OSF, it was finalised before literature screening and is provided verbatim in Supplementary Protocol S4 to ensure methodological traceability and allow independent verification. The finalised protocol, together with the complete per-study and configuration-level dataset comprising 152 peer-reviewed articles, is provided in Supplementary Table S1 (Excel format) to support transparency, reuse, replication, and secondary analysis. A completed PRISMA checklist and PRISMA flow diagram are also included in the Supplementary Materials to document full adherence to the PRISMA 2020 framework.

Importantly, while recent studies (2022–2026) are cited in the Introduction to provide an up-to-date contextual overview of advances in satellite-based water quality monitoring, the systematic review dataset was defined using a fixed search period in line with the PRISMA protocol. Accordingly, only studies meeting the predefined eligibility criteria within this search window were included in the quantitative synthesis, resulting in a total of 152 peer-reviewed articles. This distinction ensures methodological consistency while

maintaining an up-to-date scientific context. The overarching methodological objective was to construct a rigorously traceable empirical synthesis of satellite-based water quality (WQ) monitoring studies published between 1980 and August 2025, to identify structural and validation-related research gaps, and to establish a defensible analytical foundation for transferable and uncertainty-aware monitoring frameworks. To operationalise this objective, the review was implemented through six interlinked components aligned with the PRISMA workflow: (i) systematic literature search; (ii) screening and eligibility assessment; (iii) reviewer reliability and bias mitigation; (iv) structured quality appraisal; (v) standardised data extraction and coding; and (vi) transparent workflow documentation.

2.1. Search Strategy

A comprehensive and fully reproducible literature search was conducted across four major scholarly databases: Web of Science Core Collection, Scopus, IEEE Xplore, and PubMed. These databases were selected to ensure complementary coverage of environmental sciences, remote sensing, engineering, and applied geosciences. Google Scholar was used exclusively for backward and forward citation chaining to identify additional eligible peer-reviewed journal articles not retrieved through primary database searches. Grey literature, including technical reports, theses, and conference proceedings, was excluded except where necessary to trace subsequent peer-reviewed journal publications. The temporal scope extended from January 1980 to 20 August 2025, capturing the full methodological evolution of satellite-based WQ monitoring, from early multispectral Landsat applications to contemporary multi-sensor, machine-learning (ML), deep-learning (DL), and uncertainty-aware modelling frameworks [22]. Search queries were constructed using structured Boolean logic and clustered thematic terms organized into four concept groups: (i) remote sensing and Earth observation; (ii) inland and coastal water systems; (iii) water quality parameters such as chlorophyll-a, turbidity, total suspended matter, colored dissolved organic matter, and nutrients; and (iv) modelling paradigms including empirical regression, physics-based and radiative-transfer approaches, ML, DL, and hybrid frameworks. Synonyms, abbreviations, and alternative spellings (e.g., “chl-a”) were systematically incorporated to maximise recall while maintaining thematic specificity. Complete database-specific search strings, field tags, applied filters, and execution dates are reported verbatim in Supplementary Protocol S4.

Database searches yielded 2184 records (Web of Science: 839; Scopus: 620; IEEE Xplore: 420; PubMed: 305), with an additional 33 records identified through citation screening. After removal of 161 duplicate records and 38 retracted publications, 2018 unique records proceeded to title and abstract screening. Of these, 1682 were excluded based on predefined criteria, and 336 articles underwent full-text assessment. Following full-text evaluation, 152 peer-reviewed studies satisfied all inclusion criteria and were retained for qualitative synthesis. The process is summarised in Figure 1, which illustrates the PRISMA flow diagram of literature search and study selection.

2.2. Screening and Eligibility Criteria

Eligibility criteria were defined using a structured PEO framework adapted for environmental evidence synthesis. The population/problem dimension comprised inland and coastal aquatic systems. The exposure dimension was restricted to optical remote sensing platforms, including satellite, UAV, and airborne systems. The outcome dimension required explicit quantitative validation of retrieved water quality parameters against in situ measurements or independent reference datasets. Eligible study types were limited to peer-reviewed journal articles providing sufficient methodological transparency regarding sensor characteristics, modelling approach, validation protocol, and reported performance

metrics. Screening proceeded in two sequential phases. Title and abstract screening excluded studies that were not relevant to water quality monitoring, used non-optical sensing approaches, were not in English, were not peer-reviewed, or had incomplete reporting. Full-text eligibility required explicit reporting of modelling methodology and quantitative validation metrics (e.g., R^2 , RMSE, MAE, NSE) derived from independent or in situ datasets. Studies with fewer than ten validation observations, insufficient methodological transparency, or absent validation reporting were excluded to preserve analytical credibility.

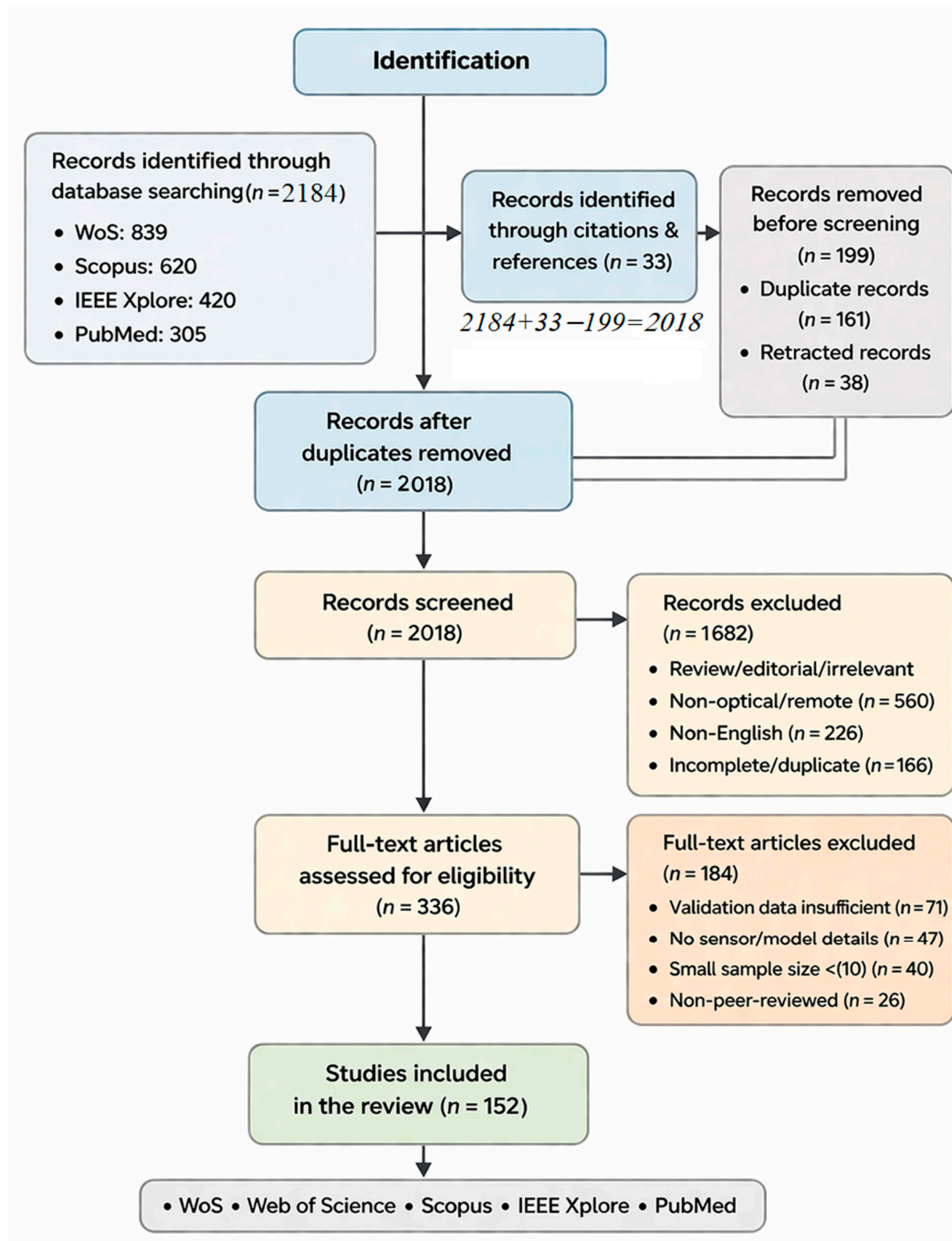


Figure 1. PRISMA flow diagram of literature search and study selection. The diagram summarises the identification, screening, eligibility assessment, and final inclusion of studies in this systematic review. Records were retrieved from Web of Science, Scopus, IEEE Xplore, and PubMed, complemented by citation and reference screening. After duplicate and retracted records were removed, titles and abstracts were screened, followed by full-text eligibility assessment based on predefined inclusion criteria. A total of 152 peer-reviewed studies were included in the qualitative synthesis.

2.3. Reviewer Reliability and Bias Mitigation

All screening and eligibility assessments were conducted independently by two reviewers. Disagreements were resolved through structured discussion and, when necessary, arbitration by a third reviewer. Inter-rater reliability was evaluated on a randomly selected 10% subset of screened records, yielding Cohen's $\kappa = 0.82$, indicating strong agreement. To further ensure methodological rigour, the reproducibility of the structured coding dimensions introduced in this review was assessed independently. For the physics-integration coding scheme (P0–P4) and uncertainty-awareness classification (U0–U4), two reviewers independently coded a randomly selected 15% subset of included studies. Agreement statistics yielded $\kappa_P = 0.87$ and $\kappa_U = 0.84$, indicating substantial to near-perfect agreement. Discrepancies were resolved through consensus review using predefined operational definitions. Beyond reviewer agreement, multiple bias-mitigation measures were implemented. Publication bias was minimised by including all relevant peer-reviewed articles irrespective of reported performance levels. Language bias was reduced by screening non-English abstracts where available. Additionally, retracted publications were explicitly identified and excluded before screening to prevent contamination of the evidence base.

2.4. Quality Appraisal and Domain-Specific Risk-of-Bias Assessment

All included studies underwent structured quality appraisal using the Critical Appraisal Skills Program (CASP) framework. Assessment focused on methodological transparency, adequacy and independence of validation data, clarity and consistency of performance reporting, and reproducibility of modelling workflows. In addition to CASP, a remote-sensing-specific risk-of-bias checklist was applied to capture domain-relevant methodological risks frequently identified in satellite-based WQ literature [5,6,76]. This domain-specific appraisal evaluated the independence of the validation design (e.g., avoidance of spatial/temporal leakage in cross-validation), the transparency of atmospheric correction procedures, the clarity of match-up protocols between satellite and in situ data, sample-size adequacy, the reporting of preprocessing pipelines, and the explicit handling of uncertainty and calibration diagnostics. These criteria were used to inform interpretive weighting and sensitivity analyses, but did not serve as exclusion thresholds. Each study was categorised into High, Moderate, or Low-quality tiers based on the combined CASP and domain-specific appraisal (Table 3).

CASP ratings and domain-specific risk flags were incorporated into sensitivity analyses (Section 4.2) to evaluate whether comparative performance patterns remained stable when lower-quality or higher-risk studies were excluded.

2.5. Data Extraction, Coding, and Quantitative Synthesis Design

A standardised data-extraction protocol was implemented to support stratified and distribution-aware quantitative synthesis. For each eligible article, structured variables were extracted covering bibliographic metadata, geographic context, climate zone, water-body type, sensor platform, retrieved parameter(s), modelling paradigm (empirical, physics-based, ML, DL, hybrid), validation strategy (random cross-validation, spatial holdout, temporal holdout, independent in situ match-up), reported performance metrics (R^2 , RMSE, MAE, NSE), and explicitly reported methodological limitations or uncertainty sources. Data extraction was conducted at the model-configuration level, defined as the unique combination of study, sensor input, target parameter, modelling approach, and validation protocol. This preserves analytical granularity while maintaining full traceability to the originating publication. Because configuration-level observations within a single study are not statistically independent, within-study dependency was explicitly controlled. Configuration-level results were not treated as independent effect sizes. Study

identity was retained throughout all analyses, and quantitative synthesis relied on stratified distribution-based summaries rather than pooled means to avoid pseudo-replication. Potential datasets that overlapped across studies were assessed using rule-based provenance tracking based on reported study locations, temporal coverage, sensor acquisition periods, and data sources. When substantial overlap was identified, unique Dataset-IDs were assigned and flagged in Supplementary Table S3 to prevent artificial inflation of evidence. Given substantial heterogeneity in reported performance metrics, validation designs, sample sizes, and reporting conventions across the literature, formal effect-size meta-analysis or meta-regression was deemed methodologically inappropriate. Reported metrics were frequently derived from non-comparable validation schemes, heterogeneous sample sizes, and selective reporting of best-performing configurations, introducing potential best-model reporting bias. Accordingly, a distribution-aware quantitative synthesis approach was adopted, emphasising medians, interquartile ranges, and dispersion patterns within stratified groups [22]. This design preserves between-study variability and avoids over-interpretation of heterogeneous evidence while maintaining consistency with prevailing reporting practices in satellite-based WQ research. The complete extraction database, including Dataset-IDs and coding variables, is provided in Supplementary Table S1 to ensure full methodological transparency.

Table 3. Quality appraisal was conducted using the Critical Appraisal Skills Programme (CASP) checklist, focusing on transparency, validation adequacy, reporting clarity, and reproducibility. While all studies that met the inclusion criteria were retained for qualitative synthesis, sensitivity analyses were performed to assess the influence of lower-quality studies on quantitative performance summaries.

Quality Tier	Number of Studies (n)	Percentage of Total (%)	Interpretation for Quantitative Synthesis
High	$n(\text{High})$	$(n(\text{High})/152) \times 100$	Comprehensive methodological reporting; independent or well-defined validation; high reproducibility. Results are considered highly reliable for comparative synthesis.
Moderate	$n(\text{Moderate})$	$(n(\text{Moderate})/152) \times 100$	Adequate methodological detail and validation, with minor limitations in reporting or reproducibility. Included in synthesis with caution in interpretation.
Low	$n(\text{Low})$	$(n(\text{Low})/152) \times 100$	Limited transparency and/or weak validation design. Retained for qualitative completeness but evaluated separately in sensitivity analyses.
Total	152	100	—

2.6. Workflow Transparency

All stages of identification, screening, eligibility assessment, coding, risk-of-bias appraisal, and data handling adhered strictly to PRISMA 2020 standards. Detailed documentation of decision rules, operational definitions, coding frameworks (P0–P4; U0–U4), provenance-tracking procedures, and sensitivity-analysis logic is provided in the Supplementary Materials to ensure complete methodological traceability and reproducibility.

3. Results/Thematic Synthesis

Based on a systematic review of 152 peer-reviewed studies, this section synthesises evidence at the model-configuration level to identify structural patterns that condition the robustness and operational credibility of satellite-based water quality (WQ) monitoring. Rather than cataloguing individual results, the synthesis integrates distributions of

reported performance and study designs across sensor classes, target parameters, modelling paradigms, and validation protocols, with prior reviews used only to contextualise recurring themes [19,77]. Evidence maps and stratified meta-summaries reveal consistent co-occurrence between modelling choices and reported skill, including (i) strong dependence of reported accuracy on validation design (e.g., internal cross-validation versus spatial/temporal hold-out), (ii) uneven evidence coverage across sensors and parameters, and (iii) a persistent disconnect between methodological sophistication and explicit uncertainty reporting. The organising logic for the results is summarised in Figure 2, which provides an end-to-end conceptual sequence linking (a) multi-source observations (satellite, UAV, in situ), (b) preprocessing and feature construction, (c) model structure and physics integration, (d) validation design, and (e) uncertainty representation and decision-support suitability [78]. The subsequent subsections follow this sequence: we first characterise the evidence base in terms of sensor platforms and study geography, then summarise parameter-dependent retrievability and model families, and finally quantify how validation design and uncertainty practices systematically shape the credibility and transferability of reported WQ retrieval performance.

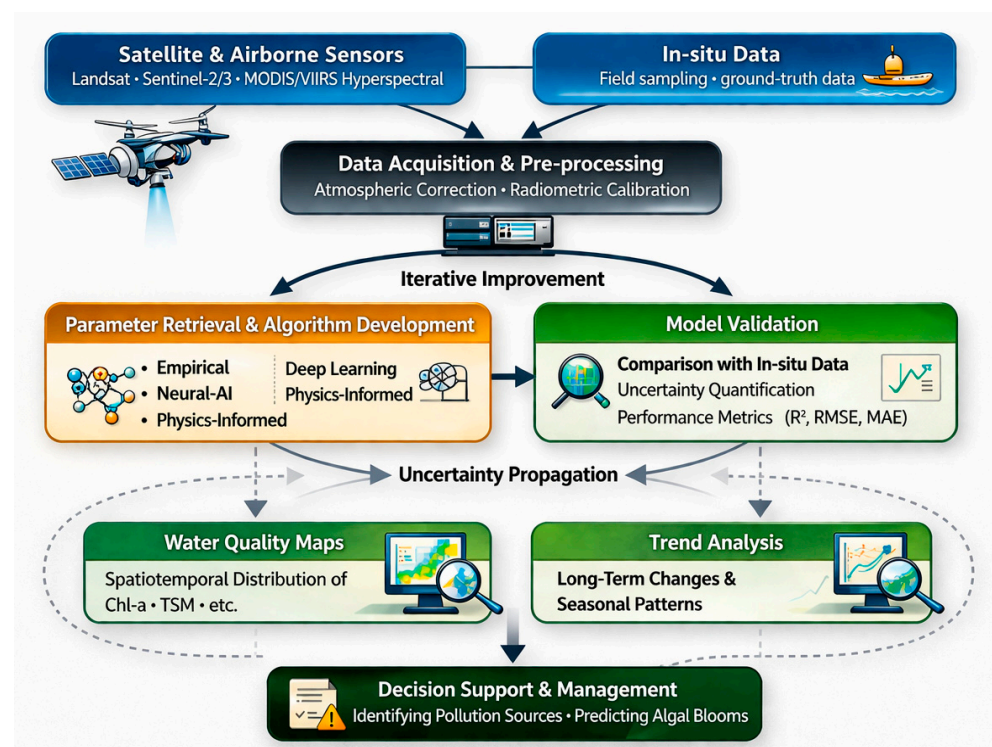


Figure 2. Conceptual framework for uncertainty-aware satellite-based water quality monitoring, integrating multi-source Earth observation and in situ data with physics-informed and data-driven modelling, iterative validation, and downstream mapping, trend analysis, and decision-support applications. Solid arrows indicate the primary data processing and modelling workflow, bidirectional arrows represent iterative refinement between model development and validation, and dashed arrows denote uncertainty propagation and indirect linkages between components. The visual layout of this conceptual figure was developed using a generative artificial intelligence tool to support graphical composition; all scientific content, structure, and interpretation were fully designed, verified, and validated by the authors. The use of AI was limited strictly to visualisation support and did not influence the study design, data analysis, or scientific conclusions.

3.1. Sensor Platforms

The configuration-level synthesis of 152 peer-reviewed studies reveals a highly structured yet non-uniform distribution of sensor usage in satellite-based water quality (WQ) monitoring. Open-access multispectral missions, most prominently the Landsat and Sentinel series, account for the majority of reported retrieval configurations in inland systems [79–81]. Their recurring selection reflects a consistent combination of moderate spatial resolution (10–30 m), radiometric stability, multi-decadal continuity, and unrestricted global accessibility, characteristics that enable reproducible regional-to-continental analyses of lakes and rivers under heterogeneous environmental conditions [82,83]. Across the reviewed evidence base, these attributes are repeatedly associated with stable temporal comparability of reflectance products and with the feasibility of inter-annual trend analyses, particularly when validation protocols explicitly incorporate multi-season or multi-year matchups.

Within this dominant multispectral class, Sentinel-2 MSI is particularly prominent in studies targeting small, morphologically complex, or optically dynamic inland waters. Its 10–20 m spatial resolution, inclusion of red-edge bands sensitive to chlorophyll-*a* absorption–fluorescence dynamics, and shortwave infrared (SWIR) bands (~1.61 and ~2.19 μm) used for turbidity discrimination, atmospheric correction, and land–water masking provide spectral configurations that align closely with optically active water quality parameters [30,84]. Across reviewed model configurations, Sentinel-2–based retrievals are frequently reported in conjunction with machine-learning or hybrid approaches, and several recent deployment-oriented studies indicate that quantisation-aware and hardware-optimised architectures can reduce computational overhead without substantial degradation of predictive skill [68,85]. In parallel, Landsat-9 ensures continuity of the longest global reflectance archive for inland waters, supporting multi-decadal assessments of trophic evolution, sediment dynamics, and temporal generalisation experiments that rely on long-term radiometric consistency [86]. For large lakes and coastal environments, the configuration distribution shifts toward sensors emphasising temporal density over spatial detail. MODIS and VIIRS are frequently used in basin- and global-scale analyses due to their near-daily revisit frequency and long-term observational continuity, particularly in studies of algal bloom dynamics and sediment transport [87,88]. However, their coarse spatial resolution (250–1000 m) limits applicability to narrow, heterogeneous, or small inland systems, reinforcing a persistent spatiotemporal trade-off that constrains reported transferability across water-body types [1,45]. Sentinel-3 OLCI partially mitigates this trade-off by combining daily revisit capability with enhanced spectral resolution and is frequently used for total suspended matter (TSM) retrieval over large inland and coastal waters [38,41]. Emerging missions such as GCOM-C/SGLI and SDGSAT-1 appear in a smaller but growing subset of studies. Multi-year evaluations report effective trend capture for selected parameters, yet also document wavelength-dependent biases and sensitivity constraints associated with atmospheric correction schemes and band configuration [47,48].

Beyond single-sensor usage, multi-sensor harmonisation and data fusion are recurrent methodological strategies within the reviewed corpus. Integrating Landsat, Sentinel, MODIS, and UAV or hyperspectral observations is commonly used to construct time series that balance spatial resolution, temporal density, and continuity [89–91]. However, configuration-level coding indicates that such integration introduces additional sources of uncertainty related to cross-sensor radiometric calibration, atmospheric correction consistency, and preprocessing heterogeneity. Studies incorporating meteorological or land-use covariates report improved explanatory performance for parameters such as turbidity and nutrients in selected contexts. Yet these gains remain region-specific and are not uniformly validated in spatially independent tests [75]. At sub-satellite scales, UAV-mounted

hyperspectral platforms and ground-based micro-hyperspectral spectrometers appear primarily in calibration-intensive or algorithm-development studies. UAV systems consistently demonstrate enhanced constituent discrimination under controlled conditions, attributable to high spectral resolution and flexible acquisition timing. Still, their limited spatial coverage and calibration burden restrict their operational scalability [22,46]. Ground-based micro-hyper spectrometers, by contrast, enable high-frequency autonomous measurements (350–950 nm range), thereby increasing match-up density for calibration, validation, and uncertainty analysis of satellite-derived products [49]. Taken collectively, the evidence does not support the primacy of any single observational platform across all parameter classes, spatial scales, and validation designs. Instead, the reviewed studies demonstrate a structured alignment between sensor characteristics and target application domains: moderate-resolution open-access satellites dominate inland parameter retrieval, high-revisit sensors support large-scale bloom and sediment monitoring, and hyperspectral or in situ systems serve diagnostic and calibration functions. Table 4 synthesises these platform characteristics and their empirically observed roles, while Figure 3 illustrates the geographic clustering of sensor usage and highlights spatial biases within the evidence base. These distributional patterns provide the empirical foundation for subsequent analysis of transferability, validation design, and uncertainty practices across sensor–model configurations.

Table 4. Comparative synthesis of satellite, UAV, and complementary ground platforms for water quality monitoring.

Platform	Key Strengths	Primary Limitations	Best Use Cases	Representative References
Landsat series (e.g., L8, L9)	Long-term archive since 1984; radiometric stability; global free access; includes thermal & SWIR bands for multiple WQPs	30 m limits small rivers/lakes; 16-day revisit; cloud cover	Multi-decadal trend analysis; regional monitoring; historical reconstructions	[37,38,79]
Sentinel-2 MSI	10–20 m detail; red-edge & SWIR bands; global free access; strong for Chl-a, TSM; edge-ready ML deployment (QAT)	5-day revisit may miss rapid change; sensitive to cloud/adjacency; SWIR bands broader than hyperspectral	Small/dynamic inland waters; HAB detection; turbidity/TSM mapping; real-time inference pilots	[39–41,43,64,68]
Sentinel-3 OLCI	Daily revisit; rich spectral coverage; strong for large lakes/coastal TSM	300 m coarse resolution; limited small-water applicability	Large-lake/coastal monitoring; bloom & sediment dynamics	[38,41]
MODIS/VIIRS	Daily to near-daily global coverage; multi-decadal archive; effective for blooms and sediment transport	250–1000 m pixel size; unsuitable for small waters; non-optical WQPs poorly resolved	Global HAB monitoring; climate-trend detection	[43–45,92]
GCOM-C/SGLI	Broad spectral range; demonstrated interannual trend detection	AC-scheme-dependent biases; weaker at UV–blue	Complementary to MODIS/S3 for Chl-a & TSM	[47]
SDGSAT-1 MII	High SNR (4–7× S2 MSI); 10 m spatial fidelity; effective for SPM	Broader bands reduce Chl-a sensitivity; AC limitations in NIR	Fine-scale SPM mapping; urban/coastal regions	[93]
UAV/Airborne Hyperspectral	Very high spatial & spectral resolution; on-demand; strong for constituent discrimination; outperforms in situ heavy sampling	Limited spatial coverage; expensive; calibration-intensive	Pilot studies; algorithm dev.; calibration/validation	[20,22,46,94]
Ground-based Micro-hyperspectrometer	Autonomous, high-frequency, real-time (350–950 nm); high R ² for Chl-a, TP, TN	Stationary; interference; hardware cost	Continuous local monitoring; calibration/validation; ML training	[49]

To complement the sensor-based overview presented in Table 4, the spatial distribution of the reviewed studies is summarised in Figure 3. The figure provides an aggregated representation of the geographic coverage of the 152 studies included in this review, based on the study regions reported in the original publications. Study locations are shown at con-

tinental and sub-continental scales to ensure consistency across studies with heterogeneous spatial reporting. Water quality parameters investigated at each location are indicated to reflect the breadth of parameter coverage across regions, providing a descriptive overview of the spatial footprint of the reviewed literature before the presentation of algorithmic, validation, and uncertainty-related results.

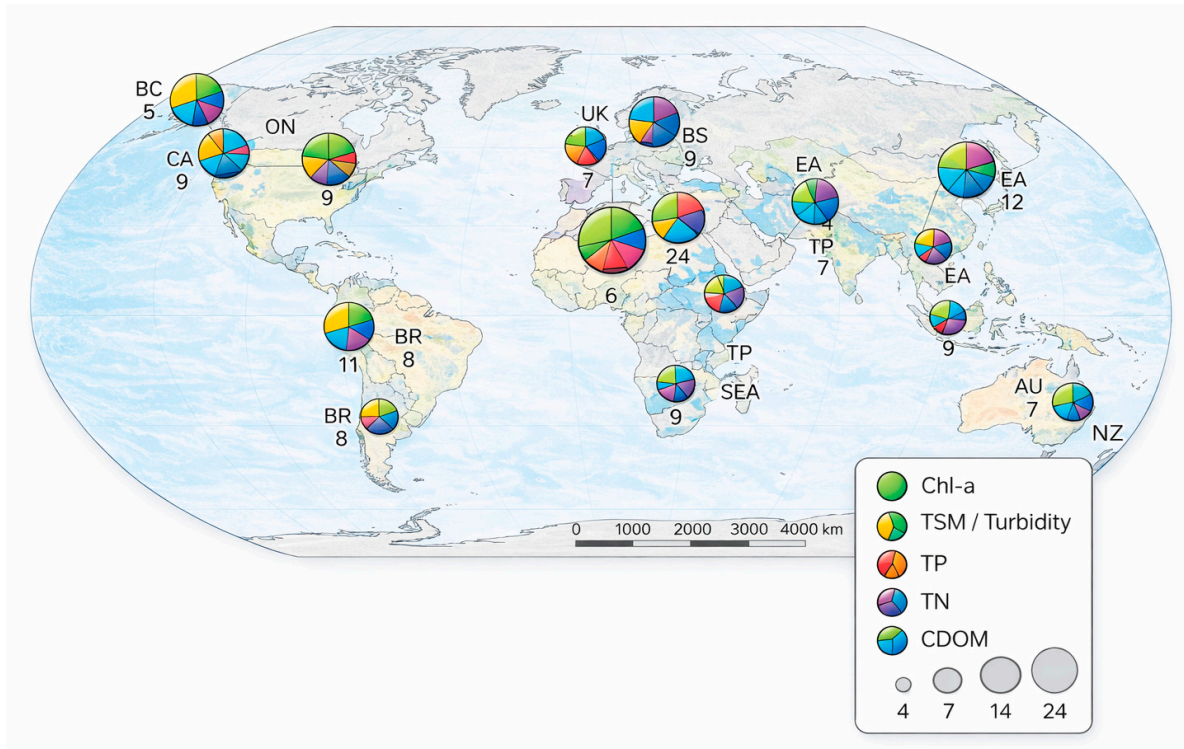


Figure 3. Geographic distribution and regional bias of satellite-based water quality retrieval studies. The map summarises the spatial distribution of the 152 reviewed studies, aggregated at continental and sub-continental scales based on reported study locations. Symbol colours denote the primary water quality parameters investigated (e.g., chlorophyll-a, turbidity/TSM, nutrients, CDOM), revealing pronounced geographic clustering and the under-representation of several regions. Locations are presented at an aggregated regional level, as many studies do not report precise geographic coordinates. The graphical composition of this figure was supported by a generative artificial intelligence tool to facilitate cartographic visualisation; however, all underlying data compilation, spatial aggregation, categorisation, and scientific interpretation were conducted and rigorously verified by the authors. The use of AI was limited to visual rendering and did not influence the study’s data, analysis, or conclusions.

3.2. Target Parameters

The configuration-level synthesis of 152 peer-reviewed studies demonstrates a clear structural concentration of satellite-based water quality (WQ) retrieval on optically active constituents characterised by diagnostically strong spectral signatures. As summarized in Table 5, chlorophyll-a (Chl-a), total suspended matter (TSM) and turbidity, and colored dissolved organic matter (CDOM) account for the majority of reported retrieval configurations, whereas non-optical or weakly optical parameters including total phosphorus (TP), total nitrogen (TN), ammoniacal nitrogen ($\text{NH}_3\text{-N}$), dissolved organic carbon (DOC), and related chemical indicators remain comparatively underrepresented [95,96]. This distribution reflects a fundamental radiative constraint: parameters exhibiting distinct absorption or scattering features within multispectral band configurations are inherently more retrievable than those requiring indirect statistical inference. As further detailed in Table 5, this structural asymmetry is consistently reflected in the alignment between parameter type, sensor

capability, and dominant modelling strategies across the reviewed studies. Chlorophyll-*a* remains the most frequently targeted indicator, consistent with its ecological role as a proxy for algal biomass and trophic state [97]. Retrieval approaches consistently exploit its characteristic absorption features in the blue (~440 nm) and red (~665 nm) regions, as well as its reflectance and fluorescence peaks in the red-edge domain (~705–710 nm) [30,31]. The introduction of red-edge bands in Sentinel-2 MSI has expanded retrieval capability in optically complex inland waters, particularly in conditions where blue–green algorithms are confounded by CDOM or suspended sediments [17,84]. Recent machine learning (ML) and deep learning (DL) implementations, including Random Forest and CNN-based architectures, report high predictive performance under internally validated or site-specific conditions, with several studies documenting R^2 values exceeding 0.9 [64,67]. However, stratified evidence across the reviewed configurations indicates that such performance frequently declines under spatially or temporally independent validation, highlighting the sensitivity of reported accuracy to validation design and domain representativeness.

TSM and turbidity form the second dominant parameter class, reflecting their importance for water clarity, sediment transport, and ecosystem light regimes [98]. Retrieval exploits strong particle backscattering in the red and near-infrared (NIR) spectral regions [99,100]. Empirical band ratios remain prevalent in locally calibrated contexts, whereas physics-based radiative transfer (RT) models and ML approaches extend their applicability to broader concentration ranges. Multi-year Sentinel-2 analyses incorporating meteorological drivers have improved explanatory characterisation of turbidity regimes in reservoir systems [75], and quantisation-aware CNN architectures have demonstrated edge-deployable TSS retrieval under constrained computational settings [68]. Nonetheless, reflectance saturation at high sediment loads, variability in particle composition, adjacency effects, and sensitivity to atmospheric correction remain recurrent constraints on transferability. CDOM retrieval remains structurally more challenging despite its strong UV–blue absorption (~350–450 nm) [101,102]. In inland waters, spectral covariance with Chl-*a* and TSM complicates signal separation [103,104]. Physics-based inversion methods and multi-sensor fusion have improved robustness in selected systems [32,105], and UAV-based hyperspectral campaigns enhance diagnostic discrimination due to finer spectral resolution [46]. However, uncertainty in atmospheric correction and low signal-to-noise ratios in the blue region continue to constrain cross-site generalisation. Nutrient retrieval (TP, TN, $\text{NH}_3\text{-N}$) and other optically inactive constituents represent the current methodological frontier and highlight the structural limitations of purely radiative approaches. As reflected in Table 5, these parameters are not directly observable but are inferred through indirect relationships with optical proxies (e.g., Chl-*a*, CDOM, TSM) and auxiliary environmental drivers. Three dominant inference pathways can be identified across the reviewed studies. First, the biological proxy pathway links nutrients to phytoplankton dynamics via Chl-*a*, typically exhibiting stronger correlations in eutrophic or bloom-dominated systems but weakening under decoupled conditions. Second, the organic matter pathway links nutrients to CDOM or DOC signals, although these relationships are highly context-dependent and often confounded by terrestrial inputs. Third, the hydrological sediment pathway connects nutrient dynamics to suspended matter and runoff-driven processes, resulting in temporally variable and event-sensitive relationships. These pathways indicate that correlations between non-optical parameters and optical proxies are neither uniform nor stable, but instead depend on optical regime, seasonality, and watershed forcing. Consequently, reported retrieval performance reflects not only model capability but also the stability of these proxy relationships. Recent ML-driven studies report moderate-to-high in-domain performance (e.g., $R^2 \approx 0.78$ for TP using Random Forest), yet these results are rarely transferable across regions or seasons. As summarised in Table 5, successful

retrieval of non-optical parameters typically requires integrating auxiliary data sources, including meteorological variables, land-use characteristics, and in situ observations, which improve model performance by constraining contextual variability rather than directly enhancing observability. Overall, the synthesis in Table 5 highlights a fundamental structural distinction: while optically active parameters benefit from strong radiative signals that support transferable retrieval, non-optical parameters rely on indirect, context-dependent inference pathways that limit generalizability. This asymmetry forms a key basis for the transferability and uncertainty patterns examined in subsequent sections.

Table 5. Comparative synthesis of key water quality parameters, their dominant optical properties, commonly used sensors, retrieval approaches, and representative references.

Parameter	Primary Optical Property	Key Sensors Used	Common Methodological Approaches	Representative References
Chlorophyll-a (Chl-a)	Absorption in blue (~440 nm), reflectance/fluorescence in red-edge (~670–710 nm)	Sentinel-2, Landsat 8/9, Sentinel-3, MODIS, Hyperspectral	Band ratios; ML (RF, SVR); DL (CNNs); global CNN frameworks	[38,60,64,67]
TSM & Turbidity	Strong backscattering in red/NIR	Sentinel-2, Landsat 8/9, Sentinel-3, MODIS, UAV hyperspectral	Empirical ratios; physics-based RT models; ML (RF, XGB); edge-ready DL	[43,45,68,75]
CDOM	Strong absorption in UV-blue (~350–450 nm)	Sentinel-2, Landsat 8/9, UAV/Hyperspectral	Physics-based inversion; ML/DL; multi-sensor fusion	[11,46,58,91,106]
Nutrients (TP, TN, NH ₃ -N, DOC)	Optically inactive; inferred via correlations with surrogates (Chl-a, CDOM) + contextual drivers	Sentinel-2, Landsat 8/9, Hyperspectral + ancillary (land use, meteorology, IoT)	ML/DL (RF, Extra Trees, XGBoost, CNN); SHAP feature attribution; indirect statistical models	[35,42,62,78,96]

While Table 5 organises retrieval patterns by parameter class and methodological family, a complementary spectral physical synthesis is required to explain why certain parameters exhibit higher stability and transferability across sensor platforms. Retrieval success is fundamentally governed by radiative transfer interactions between water constituents and incident radiation, with robustness depending on the strength, separability, and signal-to-noise characteristics of diagnostic spectral features within the available band configurations. As explicitly synthesised in Table 6, these radiative–physical properties determine not only detectability but also the extent to which retrieval relationships can be generalised across sensors and environmental conditions. Parameters with strong and well-separated spectral features, such as chlorophyll-a, benefit from physically interpretable absorption and reflectance signals (e.g., blue and red absorption, red-edge reflectance peaks), enabling relatively robust retrieval across platforms, albeit with known limitations such as adjacency effects and constituent covariance [38,84]. Similarly, TSM retrieval leverages monotonic red/NIR backscattering responses, but, as summarised in Table 6, reflectance saturation, particle-composition variability, and atmospheric-correction sensitivity constrain extrapolation beyond calibrated ranges [98,99]. In contrast, parameters with weaker or overlapping spectral signatures exhibit substantially greater sensitivity to confounding effects. CDOM, despite its strong UV–blue absorption, is highly susceptible to spectral overlap with Chl-a and TSM, as well as to uncertainties in atmospheric correction, particularly in inland waters with low signal-to-noise ratios. These limitations, detailed in Table 6, reduce the separability of spectral signals and constrain cross-site generalisation [101,103,104].

The limitations become most pronounced for optically inactive parameters. As indicated in Table 6, nutrient-related variables (TP, TN) and other non-optical indicators lack intrinsic spectral responses and are therefore inferred through indirect, context-dependent relationships with optical proxies and auxiliary drivers. This weak radiative linkage results in high sensitivity to domain shift, calibration bias, and variability in proxy stability. Simi-

larly, dissolved oxygen and Secchi depth are indirectly inferred from surrogate indicators of water clarity or biological activity, which limits their robustness and transferability [67]. Collectively, the synthesis across Tables 5 and 6 demonstrates that parameter retrievability in satellite-based WQ monitoring is constrained not only by algorithmic sophistication but fundamentally by radiative–physical structure, spectral separability, and the stability of proxy relationships. These structural asymmetries explain why optically active parameters exhibit higher transferability, while non-optical parameters remain strongly context dependent. This distinction provides the empirical foundation for the transferability limitations and uncertainty propagation patterns analysed in subsequent sections.

Table 6. Spectral sensitivity and band summary for major water quality parameters retrieved from satellite observations. The table synthesises dominant spectral features, key wavelength ranges, and sensor support across the reviewed literature, while highlighting common failure modes that constrain transferability and robustness, particularly for optically inactive parameters.

Parameter	Dominant Spectral Features	Key Bands/Wavelength Ranges (nm)	Best-Supported Sensors	Typical Failure Modes & Limitations	Representative References
Chlorophyll-a (Chl-a)	Strong absorption in blue and red; reflectance/fluorescence peak in red-edge	Absorption: ~440, ~665; Red-edge peak: ~705–710; NIR shoulder: ~740	Sentinel-2 MSI (red-edge bands); Sentinel-3 OLCI; Landsat-8/9 (limited red-edge via proxies); MODIS/VIIRS (large waters); UAV hyperspectral	Adjacency effects in small waters; AC sensitivity in blue; saturation at high biomass; CDOM/TSM confounding	[17,30,38,84]
Total Suspended Matter (TSM)/Turbidity	Strong backscattering in red and NIR; monotonic reflectance increase with concentration	Red: ~620–670; NIR: ~700–900; SWIR (~1610–2190) for masking & high loads	Sentinel-2 MSI; Landsat-8/9 OLI; Sentinel-3 OLCI (large lakes/coastal); MODIS/VIIRS; UAV hyperspectral	Reflectance saturation at high loads; particle composition variability; AC errors; adjacency effects	[45,90,98,99]
CDOM	Strong absorption in UV-blue; weak direct signal beyond ~500 nm	UV-blue: ~350–450; blue–green ratios ~412–490	Sentinel-2 MSI; Landsat-8/9 (limited blue); UAV hyperspectral; Sentinel-3 OLCI (large waters)	Severe confounding with Chl-a and TSM; high AC sensitivity in blue; low SNR in inland waters	[32,101,103,104]
Nutrients (TP, TN)	Optically inactive; inferred via correlations with optical surrogates and contextual drivers	Indirect: Chl-a (~705), CDOM (~412–443), TSM (red/NIR) + ancillary variables	Sentinel-2 MSI; Landsat-8/9; UAV hyperspectral (local); fusion with in situ/ancillary data	Weak physical linkage; site/season dependence; limited generalizability; strong model bias risk	[13,62,78,96]
Dissolved Oxygen (DO)/Secchi depth	No direct spectral signature; inferred via clarity or biological proxies	Indirect: blue–green ratios; NIR backscattering (Secchi)	Landsat-8/9; Sentinel-2 MSI; MODIS (large waters)	Very low robustness; indirect inference only; strong dependence on calibration context	[12,13,67]

To further systematise the inversion mechanisms of optically inactive water quality parameters, Table 7 provides an evidence-informed synthesis of dominant proxy relationships, underlying inference pathways, and the role of auxiliary data in improving retrieval performance. While Table 6 highlights the spectral limitations and weak radiative linkage of these parameters, the synthesis in Table 7 focuses on how indirect, context-dependent relationships are operationalised within retrieval models. Emphasis is placed on the variability of proxy strength, the contribution of auxiliary data (e.g., meteorological and land-use information), and the resulting implications for model selection and transferability.

3.3. Algorithms & Models

The effectiveness of remote sensing–based retrieval of water quality parameters is fundamentally governed by algorithmic design, which shapes not only predictive accuracy but also transferability, interpretability, and operational feasibility. Across the reviewed

literature, a consistent methodological evolution is evident: early empirical regressions and semi-analytical formulations were progressively complemented first by physics-based bio-optical approaches and, more recently, by machine learning (ML) and deep learning (DL) frameworks. This trajectory is widely recognised as reflecting persistent trade-offs among simplicity, physical consistency, generalisation across optical regimes, and operational scalability [41,72]. Empirical and semi-analytical models dominated early remote sensing applications for water quality monitoring. These methods typically employ band ratios or regression relationships between spectral reflectance and in situ measurements and remain attractive for their computational efficiency and simplicity of implementation. However, systematic evidence indicates that their reliance on site-specific calibration often limits transferability, especially in optically complex inland and coastal waters, where constituent interactions vary substantially [107–109]. Physics-based models, most notably bio-optical radiative transfer formulations and semi-analytical inversion schemes, offer a mechanistic foundation by explicitly representing light–constituent interactions. When underlying optical assumptions and inherent optical properties are appropriately characterised, these approaches provide stronger interpretability and physical consistency than purely empirical regressions. Nonetheless, their reliance on ancillary inputs, sensitivity to atmospheric correction and parameterisation, and comparatively high computational complexity can restrict routine operational deployment at scale [110,111]. The rise in ML paralleled advances in computational capacity and the increasing availability of multi-sensor datasets and cloud-based processing environments. Algorithms such as Random Forest (RF), Support Vector Regression (SVR), and boosting methods have repeatedly demonstrated a strong capacity to capture nonlinear relationships between reflectance and water quality parameters [59,61]. Notably, tree-based ensembles have shown practical utility for optically weak or non-optical parameters (e.g., nutrients) when contextual covariates or multi-source inputs are available [62,92]. However, a recurring limitation is sensitivity to training-data representativeness: strong cross-validation performance may not translate to out-of-domain deployment across regions, seasons, or sensors, highlighting persistent generalisation risks [96,112].

DL represents the most recent methodological frontier in this domain. CNN-based models and spatiotemporal hybrid architectures (e.g., CNN–LSTM) enable automated feature extraction and can outperform conventional baselines across heterogeneous aquatic environments, including applications leveraging long-term Landsat archives and Sentinel platforms [113]. Nevertheless, DL models often require large and representative labelled datasets, can be computationally demanding, and frequently remain difficult to interpret, which constitutes a critical limitation in decision-support contexts [34,114]. Recent work on deployment-oriented optimisation (e.g., quantisation and edge-oriented inference) indicates that operational constraints are increasingly being addressed alongside accuracy considerations [68]. Beyond the conventional labelling of models as “empirical,” “physics-based,” or “learning-based,” the reviewed literature demonstrates that physical knowledge can be incorporated into learning workflows at multiple, qualitatively distinct levels. These include physics-guided feature design or optical water type guidance, simulation- or physics-motivated priors, and more tightly coupled hybridisation strategies that seek to improve physical consistency and robustness beyond calibration domains [96,114]. To consistently distinguish the degree to which physical knowledge is operationalised within model configurations, this review adopts the physics-integration coding scheme summarised in Table 8 (P0–P4). This coding is applied at the model-configuration level to preserve analytical granularity when multiple approaches are reported within a single study.

Table 7. Evidence-informed synthesis of inversion pathways for optically inactive water quality parameters. The table summarises dominant optical proxies, underlying inference pathways, and the role of auxiliary data in supporting the retrieval of non-optical constituents (e.g., nutrients and DOC). Rather than representing direct spectral observability, these relationships reflect context-dependent associations mediated by biological, organic, and hydrological processes. The synthesis highlights how auxiliary data (e.g., meteorological variables, land use, and in situ observations) improve retrieval performance by constraining environmental variability, while also emphasising the inherent limitations of transferability due to proxy instability and domain dependence.

Target Parameter	Dominant Optical Proxies	Inference Pathway	Typical Auxiliary Data	Reported Role/Improvement Mechanism	Main Limitations to Generalisation	Suitable Modelling Strategies
TP (Total Phosphorus)	Chl-a (red-edge), CDOM (blue), TSM (red/NIR)	Biological + hydrological coupling	Land use, precipitation, runoff, temperature	Improves contextual interpretation of nutrient enrichment and runoff-driven variability	Strong seasonal dependence; proxy instability; watershed heterogeneity	ML/DL (RF, XGBoost, CNN); hybrid models with contextual inputs
TN (Total Nitrogen)	Chl-a (red-edge), CDOM (blue)	Biological + organic matter linkage	Meteorological variables, watershed characteristics, and rainfall events	Enhances detection under bloom or rainfall-driven conditions	Weak direct linkage; strong domain shift sensitivity; event-dependence	ML/DL with stratification (seasonal/regime-aware models)
NH ₃ -N	Chl-a (indirect), CDOM	Weak biological coupling	Temperature, pH, and seasonal indicators	Supports indirect inference via ecosystem state	Highly unstable correlations; strong temporal variability; limited transferability	ML (Extra Trees, XGBoost) with feature attribution (e.g., SHAP)
DOC	CDOM (UV-blue), Chl-a (secondary)	Organic matter co-variation	Land use, hydrology, watershed inputs	Improves estimation where terrestrial DOM dominates	Confounding from mixed sources; spectral overlap; AC sensitivity	ML/DL + multi-source data fusion
Nutrients (General)	Chl-a, CDOM, TSM	Combined biological–organic–hydrological pathways	Meteorology, land use, IoT/in situ data	Reduces ambiguity in proxy relationships; improves model robustness under regime variability	Context-dependent relationships; poor cross-region generalisation; high model bias risk	ML/DL + hybrid frameworks; OWT-conditioned or regime-aware models

Table 8. Physics-integration coding scheme for classifying physics-informed learning in satellite-based water quality studies. Codes P0–P4 denote increasing levels of physical knowledge integration, from purely data-driven models to hybrid architectures that structurally couple physical operators with learning components. Coding is evidence-based, applied at the model-configuration level, and relies exclusively on information explicitly reported in each study.

Code	Physics Integration Category	Operational Definition (Coding Rule)	Minimum Evidence Required in the Paper	Representative Foundational References
P0	Purely data-driven (no explicit physics)	No explicit physical knowledge (e.g., radiative transfer, conservation constraints, physically consistent penalties) is used in inputs, training objective, model structure, or evaluation.	Methods describe only statistical/ML/DL mapping from inputs to WQPs; no RT/physics constraints beyond narrative discussion.	Physics-informed NN overview context (baseline contrast): [115]
P1	Physics-guided features/inputs	Physical knowledge is used to derive/select input features or predictors but not embedded in the loss function or architecture (no constraint enforcement).	Use of physics-motivated band ratios/indices, IOP-inspired features, or physically motivated preprocessing as model inputs; the objective remains purely data-fit.	Physics-guided neural networks leveraging physics-based model simulations [116] (https://arxiv.org/abs/1710.11431 , accessed on 1 April 2026).
P2	Physics-based simulation for training/augmentation	A physical/RT or process-based model is used to generate synthetic samples for training, pretraining, or augmentation; the learner itself is not physically constrained.	Clear statement that RT/process simulations were used to create spectra/training labels or augment training distribution.	Physics-guided neural networks leveraging physics-based model simulations [116] (https://arxiv.org/abs/1710.11431 , accessed on 1 April 2026).
P3	Physics-informed objective/regularization	Physical relationships are explicitly encoded as loss terms/constraints/regularizers during training (soft or hard constraints), improving physical consistency and extrapolation behavior.	Explicit formulation or description of a physics term in the objective (e.g., penalty for physically implausible spectra–WQP relations; conservation constraints).	PINNs’ canonical formulation: [115]
P4	Hybrid coupled physics-ML architecture (incl. differentiable physics layer)	Physics and ML are structurally coupled (e.g., differentiable physics/RT layer inside the network; ML emulates or corrects a physics component) within a single end-to-end or tightly coupled pipeline.	Clear description of a coupled architecture (physics operator or RT module integrated with ML, or end-to-end differentiable physical component).	Differentiable RT paradigm example (remote sensing): [117]; PINNs’ foundations: [115]

While Table 8 clarifies how physical knowledge is explicitly embedded in individual model configurations, ranging from purely data-driven formulations to hybrid architec-

tures with structural physical coupling, a complementary synthesis is required to situate these configurations within the broader methodological families that dominate the water quality remote sensing literature. Accordingly, Table 9 maps the major algorithmic families (empirical and semi-analytical models, physics-based approaches, machine learning, and deep learning) onto the physics-integration spectrum defined in Table 8, while simultaneously summarising their characteristic strengths and limitations as consistently reported across the reviewed studies and established syntheses. This mapping clarifies that although recent advances increasingly explore higher levels of physics integration, a substantial fraction of contemporary ML and DL applications remain concentrated at lower or intermediate levels of integration, relying primarily on data-driven representations. By explicitly distinguishing between model families and physics-integration levels, Table 9 avoids redundancy with Table 8 and instead provides a structured lens for interpreting the methodological evolution of the field. This synthesis illustrates the progressive shift from locally calibrated, empirically driven approaches toward more physically consistent and deployment-oriented frameworks, and it establishes a clear conceptual bridge to the transferable, physics-informed, and uncertainty-aware framework proposed in Section 5.

Table 9. Mapping of major algorithmic families used in satellite-based water quality retrieval to the physics-integration coding framework (P0–P4) defined in Table 8. Rather than reiterating methodological details, this table provides a structured synthesis that positions each model family along the physics-integration spectrum, illustrating how physical knowledge is incorporated into modelling strategies in practice. (* “Dominant physics-integration level” refers to the representative or most commonly observed level (P0–P4) within each model family based on the evidence synthesis, and does not imply that all studies strictly conform to a single level).

Model Family	Typical Modelling Strategy	Dominant Physics-Integration Level *	Key Strengths	Primary Limitations	Representative References
Empirical & Semi-analytical	Band ratios; linear/nonlinear regressions calibrated to in situ data	P0–P1	Simple, computationally efficient, transparent formulation, effective for local applications	Strong site-specificity; poor transferability; requires frequent recalibration	[12,37,38]
Physics-based (Bio-optical/RT)	Radiative transfer inversion; semi-analytical bio-optical models	P2	Physically interpretable; transferable if IOPs are well constrained; mechanistic consistency	Computationally intensive; extensive input requirements; difficult to operationalise at scale	[57,76,110]
Machine Learning (ML)	RF, SVR, XGBoost, CatBoost using spectral + contextual predictors	P1–P2	Captures nonlinear relationships; integrates multi-sensor and ancillary data; effective for non-optical WQPs	Requires large, representative datasets; extrapolation risk; interpretability is limited	[40,62,78]
Deep Learning (DL)	CNNs, CNN-LSTM, global DL frameworks	P1–P3	State-of-the-art predictive performance; automated feature extraction; emerging cross-site generalisation	Very high data demand; computational cost; black-box behaviour	[64,65,67]
Hybrid/Physics-informed ML–DL	RT-guided features; physics-based regularisation; differentiable physical components	P3–P4	Balances accuracy with physical consistency; improved interpretability and transferability; suitable for decision-grade outputs	Higher implementation complexity; limited number of fully operational examples	[38,58,74]

Collectively, this synthesis indicates that the field is transitioning from models optimised primarily for local calibration and methodological simplicity toward frameworks that explicitly balance predictive performance with physical consistency, transferability, and operational readiness. While early empirical and semi-analytical approaches remain valuable for site-specific applications, the growing reliance on machine learning and deep learning reflects an effort to capture complex, nonlinear relationships and to scale anal-

yses across heterogeneous aquatic environments. However, when viewed through the physics-integration framework defined in Table 8 and the corresponding evidence map summarising sensor method parameter coverage (Figure 4), it becomes evident that many conventional ML and DL implementations remain concentrated at lower or intermediate integration levels, relying predominantly on data-driven representations without explicit physical constraints. The evidence map further reveals that higher levels of physics integration are unevenly distributed across sensors and parameters, with a strong methodological concentration on optically active variables and comparatively limited coverage of non-optical constituents. In contrast, recent developments increasingly explore higher levels of integration and deployment-oriented design choices, including physics-guided learning, hybrid radiative-transfer-informed architectures, and computational strategies for near-real-time operation. As highlighted by recent syntheses, rigorous uncertainty handling and validation design remain essential prerequisites for translating these methodological advances into decision-grade monitoring systems; these issues are addressed in detail in Section 3.4.

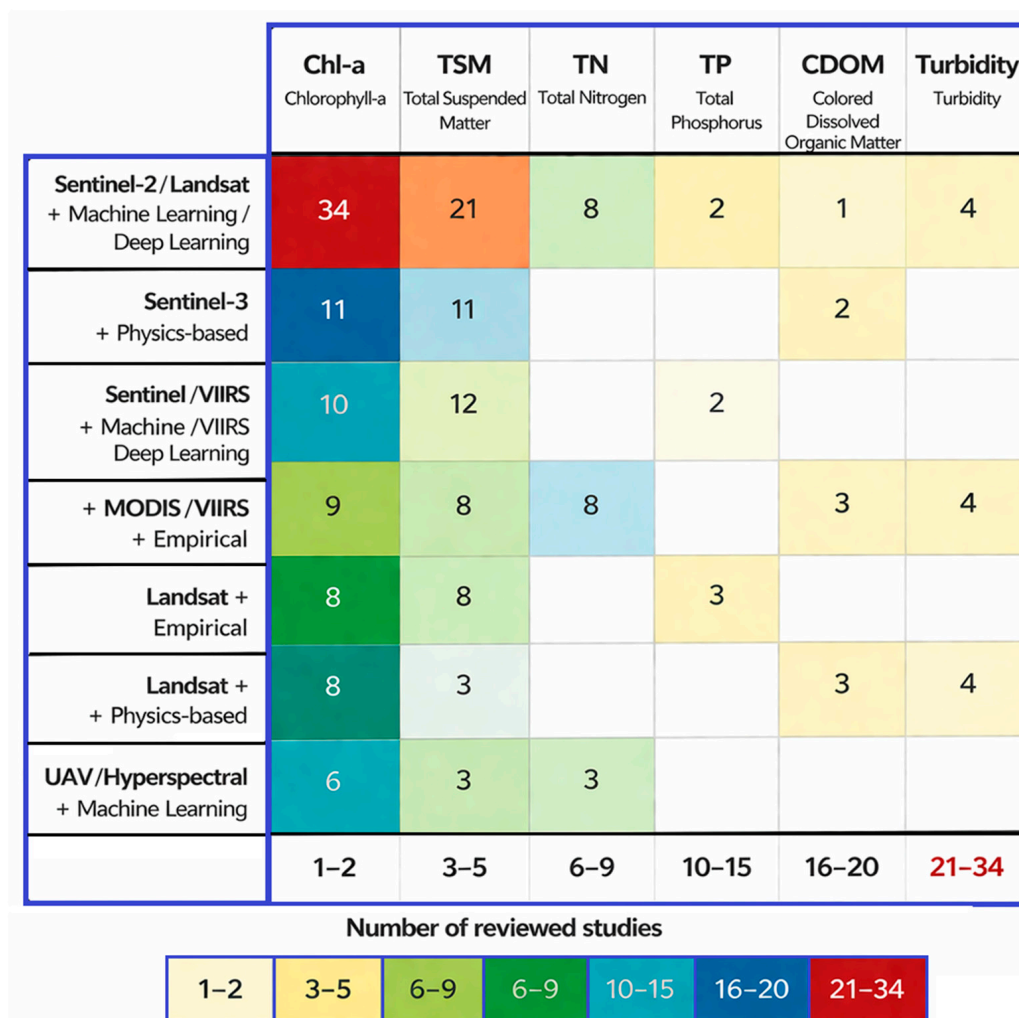


Figure 4. Evidence map summarising the number of reviewed satellite-based water quality retrieval studies across sensor classes, retrieval methods, and target parameters. Cell values and colour intensity indicate the count of studies reporting each sensor method parameter combination. Blank cells denote the absence of reported evidence. The map highlights uneven coverage across literature, with strong emphasis on multispectral sensors and machine learning approaches for optically active parameters, and limited representation of UAV/hyperspectral studies and optically inactive parameters.

Figure 5 illustrates the aggregate temporal evolution of modelling approaches applied in remote sensing-based water quality retrieval, reinforcing the methodological trajectory outlined above. The temporal distribution of studies reveals a clear progression from empirically driven and locally calibrated models toward increasingly sophisticated machine learning, deep learning, and hybrid formulations, thereby providing quantitative evidence synthesis for a field-wide shift toward more physically consistent and operationally oriented frameworks.

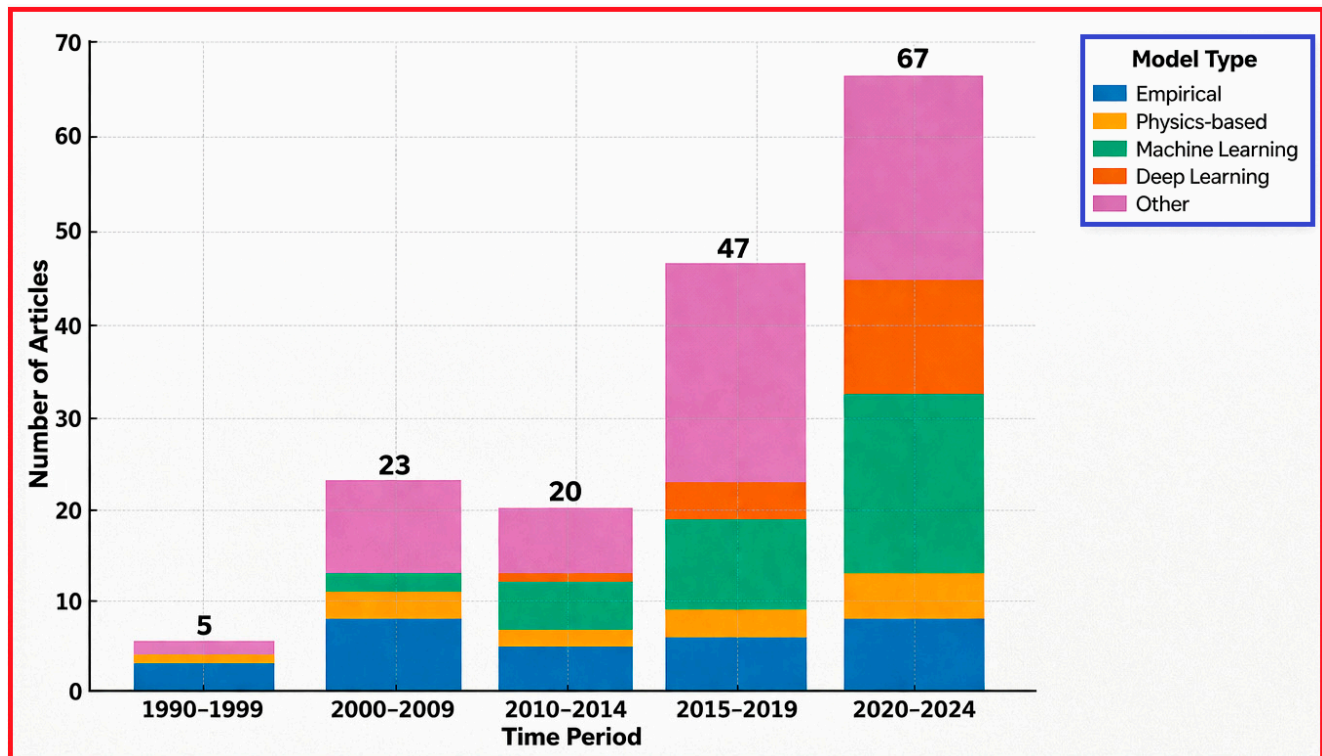


Figure 5. Aggregate temporal evolution of modelling approaches used in remote sensing-based water quality retrieval across five time periods. The distribution of studies highlights a progressive shift from empirically driven, physics-based models toward machine learning, deep learning, and hybrid approaches, supporting the synthesis presented in Tables 8 and 9 and illustrating the field’s maturation toward transferable, deployment-oriented frameworks.

While Figure 5 quantitatively illustrates the temporal evolution of modelling approaches in satellite-based water quality retrieval from locally calibrated empirical methods toward increasingly sophisticated machine learning, deep learning, and hybrid formulations, it does not by itself resolve how these heterogeneous approaches are structurally organised within a unified modelling pipeline. In particular, the temporal aggregation highlights what classes of models have emerged and proliferated, but remains agnostic to how physical knowledge, data-driven learning, validation design, and uncertainty handling are operationally combined within individual model configurations. To bridge this gap, Figure 6 provides a formalised conceptual synthesis that translates the evidence summarised in Tables 8 and 9, and Figure 5 into an explicit end-to-end framework. By positioning empirical, physics-based, and learning-based components within a single architectural flow linking multisensor inputs, feature extraction, hybrid modelling strategies, loss formulation, validation, calibration, and uncertainty quantification, the framework clarifies how the methodological trajectory observed over time can be instantiated in practice as a transferable, physics-informed, and uncertainty-aware monitoring system.

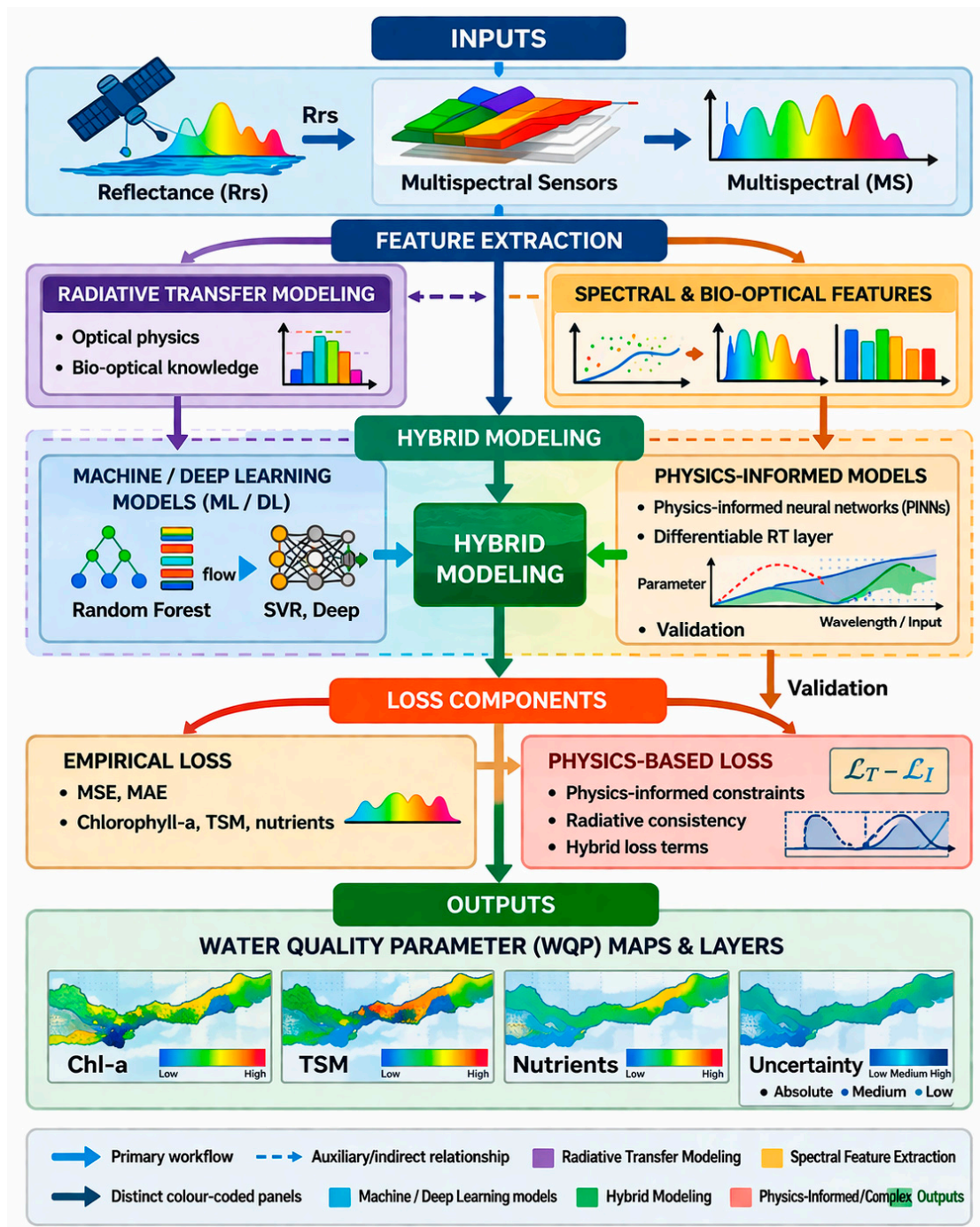


Figure 6. Conceptual framework for hybrid and physics-informed satellite-based water quality retrieval, integrating multisensory Earth observation inputs, feature extraction, and coupled modelling strategies with validation workflows and uncertainty-aware outputs. The framework illustrates how physics-informed and data-driven approaches can be combined to generate water quality products with associated uncertainty layers, enabling robust mapping, trend analysis, and decision-relevant applications. Color-coded panels represent the main functional modules of the framework, solid arrows indicate the primary workflow, and dashed lines denote auxiliary or indirect relationships between components. The visual design of this conceptual framework was developed using a generative artificial intelligence tool to support graphical structuring and layout. All methodological components, scientific relationships, and interpretative elements were explicitly defined, critically evaluated, and validated by the authors. The use of AI was strictly limited to visualisation support and had no role in data processing, model development, analysis, or the derivation of scientific conclusions.

While the preceding sections have detailed the evolution of algorithmic strategies and levels of physical integration in satellite-based water quality retrieval, a synthesis of these developments alone does not fully capture the field's current level of practical readiness or the persistence of critical bottlenecks. Notably, the reviewed evidence consistently indicates that no single model family, whether empirical, physics-based, or data-driven, achieves robust performance across all water quality parameters, optical regimes, sensor configurations, and deployment contexts. Model performance is inherently conditioned by domain-specific factors, including optical complexity, data availability, and environmental variability, leading to context-dependent strengths and failure modes across different modelling approaches. In this context, ensemble and multi-model fusion strategies emerge as a promising pathway toward improved robustness and generalisation. By integrating the complementary strengths of empirical formulations, physics-informed models, and machine-learning or deep-learning approaches, fusion frameworks can mitigate model-specific biases and enhance adaptability across heterogeneous conditions. Such integration may occur at multiple levels, including feature-level fusion, model-level ensemble learning (e.g., stacking or mixture-of-experts), and decision-level combination of outputs. Although relatively underexplored and inconsistently implemented across the reviewed literature, these approaches provide a conceptual bridge between methodological diversity and operational robustness, particularly in environments characterised by strong optical variability and domain shifts. To bridge the gap between methodological advancement and operational readiness, a structured roadmap is required to explicitly align achieved capabilities with persistent system-level constraints. Accordingly, Table 10 consolidates the available evidence into a milestone gap framework, summarising how advances in algorithms, physics integration, and computational strategies have progressed, while simultaneously highlighting unresolved challenges related to limited transferability, weak interpretability, insufficient uncertainty quantification, and fragmented multi-sensor integration and operational infrastructure. By framing recent progress and remaining constraints within a unified structure while also recognising the emerging role of multi-model fusion as a robustness-enhancing strategy, this roadmap provides a coherent context for the subsequent emphasis on validation and uncertainty considerations, which remain central to the transition from research-oriented demonstrations to decision-grade and operationally reliable water quality monitoring systems.

3.4. Validation and Uncertainty

Validation represents the critical interface between satellite-derived water quality retrievals and ground-based observations, providing the methodological rigour required to bridge research-oriented demonstrations and decision-relevant operational monitoring systems [33,118]. Across the reviewed literature, comparisons with near-synchronous in situ measurements remain the dominant approach for performance evaluation, implicitly defining the degree to which retrieval models generalise across ecosystems, seasons, and temporal domains [112,119]. However, validation outcomes are commonly communicated through scatter-based diagnostics (e.g., agreement relative to the 1:1 line) and summary accuracy metrics, which, when reported in isolation, obscure substantial variability arising from sensor characteristics, target parameters, modelling paradigms, and, critically, validation design. To move beyond narrative reporting and enable systematic cross-study comparability, this review adopts a distribution-aware performance synthesis across the 152 retained studies. Given the pronounced heterogeneity in reported metrics, sample sizes, and evaluation protocols, validation outcomes are summarised using robust statistics (median and interquartile range), rather than mean-based aggregation, thereby reducing sensitivity to extreme values and study-specific reporting practices. Performance distribu-

tions are explicitly stratified by water quality parameter, sensor class, model family, and validation protocol (random cross-validation versus spatially or temporally independent evaluation), enabling identification of structural patterns such as performance inflation under cross-validation, parameter-dependent limits to generalisation, and sensor-specific variability in reported accuracy. These challenges are most pronounced for non-optically active parameters, particularly total phosphorus (TP) and total nitrogen (TN), where retrieval accuracy depends strongly on optical water regime, concentration gradients, and the availability of well-distributed and temporally representative calibration datasets [120]. Persistent scarcity of spatially diverse and temporally consistent in situ observations continues to constrain calibration, extrapolation, and cross-site transferability [121,122]. As a result, even studies reporting high predictive accuracy frequently rely on site-specific calibration strategies, limiting scalability and long-term operational relevance, particularly for empirically driven approaches that require repeated recalibration across locations or seasons [123,124].

Table 10. Roadmap summarizing key milestones, persistent gaps, and priority actions for advancing satellite-based water quality monitoring toward transferable, physics-informed, and uncertainty-aware operational systems.

Domain	Current State (Milestones Achieved)	Persistent Gaps/Limitations	Priority Actions (Next Milestones)	Implications for Operational Readiness
Sensor Platforms & Data	Long-term open-access archives (Landsat, Sentinel); daily global coverage (MODIS/VIIIRS); emerging hyperspectral & UAV platforms	Spatial-temporal trade-offs; limited harmonisation across sensors; underrepresentation of UAV/hyperspectral studies	Standardised cross-sensor harmonisation; benchmark multi-sensor datasets; coordinated satellite UAV in situ designs	Enables consistent spatio-temporal coverage and scalable monitoring
Target Parameters	Robust retrieval of optically active parameters (Chl-a, TSM, turbidity); improving CDOM inversion	Weak generalisation for nutrients (TP, TN); indirect inference via surrogates; limited optical diversity coverage	Optical Water Type (OWT)-aware modelling; expanded in situ datasets for nutrients; hybrid surrogate-physics approaches	Extends monitoring beyond optical proxies toward biogeochemical relevance
Algorithms & Models	ML/DL models outperform empirical models; growing use of hybrid and physics-guided learning	Many ML/DL models remain data-driven (P0–P2); limited interpretability and transferability	Formal integration of physics (P3–P4); differentiable RT layers; deployment-oriented architectures	Improves physical consistency, robustness, and regulatory trust
Validation Design	Cross-validation is widely reported; increasing awareness of independent validation	Performance inflation under CV; scarce inter-site/inter-sensor testing; inconsistent reporting	Mandatory stratification (CV vs. independent); multi-region benchmarks; transparent reporting standards	Prevents over-optimistic claims; supports decision-grade evaluation
Uncertainty Quantification (UQ)	Growing recognition of the importance of uncertainty; isolated and non-standardised adoption of probabilistic models	Dominance of accuracy-only reporting (U0); lack of uncertainty propagation	Explicit probabilistic UQ (U3–U4); standardized CI/PI reporting; uncertainty-aware loss functions	Enables risk-informed interpretation and policy use
Calibration & In Situ Integration	Improved satellite field matchups; emerging autonomous sensors	Sparse, biased in situ data; weak coupling to UQ and validation	Continuous in situ networks; joint calibration UQ frameworks	Strengthens credibility and long-term reliability
Operational Deployment	Cloud platforms and edge-ready ML are emerging	Limited discussion of latency, versioning, governance, and user needs	End-to-end pipelines; model versioning; stakeholder co-design	Transition models from proof-of-concept to operational services
Reproducibility & Standards	PRISMA-guided reviews; increasing open datasets	Heterogeneous protocols; limited reproducible benchmarks	Open-source pipelines; standardised benchmarks and metadata	Facilitates comparability and cumulative progress

Importantly, uncertainty in satellite-based water quality monitoring does not arise solely at the model-prediction stage but accumulates across the full observation-to-inference

chain. Upstream sources include data acquisition uncertainty, such as sensor noise, viewing-illumination variability, geolocation mismatch, and temporal mismatch between satellite overpass and field sampling; preprocessing uncertainty, including masking, resampling, spectral harmonization, feature construction, and cross-sensor calibration; and atmospheric correction uncertainty, arising from AC product choice, aerosol assumptions, adjacency effects, and reduced signal stability in spectrally sensitive regions such as the blue bands. Additional uncertainty is introduced through reference-data construction, including sparse or biased in situ observations, laboratory inconsistency, and limited spatial or temporal representativeness of calibration and validation datasets. These upstream uncertainties interact with downstream model uncertainty, including epistemic and aleatoric components, and can substantially influence reported retrieval accuracy, robustness, and transferability. A comprehensive interpretation of validation outcomes, therefore, requires that uncertainty be treated as a property of the monitoring chain rather than of model outputs alone. Over the past decade, uncertainty quantification (UQ) has evolved from implicit or qualitative consideration into an increasingly explicit and formalised component of validation practice in satellite-based water quality monitoring. A growing body of studies now incorporates error bands, prediction intervals, ensemble dispersion, residual diagnostics, and probabilistic uncertainty surfaces to support risk-aware interpretation and decision-making under dynamic environmental conditions [104,125]. Nevertheless, adoption remains uneven, and uncertainty handling is still frequently concentrated at the prediction stage, with limited explicit treatment of uncertainty propagation from acquisition, preprocessing, atmospheric correction, and reference-data construction. To systematically synthesise how uncertainty is operationalised across the reviewed literature, Table 11 introduces an evidence-based coding framework that classifies UQ practices from purely deterministic accuracy reporting to fully integrated, decision-oriented uncertainty-aware validation designs. While this taxonomy provides a structured overview of reported practice, it does not by itself resolve the broader issue of whether upstream uncertainty sources are explicitly characterised, propagated, or incorporated into performance interpretation. Accordingly, uncertainty levels (U0–U4) are treated as explanatory moderators in the performance synthesis, along with sensor class, validation strategy, and physics-integration level, as detailed in Section 3.5. Further discussion of how optical regime diversity and Optical Water Types (OWTs) shape epistemic uncertainty and cross-domain generalisation is provided in Section 3.6. Taken together, these analyses support a broader interpretation of uncertainty as a system-level property that emerges from acquisition, preprocessing, modelling, and deployment conditions, rather than as a narrow by-product of model prediction alone.

While Table 11 establishes an evidence-based taxonomy for classifying uncertainty quantification (UQ) practices in satellite-based water quality retrieval models, a tabular scheme alone does not convey how these practices are distributed across the broader modelling landscape. To explicitly situate UQ adoption relative to the degree of physical knowledge integration, Figure 7 provides a two-dimensional synthesis mapping physics-integration level (P0–P4) against UQ levels (U0–U4). This visual aggregation enables rapid identification of dominant methodological regimes and reveals systematic imbalances between modelling sophistication and uncertainty-aware validation. In particular, the clustering of reported configurations at lower physics-integration levels, with accuracy-only evaluation, underscores a persistent gap between advanced retrieval modelling and operationally meaningful uncertainty characterisation, motivating the need for more tightly coupled physics-informed and uncertainty-aware frameworks.

Table 11. An extended uncertainty quantification (UQ) coding scheme across the satellite water quality monitoring chain. The table expands the U0–U4 framework by incorporating uncertainty sources beyond model predictions, including those arising from data acquisition, preprocessing, and atmospheric correction. It highlights the limited but emerging consideration of upstream uncertainties and their implications for validation robustness and decision support.

Code	UQ Level	Definition (Evidence-Based)	Primary Uncertainty Scope	Typical Indicators in the Literature	Upstream Sources Considered	Implications for Decision Support
U0	No UQ	No uncertainty information reported	Prediction-only (deterministic)	Point estimates only; accuracy metrics without uncertainty	None explicitly considered	Not suitable for decision-grade applications
U1	Implicit/Diagnostic UQ	Uncertainty is discussed qualitatively or inferred indirectly	Prediction-focused, limited upstream awareness	Residual analysis; cross-validation errors; sensitivity discussion	Indirect recognition of data variability, but not formalised	Limited interpretability; weak risk awareness
U2	Empirical/Error-based UQ	Uncertainty quantified via empirical error statistics	Prediction + partial validation-stage uncertainty	RMSE maps; standard deviation; validation error propagation	Reference data uncertainty partially captured; upstream sources largely unaccounted	Basic confidence awareness; site-dependent reliability
U3	Probabilistic/Statistical UQ	Explicit probabilistic uncertainty estimation	Prediction-level + partial modelling uncertainty	Prediction intervals; Bayesian models; ensemble variance	Model uncertainty addressed; upstream sources (e.g., AC, preprocessing) rarely propagated	Suitable for risk-informed decisions, but incomplete uncertainty coverage
U4	Integrated/Decision-oriented UQ	Uncertainty is embedded across the model architecture and outputs, with partial system awareness	Prediction + modelling + limited pipeline integration	Bayesian DL; mixture density networks; epistemic vs. aleatoric separation	Partial consideration of input/data uncertainty; limited explicit propagation of acquisition, preprocessing, and AC uncertainty	Decision-grade outputs, though full end-to-end uncertainty remains, are rarely implemented

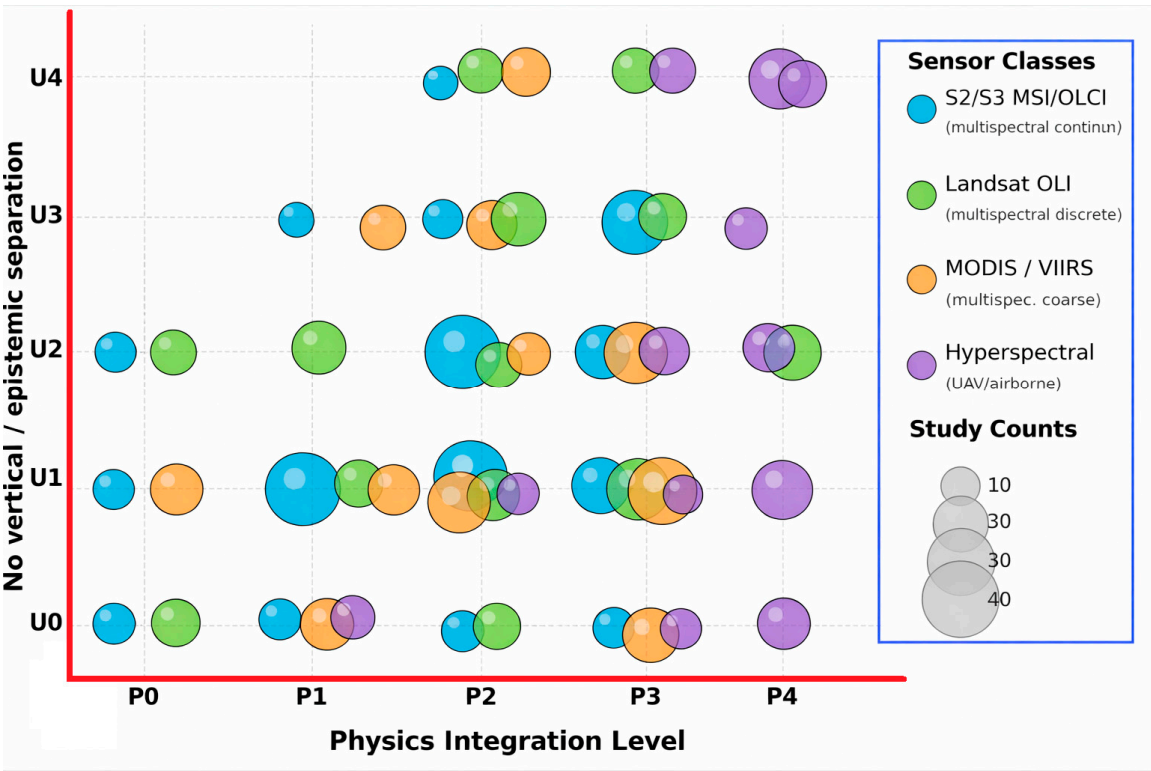


Figure 7. Distribution of uncertainty quantification (UQ) practices across physics-integration levels in satellite-based water quality retrieval models. Bubble size denotes the number of reported model configurations for each combination of physics integration (P0–P4) and UQ level (U0–U4), revealing a strong concentration in low-integration, accuracy-only regimes and a persistent gap in uncertainty-aware validation.

In multi-sensor and large-scale systems, ensemble strategies such as Bayesian model averaging and Mixture Density Networks have demonstrated strong potential for producing harmonised retrievals with explicitly quantified uncertainty [38]. Nevertheless, substantial obstacles remain. Deep learning models routinely achieve state-of-the-art accuracy for chlorophyll-a and turbidity but often function as “black boxes,” limiting interpretability and complicating regulatory acceptance [46,72]. Recent advances address this limitation through uncertainty-aware DL architectures that explicitly separate epistemic (model-driven) and aleatoric (data-driven) uncertainty, yielding calibrated probabilistic outputs suitable for decision support [62]. Complementarily, edge-oriented implementations demonstrate that uncertainty-aware, quantified models can operate efficiently on constrained hardware, bringing validation and uncertainty assessment closer to near-real-time operational contexts [68]. The literature further underscores the critical role of contextual and seasonal variability in validation outcomes. For example, Random Forest models for TN can achieve R^2 values approaching 0.92 during rainy seasons but degrade substantially during dry periods, highlighting strong seasonal dependence [42]. Long-term validation against multi-year in situ datasets shows that Landsat and GCOM-C/SGLI products reliably capture interannual trends but underperform under extreme optical conditions or high-turbidity regimes [48,126]. At the global scale, CNN-based frameworks report high overall accuracy for chlorophyll-a ($R^2 > 0.82$) across diverse sites, but their reliability declines in shallow, heterogeneous, or optically complex waters, reinforcing the need for environment-specific validation protocols [56]. Table 12 synthesises validation practices reported in the reviewed literature by organising methodological choices and associated challenges across data requirements, model performance, uncertainty and generalizability, and preprocessing impacts. While this synthesis captures how validation is approached conceptually, it does not address how consistently uncertainty information is quantified and reported in practice.

Table 12. Synthesis of validation practices and uncertainty sources in satellite-based water quality monitoring. The table extends conventional validation summaries by explicitly identifying uncertainty sources across the monitoring chain, including data acquisition, preprocessing, atmospheric correction, and modelling stages, and highlighting their implications for retrieval accuracy, transferability, and robustness.

Category	Key Practices & Methods	Associated Challenges & Limitations	Uncertainty Source Type	Implications for WQ Outputs	Representative References
Data Acquisition & Match-up	Validation with synchronised in situ matchups; cross-season and cross-ecosystem sampling; benchmark datasets	Scarcity of well-distributed samples; temporal mismatch (satellite vs. in situ); spatial representativeness; seasonal bias; geographic gaps	Acquisition uncertainty (sensor noise, revisit timing, viewing geometry, matchup inconsistency)	Bias in calibration datasets, reduced representativeness, and instability in cross-site generalisation	[35,64,96,127]
Model Performance	ML/DL outperform empirical approaches for optical WQPs; RF and CatBoost are effective for nutrient inference; hybrid models integrate contextual drivers (land use, meteorology)	Risk of overfitting; confounding correlations; degraded accuracy in cross-region or near-real-time validation; strong seasonal dependence	Model uncertainty (epistemic + aleatoric; data-driven bias)	Performance inflation under cross-validation; reduced transferability; sensitivity to domain shift	[36,37,40,55,60,75]
Uncertainty & Generalizability	Explicit UQ via prediction intervals, geostatistical mapping, Bayesian DL, and MC-dropout; complementary interpretability (e.g., SHAP)	Black-box behaviour of DL; inconsistent generalisation across ecosystems; lack of standardised uncertainty reporting	Prediction-level uncertainty (probabilistic outputs)	Improved risk awareness, but incomplete uncertainty representation across the pipeline	[38,46,62,67,68,78]

Table 12. Cont.

Category	Key Practices & Methods	Associated Challenges & Limitations	Uncertainty Source Type	Implications for WQ Outputs	Representative References
Preprocessing & Atmospheric Correction	Atmospheric correction (e.g., C2RCC, ACOLITE, LaSRC); radiometric cross-calibration; sensor harmonization; masking and resampling	Propagation of AC errors to retrievals; adjacency effects; inter-product biases; sensitivity to AC selection; spectral inconsistencies across sensors	Preprocessing & AC uncertainty (AC assumptions, aerosol models, adjacency, harmonisation errors)	Systematic bias in reflectance inputs; reduced comparability across sensors; uncertainty amplification in downstream models	[48,126,128–130]

To address this gap, Table 13 provides a complementary quantitative meta-summary, reporting median performance metrics stratified by parameter, sensor class, and model family, together with an evidence-based uncertainty quantification (UQ) flag linked to the coding scheme defined in Table 11. This joint presentation enables a clear distinction between reported predictive accuracy and the explicit treatment of uncertainty across studies. However, it is important to note that Table 13 primarily reflects uncertainty as reported at the model prediction stage and does not explicitly capture upstream sources of uncertainty arising from data acquisition, preprocessing, atmospheric correction, or reference-data construction, which are discussed separately in Sections 3.4 and 3.5 and summarised in Table 12. Collectively, recent reviews emphasise that standardised benchmarking datasets, explicit UQ reporting, and fully reproducible validation protocols are indispensable prerequisites for translating methodological advances into trusted, operational water quality products [38,96]. Without such rigour, even the most sophisticated retrieval models remain confined to scientific proof of concept rather than decision-grade deployment.

To address this gap, Table 13 provides a complementary quantitative meta-summary that reports median performance metrics stratified by water quality parameter, sensor class, and model family, together with an evidence-based uncertainty quantification (UQ) flag linked to the coding scheme defined in Table 11. Rather than implying widespread adoption of uncertainty-aware practices, this joint synthesis explicitly separates reported predictive accuracy from the largely absent or non-standardized treatment of uncertainty across the reviewed literature, thereby confirming the dominance of accuracy-only reporting identified in the preceding analyses. It should be noted, however, that the UQ information summarized in Table 13 primarily reflects uncertainty as reported at the model prediction stage and does not explicitly capture upstream sources of uncertainty arising from data acquisition, preprocessing, atmospheric correction, or reference-data construction, which are addressed separately in Sections 3.4 and 3.5 and synthesised in Table 12. By consolidating these patterns, Table 13 characterises the prevailing state of practice and reveals a persistent disconnect between methodological innovation and the routine, reproducible reporting of uncertainty. This disconnect directly motivates the structured retrieval architecture illustrated in Figure 8, which translates the empirical deficiencies identified in Tables 11–13 into an explicit, end-to-end OWT-conditional pipeline. In this framework, optical water type (OWT) classification precedes model selection, parameterisation, uncertainty quantification, and validation, thereby operationalising regime-aware retrieval, OWT-specific uncertainty propagation, and water-type-stratified validation. By embedding UQ, calibration, and validation as integral components of the retrieval process rather than optional post hoc analyses, the framework provides a concrete response to the accuracy-dominated reporting practices documented in the quantitative synthesis.

Consistent with recent synthesis studies, this alignment underscores the necessity of standardised benchmarking datasets, explicit and comparable UQ outputs, and fully reproducible validation protocols as prerequisites for translating algorithmic advances

into trusted, operational water quality products [85]. Without such rigor, even technically sophisticated retrieval models are likely to remain confined to scientific proof-of-concept rather than decision-grade deployment.

Table 13. Meta-summary of dominant quantitative performance reporting and uncertainty quantification (UQ) practices in satellite-based water quality retrieval studies. For each representative study–parameter combination, median performance metrics (e.g., R^2) are summarised, along with an evidence-based UQ classification, following the coding scheme defined in Table 11. The table reflects the dominant, standardised reporting practices that support cross-study quantitative synthesis. Although advanced uncertainty-aware approaches (e.g., Bayesian inference, ensemble modelling, or probabilistic neural networks) are reported in a limited number of studies, they are typically documented in heterogeneous or non-standardised forms and therefore do not materially alter the prevalence of accuracy-only (U0) reporting observed in the synthesised evidence. The meta-summary thus captures the prevailing state of practice rather than the full methodological spectrum.

Water Quality Parameter	Dominant Sensor(s)	Dominant Model Family	Median R^2 (IQR)	Typical Sample Size Range	Dominant UQ Reporting Practice	Assigned UQ Code
Chlorophyll-a (Chl-a)	Sentinel-2, MODIS, Landsat-8	ML/DL (CNN, RF)	0.82 (0.75–0.90)	50–>300	Accuracy metrics only (R^2 , RMSE); no predictive intervals	U0 (accuracy-only)
Total Suspended Matter (TSM)	Landsat-8, Sentinel-3	Physics-based/ML	0.80 (0.78–0.85)	40–>300	Error statistics only; uncertainty not propagated	U0 (accuracy-only)
Turbidity	Landsat-8, Sentinel-2	Empirical/ML	0.88 (0.85–0.90)	30–150	Error statistics only; site-specific calibration	U0 (accuracy-only)
Total Phosphorus (TP)	Sentinel-2	ML (RF, XGB)	0.68 (0.64–0.74)	40–120	No explicit uncertainty characterisation	U0 (accuracy-only)
Total Nitrogen (TN)	Sentinel-2	ML (RF/hybrid)	0.75 (0.70–0.80)	Seasonal subsets	Seasonal stratification only; no predictive intervals	U0 (accuracy-only)
CDOM	Sentinel-2	Physics-based/ML	0.82 (single-study)	<50	Uncertainty not reported	U0 (accuracy-only)
Dissolved Oxygen/Secchi depth	Landsat-8	Empirical/ML	0.35–0.45	<50	Accuracy metrics only	U0 (accuracy-only)
Long-term trends (multi-year)	MODIS, GCOM-C/SGLI	Product-level analysis	≥ 0.88	Global match-ups	Product comparison; no uncertainty envelopes	U0 (accuracy-only)

While Table 13 provides a quantitative meta-summary of reported performance metrics and dominant uncertainty quantification practices aggregated across studies, a purely tabular representation inevitably limits direct comparison of systematic differences across validation designs and sensor classes. Contrasts between models evaluated using internal cross-validation and those assessed against spatially or temporally independent test datasets, as well as sensor-dependent patterns in reported accuracy, are not readily discernible from summary statistics alone. To address this limitation, Figure 9 presents a stratified visual synthesis of median R^2 values and their dispersion, explicitly contrasting cross-validation and independent validation outcomes across water quality parameters and sensor classes. This visualisation enables rapid identification of systematic performance inflation associated with internal validation strategies and highlights sensor- and parameter-specific variability that is central to interpreting the robustness, generalizability, and transferability of reported water quality retrieval models. It is important to note, how-

ever, that systematic, head-to-head quantitative comparisons of model performance across different OWT classification frameworks remain limited in the reviewed literature. As a result, performance differences associated with OWT frameworks cannot be robustly assessed through aggregated metrics alone. The OWT-related synthesis presented in this study is therefore comparative and evidence-informed rather than strictly meta-analytic and should be interpreted as identifying consistent patterns and context-dependent tendencies rather than definitive quantitative rankings.

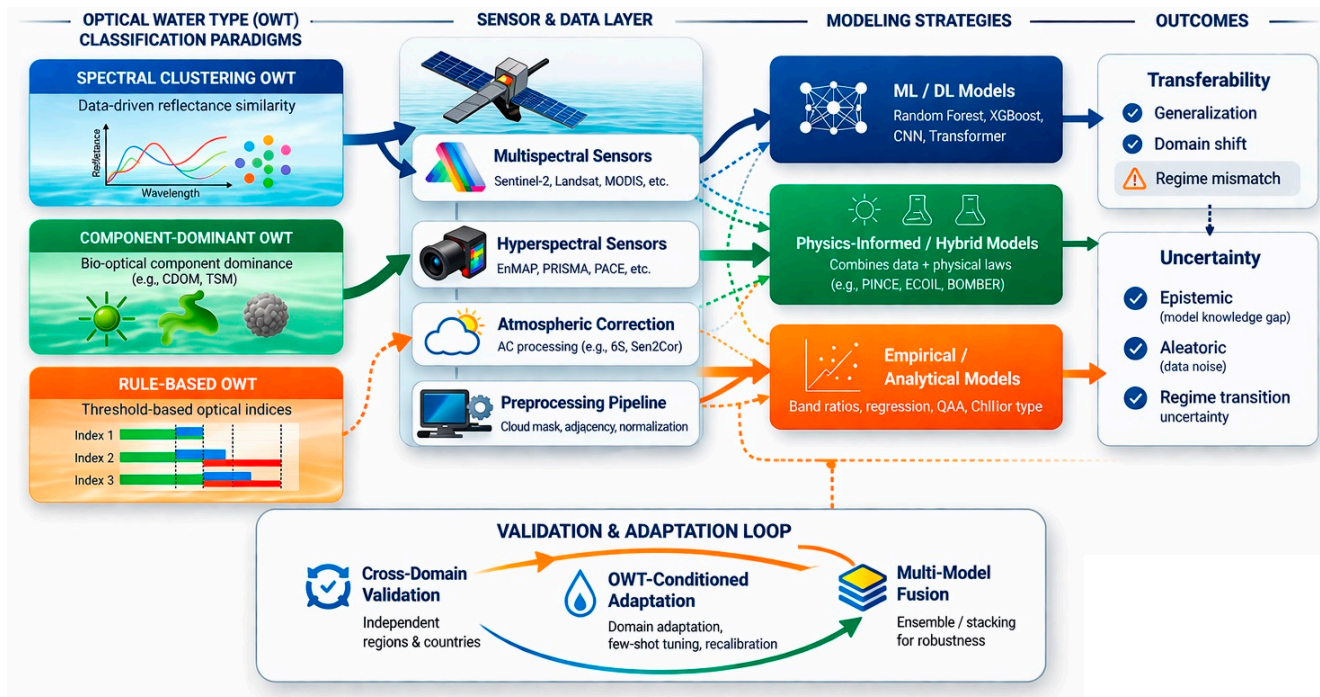


Figure 8. OWT–sensor–model alignment framework for satellite-based water quality retrieval. Optical Water Types (OWTs) are used as conditioning variables to guide model selection, parameterisation, validation, and uncertainty quantification. The framework highlights context-dependent alignment between OWT classification, sensor characteristics, and modelling strategies to enhance transferability, physical consistency, and robustness across heterogeneous water bodies. Color-coded panels represent major functional components (OWT classification paradigms, sensor and data layers, modelling strategies, and outcomes), while solid arrows indicate primary data and processing flows and dashed arrows denote auxiliary or feedback relationships within the validation and adaptation loop. The graphical layout of this conceptual figure was assisted by a generative AI tool; all scientific content, relationships, and interpretations were defined and validated by the authors.

Table 14 further refines the quantitative synthesis by stratifying reported model performance according to validation protocol (cross-validation versus independent validation), water quality parameter, sensor class, and model family. For each stratum, the table reports the number of model configurations, the median R^2 , the interquartile range (IQR), and the dominant uncertainty quantification (UQ) code assigned according to the evidence-based scheme defined in Table 11. While Table 14 provides a detailed numerical breakdown, Figure 9 complements this synthesis by visually contrasting performance distributions across sensor classes and validation designs, enabling rapid identification of systematic performance inflation under cross-validation and sensor-dependent variability. Together with Tables 12 and 13, this layered presentation offers a structured and internally consistent overview of how validation design, performance reporting, and uncertainty treatment co-occur through the reviewed literature.

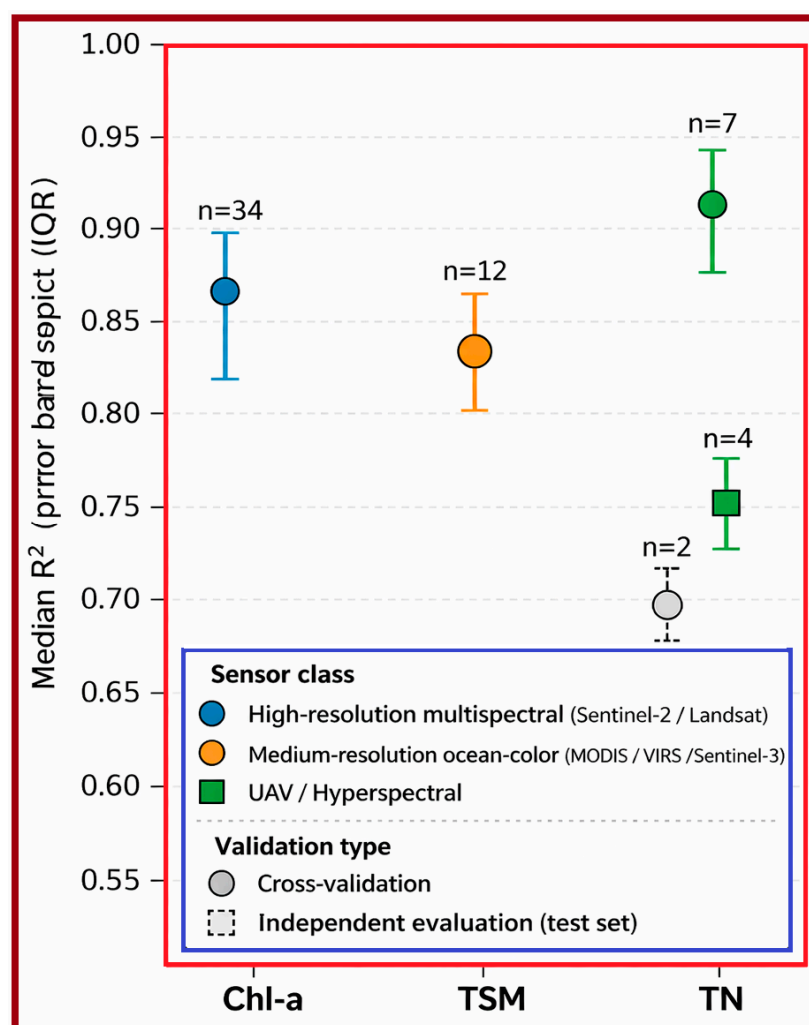


Figure 9. Stratified meta-summary of model performance in satellite-based water quality retrieval. Median R^2 and interquartile ranges are shown across parameters, sensor classes, and validation protocols, highlighting systematic performance inflation under cross-validation and sensor- and parameter-dependent variability.

Table 14. Quantitative meta-summary of model performance across reviewed water quality studies, stratified by parameter, sensor class, model family, and validation protocol. The table reports model counts, median and interquartile ranges of R^2 , and the dominant category of uncertainty quantification, revealing a systematic lack of explicit uncertainty reporting across most configurations.

Parameter	Sensor Class	Model Family	Validation Type	n (Models)	Median R^2	IQR (R^2)	Dominant UQ Code (Mode; n Studies)
Chl-a	Sentinel-2/Landsat	ML/DL	Cross-validation	34	0.88	0.06	U0 (accuracy-only; n = 29)
Chl-a	Sentinel-2/Landsat	ML/DL	Independent	18	0.79	0.09	U0 (accuracy-only; n = 16)
TSM	Sentinel-3	Physics-based	Independent	12	0.85	0.05	U0 (accuracy-only; n = 11)
TN	Sentinel-2	ML	Cross-validation	7	0.92	0.04	U0 (accuracy-only; n = 6)
TN	Sentinel-2	ML	Independent	4	0.81	0.08	U0 (accuracy-only; n = 4)

3.5. Multi-Source Data Integration

Configuration-level synthesis across the reviewed studies indicates that multi-source data integration has transitioned from an auxiliary enhancement strategy to a structural component of contemporary satellite-based water quality (WQ) monitoring architectures. Rather than merely aggregating heterogeneous datasets, integration is increasingly opera-

tionalised as a deliberate mechanism to mitigate structural constraints inherent in single-sensor systems, particularly the persistent trade-offs among spatial resolution, temporal density, spectral richness, and archive continuity [90]. No individual platform simultaneously satisfies the resolution–revisit–continuity triad: Sentinel-2 provides fine spatial detail, MODIS/VIIRS ensures high temporal frequency, and Landsat provides multi-decadal stability [131–133]. The reviewed evidence demonstrates that integration strategies are therefore motivated less by incremental accuracy gains than by the need to reconcile these structural asymmetries within transferable and operational monitoring systems. Across the literature, integration manifests along a spectrum of increasing structural coupling. At the foundational level, sensor harmonisation addresses radiometric calibration discrepancies, atmospheric correction (AC) inconsistencies, spectral response mismatches, and geolocation biases that otherwise propagate systematic errors into downstream retrieval models [129,134]. Without such preprocessing alignment, fusion-induced performance gains are often offset by cross-mission bias amplification. At an intermediate level, spatio-temporal fusion strategies attempt to resolve resolution–revisit conflicts by combining moderate-resolution imagery (e.g., Landsat, Sentinel) with high-frequency observations (e.g., MODIS) through compositing, downscaling, or learning-based super-resolution frameworks [131,135]. These approaches expand temporal coverage and reduce data gaps, yet introduce additional uncertainty associated with cross-scale interpolation and spectral transfer assumptions. At the highest level of integration, model architectures explicitly incorporate multi-source inputs, including satellite imagery, meteorological drivers, land-use variables, and in situ time series within unified learning frameworks. CNN–LSTM hybrids and ensemble-based configurations exemplify this structurally coupled paradigm, linking spatial reflectance patterns with temporal and contextual dynamics [101,136].

Importantly, multi-source data integration should be conceptually distinguished from multi-model fusion, although the two are often co-implemented in practice. While data integration focuses on combining heterogeneous input sources, multi-model fusion aims to integrate models with complementary inductive biases and strengths. For example, empirical models provide transparency and computational efficiency in locally constrained settings, physics-based models ensure consistency with radiative transfer processes, and ML/DL approaches capture nonlinear relationships and high-dimensional interactions. Fusion strategies, including ensemble learning, stacking, and mixture-of-experts, can combine these complementary capabilities to improve robustness under varying optical regimes, reduce sensitivity to domain shifts, and mitigate reliance on any single modelling assumption. Although such approaches remain relatively underexplored and inconsistently standardised across the reviewed literature, they represent a promising direction for enhancing model generalisation and operational reliability. Context-aware models incorporating precipitation, land cover, and hydrological drivers further demonstrate that the explanatory power for optically weak parameters such as nutrients is often contingent on multi-source contextualization rather than on spectral signal strength alone [62,75]. In this context, auxiliary data can be broadly classified into meteorological variables (e.g., precipitation, temperature), land-use and watershed characteristics, hydrological indicators (e.g., runoff, discharge), and in situ or IoT-based observations. Each class contributes to reducing different sources of uncertainty: meteorological inputs help capture event-driven variability and temporal dynamics; land-use data constrain spatial heterogeneity and watershed-driven nutrient loading; hydrological variables improve representation of transport processes; and in situ/IoT observations enhance calibration density and temporal alignment. Importantly, these improvements do not arise from enhanced spectral observability, but rather from contextual disambiguation of proxy relationships, particularly for non-optical parameters whose inversion depends on indirect inference pathways.

Importantly, integration extends beyond satellite–satellite fusion. The reviewed literature increasingly documents the coupling between Earth observation data and ground-based infrastructure, including hydrological stations, autonomous micro-hyperspectral spectrometers, and IoT sensor networks [137,138]. These configurations enhance calibration density, enable near-synchronous validation, and partially address the temporal mismatch between satellite overpass schedules and in situ sampling. Long-term benchmarking against field stations has demonstrated improved characterisation of interannual trends and seasonal variability [126] yet also revealed sensitivity to AC scheme selection and cross-product inconsistencies [48]. Consequently, multi-source integration does not eliminate structural uncertainty; rather, it redistributes uncertainty across harmonisation, scaling, and contextual modelling stages. When evaluated using the physics-integration (P0–P4) and uncertainty (U0–U4) coding frameworks introduced earlier, multi-source configurations frequently cluster in intermediate integration regimes, where contextual or physics-guided inputs are incorporated without fully embedding physical constraints into the model objective. While such configurations often report improved in-domain performance relative to single-sensor baselines [38,133], independent validation across regions and optical regimes remains comparatively limited. Evidence of enhanced transferability is therefore conditional rather than universal, and performance gains must be interpreted in relation to validation design and uncertainty handling. In several cases, ensemble-based fusion and mixture density networks have begun to couple harmonisation with probabilistic output generation, thereby partially aligning multi-source strategies with uncertainty-aware operational frameworks [68]. However, explicit uncertainty propagation across harmonisation and fusion stages remains inconsistently implemented.

Operational scalability introduces an additional layer of complexity. Multi-source pipelines must accommodate heterogeneous file formats, asynchronous acquisition schedules, large data volumes, and cloud-based computational orchestration [139]. Hybrid ML/DL architectures, while flexible, remain dependent on curated and representative training datasets, and their interpretability may degrade as input dimensionality increases [67,72]. Consequently, the integration process itself becomes a potential source of epistemic uncertainty if harmonisation assumptions, scaling transformations, and contextual covariates are not transparently documented. Collectively, the synthesised evidence indicates that multi-source integration is neither a peripheral refinement nor a guaranteed pathway to universal transferability. Rather, it represents a structurally necessary yet methodologically demanding evolution toward operational WQ monitoring systems that balance spatial detail, temporal continuity, contextual awareness, and long-term stability. When harmonisation rigour, validation stratification, and explicit uncertainty quantification are jointly implemented, the integration architecture demonstrates clear potential for enhancing robustness and decision relevance. In the absence of such safeguards, however, fusion strategies, whether data-driven or model-based, risk amplifying bias and obscuring model limitations. The conceptual workflow underlying these integration stages is summarised in Figure 10, and a structured synthesis of representative multi-source configurations, including their data sources, methodological coupling, and primary contributions, is provided in Table 15.

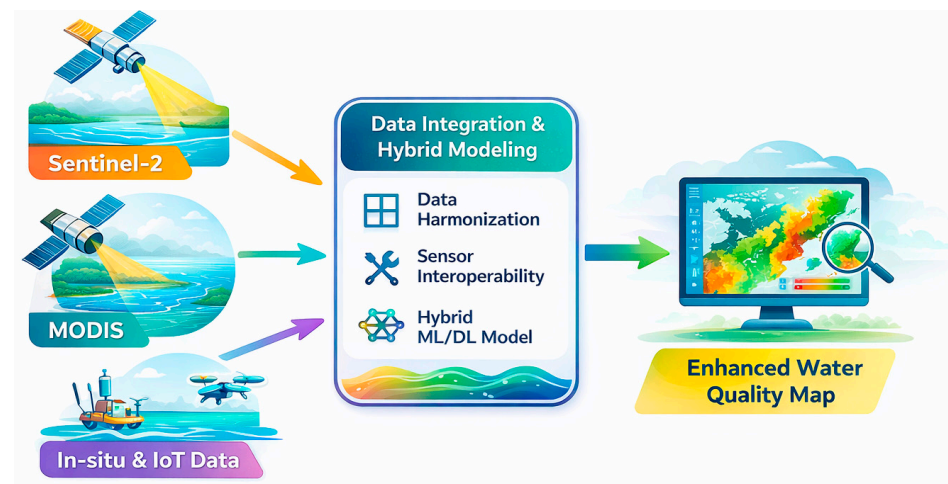


Figure 10. Conceptual schematic of a multi-source data integration and hybrid modelling framework for satellite-based water quality monitoring. The framework integrates complementary Earth observation data (e.g., Sentinel-2, MODIS) with in situ and IoT measurements through sensor harmonisation and interoperability, coupled with hybrid (machine learning and deep learning) models to generate enhanced water quality products. This approach improves spatial–temporal coverage, robustness, and consistency across diverse environmental conditions. The visual layout was developed with the assistance of a generative artificial intelligence tool for graphical design; all scientific content and interpretations were defined and validated by the authors, with AI use strictly to visualisation support.

Table 15. High-level synthesis of multi-source data integration strategies in satellite-based water quality studies, summarising data sources, methodological frameworks, and key contributions across recent literature.

Study (Reference)	Integrated Data Sources	Methodology	Key Contribution
[133]	Landsat, Sentinel, MODIS	Hybrid ML + cluster-based empirical models	Improved accuracy by combining multiple satellites
[127]	Landsat + Hydrological stations	ML regression	Fusion with field data enhanced estimation reliability
[138]	Landsat-9 + GIS + IoT stations	Hybrid architecture	Framework for integrated real-world monitoring
[38]	Landsat-8, Sentinel-2/3	Ensemble (Mixture Density Networks)	Developed a technical framework for sensor harmonisation
[132]	PlanetScope, Landsat-8, Sentinel-2	Comparative analysis	Assessed trade-offs between public vs. commercial sensors
[129]	MODIS Aqua + Terra	Radiometric cross-calibration	Highlighted technical challenges of sensor harmonisation
[68]	Sentinel-2 + optimised CNN	Quantisation-aware training (QAT)	Demonstrated edge-ready integration with efficient uncertainty handling
[62]	Sentinel-2 + meteorology + land use	ML (Extra Trees, XGBoost, SHAP)	Revealed contextual drivers (precipitation, land cover) of nutrient variability
[75]	Sentinel-2 + meteorology + land management	RF/XGBoost	Showed turbidity regimes shaped by socio-hydrological drivers
[126]	Landsat-8 OLI + long-term field stations	AC (C2RCC, l2gen) + RF	Validated interannual trends; highlighted AC constraints
[48]	GCOM-C/SGLI + global AC products	Multi-product intercomparison	Identified wavelength-dependent biases and AC scheme effects
[67]	Sentinel-3 + multi-site datasets	Global CNN framework	Demonstrated cross-site generalisation and limits in shallow waters

3.6. Optical Water Types and Implications for Transferability and Uncertainty

Optical Water Types (OWTs) provide a physically grounded framework for representing the intrinsic optical heterogeneity of inland and coastal waters by classifying systems according to dominant absorption and scattering regimes, rather than geographic or hydrological descriptors. Global OWT schemes derived from in situ spectroscopy and satellite reflectance clustering capture systematic variability in watercolour, constituent composition, and radiative-transfer behaviour across diverse environments [32,40]. Within the configuration-level synthesis of this review, optical regime heterogeneity emerges as a structural determinant of model transferability, operating alongside but not reducible to sensor characteristics and algorithmic choice. Across the reviewed studies, retrieval performance is consistently conditioned by the degree of optical consistency between calibration and deployment domains. Models trained within a limited OWT envelope frequently exhibit degraded performance when applied to optically distinct waters, even when sensor characteristics and nominal target variables remain unchanged. This effect is particularly evident in empirical and purely data-driven configurations (P0–P1), where spectral–parameter relationships implicitly encode regime-specific covariance structures. Consequently, performance variability should be interpreted in relation to optical-regime diversity, which serves as a latent constraint on generalisation rather than as a secondary effect of model complexity.

OWTs are implemented across multiple classification paradigms, each with distinct implications for adaptability, interpretability, and transferability. As synthesised in Table 16, three broad categories can be distinguished: (i) spectral clustering-based OWTs, (ii) component- or constituent-dominant OWTs, and (iii) rule-based or optical-property-based schemes. These paradigms differ not only in their conceptual basis but also in their compatibility with sensor characteristics and modelling strategies, forming the basis for context-dependent OWT–sensor–model alignment. Spectral clustering-based OWTs, derived from reflectance similarity patterns, are flexible and well-suited to heterogeneous environments and multisensory applications. They are particularly compatible with machine learning and deep learning approaches, in which OWT information can be incorporated as stratification variables or as conditional inputs. However, their physical interpretability is limited, and their stability may depend on consistent atmospheric correction and sensor harmonisation. In contrast, component-dominant OWTs provide stronger bio-optical interpretability by explicitly linking classes to dominant constituents (e.g., phytoplankton, suspended matter, CDOM). These frameworks are more naturally aligned with physics-informed or hybrid modelling approaches (P2–P4) but may require spectrally rich observations and are more sensitive to preprocessing uncertainties. Rule-based OWT schemes offer operational simplicity and transparency, yet their fixed thresholds and limited adaptability can reduce robustness under optically complex or transitional conditions. The synthesis further indicates that no single OWT framework is universally optimal. Instead, their effectiveness depends on the alignment between optical regime complexity, sensor spectral capability, and model architecture. In this context, OWTs should be treated not as static pre-classification tools, but as conditioning variables within the retrieval design space. Performance degradation across regions can often be more accurately attributed to mismatches between calibration and deployment OWT domains than to intrinsic model deficiencies. Approaches that incorporate OWT information as conditional inputs, stratification layers, or regime-specific parameterisation mechanisms (e.g., mixture-of-experts) tend to yield greater robustness than globally trained, regime-agnostic configurations.

Table 16. Evidence-informed comparison of major Optical Water Type (OWT) classification frameworks and their implications for modelling strategies, sensor compatibility, transferability, and uncertainty in satellite-based water quality retrieval. The synthesis is based on qualitative patterns observed across the reviewed studies rather than direct quantitative comparisons, as systematic head-to-head evaluations of OWT frameworks remain limited. Relationships should therefore be interpreted as context-dependent tendencies influenced by sensor characteristics, optical regime variability, and model design.

OWT Framework	Core Basis	Modelling Alignment	Sensor Context	Transferability Implications	Uncertainty Implications
Spectral clustering-based	Reflectance similarity (data-driven)	Commonly used with ML/DL (P0–P1) as stratification or conditional input	Widely applied with multispectral and multisensor datasets	Reported to support transferability when used for regime-aware stratification; it depends on consistency across domains	Sensitive to domain shifts; uncertainty linked to clustering stability and representativeness
Component-/constituent-dominant	Dominant optical constituents (bio-optical basis)	Aligned with physics-informed and hybrid models (P2–P4)	More compatible with spectrally rich sensors (context-dependent)	Supports transferability where constituent regimes are preserved; may degrade in mixed conditions	Enables interpretable uncertainty linked to constituent variability and model assumptions
Rule-based/optical-property-based	Threshold-based optical indices	Typically used in empirical or simplified workflows (P0–P2)	Applicable across sensors, depending on band availability	Limited transferability under changing optical conditions; suitable for stable environments	May underestimate uncertainty in transitional regimes due to rigid classification

OWTs also play a critical role in characterising uncertainty. Transitions across optical regimes, whether spatial or temporal, are frequently associated with elevated epistemic uncertainty, reflecting gaps in training data representation rather than increased observational noise [38]. In UQ terms, such regime shifts manifest as uncertainty inflation, particularly in configurations lacking explicit probabilistic treatment (U0–U1). Validation strategies spanning multiple OWT classes consistently reveal greater performance dispersion than within-regime validation, underscoring the structural linkage between optical heterogeneity and uncertainty propagation. From an operational perspective, the synthesis suggests several context-dependent tendencies rather than universal prescriptions: spectral clustering-based OWTs are generally well suited to multisensory and heterogeneous regional applications when combined with ML/DL models; component-dominant OWTs are more appropriate for spectrally rich observations and physics-informed or hybrid retrieval frameworks; and rule-based OWTs remain effective in optically stable environments when paired with empirical or semi-analytical approaches. These alignments should be interpreted as design guidelines contingent on data availability, sensor capability, and application context. Overall, OWT-awareness functions as a structural control variable for transferability optimisation and uncertainty interpretation. Embedding OWT-informed conditioning within retrieval frameworks through adaptive model selection, regime-specific parameterisation, and stratified validation provides a principled pathway toward more transferable, physically consistent, and uncertainty-aware satellite-based water quality monitoring systems.

3.7. Temporal Generalisation, Non-Stationarity, and Climate-Driven Shifts

Temporal generalisation constitutes a distinct and under-addressed dimension of transferability in satellite-based water quality retrieval, arising from the inherently non-stationary dynamics of aquatic systems. Whereas spatial transferability is predominantly shaped by optical and sensor heterogeneity, temporal transferability is governed by seasonal cycles, episodic disturbances, interannual variability, and long-term climatic forcing that modify both biogeochemical processes and optical structure. Long-term satellite

archives document substantial variability in chlorophyll-a, turbidity, and suspended matter driven by hydrological forcing, nutrient loading, and temperature-mediated ecosystem responses [97]. Nevertheless, many retrieval configurations implicitly assume temporal stationarity through reliance on calibration datasets that underrepresent extreme events, anomalous bloom years, or evolving mixing regimes. Configuration-level evidence indicates that models trained in temporally constrained regimes often degrade in performance when exposed to unseen hydrological or climatic conditions [35,140]. This degradation manifests not only in reduced predictive accuracy but also in increased epistemic uncertainty, reflecting a divergence between learned statistical mappings and evolving system behaviour. Such effects are amplified for optically inactive parameters inferred indirectly via surrogate relationships, where temporal shifts in constituent covariance alter the stability of proxy-based inference [55,92]. Under conventional random cross-validation, which preserves temporal mixing between training and test subsets, these non-stationarity effects are often masked, resulting in optimistic performance estimates. Temporal hold-out validation, multi-year hindcasting, and seasonally stratified evaluation provide more robust diagnostics of genuine temporal generalisation capacity [35].

Climate-driven shifts further intensify temporal non-stationarity. Observed warming of lake surface waters and altered stratification regimes have been linked to changes in bloom phenology, nutrient cycling, and the optical regime [1,2]. Such structural transformations challenge the long-term validity of static retrieval models and underscore the necessity of adaptive and uncertainty-aware monitoring strategies. Emerging approaches, including recurrent and hybrid DL architectures for sequence-aware modelling [113,136], periodic retraining protocols, and drift-aware uncertainty monitoring, represent early steps toward dynamic model updating in response to evolving environmental conditions. Within the integrated framework proposed in this review, temporal generalisation is therefore conceptualised not as an inherent property of a fixed model configuration, but as an operational capability supported by temporally explicit validation, regime-aware uncertainty assessment, and iterative recalibration. This perspective aligns with the distribution-based performance synthesis in Section 3 and reinforces the requirement for temporally stratified evaluation to ensure decision-grade robustness in large-scale, climate-sensitive water quality monitoring systems.

4. Discussion: A Synthesis of Advances, Challenges, and Future Directions

This review synthesises research conducted between 1980 and 2025, tracing the methodological evolution from simple empirical regressions toward increasingly sophisticated machine learning (ML), deep learning (DL), and physics-informed approaches. The broad availability of satellite missions such as Landsat, Sentinel, MODIS/VIIRS, together with emerging sensors including SDGSAT-1 and GCOM-C/SGLI, has facilitated substantial progress in water quality monitoring. Nevertheless, the synthesis highlights four systemic obstacles that continue to constrain operationalisation: limited transferability, weak interpretability, insufficient uncertainty quantification, and the absence of standardised protocols and infrastructures.

4.1. Critical Gaps

4.1.1. Transferability

Algorithms frequently fail to generalise beyond their calibration domain, with performance deteriorating across different water types, optical conditions, and climatic settings [37,123]. Empirical models are especially site-specific [34], and even advanced ML architectures degrade when trained on geographically narrow datasets, leading to substan-

tial out-of-sample errors [60]. Reviews confirm that regional bias and the scarcity of in situ observations remain structural bottlenecks [97]. Promising directions include:

1. Applying transfer learning and domain adaptation to leverage global datasets in local contexts; developing cross-sensor and cross-region training strategies, such as the global CNN framework for Chl-a [67];
2. Constructing hybrid frameworks that integrate ML with contextual drivers (meteorology, land use, hydrology), which have improved TP/TN predictions [62,75]; and
3. Adopting multi-model fusion and ensemble learning strategies to combine the complementary strengths of empirical, physics-based, and data-driven approaches. Such fusion frameworks can improve robustness under heterogeneous optical regimes, reduce sensitivity to domain shifts, and mitigate the limitations of individual model families, particularly in cross-region and cross-sensor applications where no single modelling approach consistently performs optimally.

4.1.2. Interpretability

The shift toward ML/DL has introduced significant opacity: powerful CNNs and ensemble models often function as “black boxes,” obscuring the ecological and optical processes underlying predictions [141]. This limitation restricts adoption in regulatory and legal contexts, where transparent reasoning is essential [142,143]. Recent advances provide encouraging directions:

1. SHAP-based feature attribution, which highlights the influence of precipitation and land cover on nutrient retrieval [62].
2. Global model benchmarking, which clarifies conditions under which models succeed or fail [67];
3. Physics-informed neural networks (PINNs), which embed optical principles into learning frameworks to align outputs with biophysical reality.

4.1.3. Uncertainty Quantification

Despite increasing recognition of its importance, explicit and reproducible uncertainty quantification (UQ) remains uncommon in satellite-based water quality retrieval studies. Most published work continues to report accuracy metrics such as R^2 or RMSE without calibrated prediction intervals, uncertainty maps, or probabilistic error characterisation [118]. However, operational water quality monitoring requires not only accurate predictions but also quantified and interpretable measures of reliability to support risk-aware decision-making [120]. Importantly, uncertainty in satellite-based water quality monitoring is not generated solely at the model-output stage but accumulates across the full observation-to-inference pipeline. Substantial uncertainty arises from data acquisition (e.g., sensor noise, viewing geometry, temporal mismatch between satellite overpass and in situ sampling), match-up construction (e.g., spatial representativeness and sampling bias), preprocessing and harmonization (e.g., masking, resampling, cross-sensor calibration), and atmospheric correction (AC) (e.g., aerosol assumptions, adjacency effects, and variability across AC processors). These upstream sources are rarely quantified or propagated explicitly in existing studies, yet they can exert a first-order influence on retrieval accuracy, transferability, and validation robustness. As a result, reported performance metrics often reflect a combination of model skill and uncharacterized upstream uncertainty.

A limited subset of studies has explored uncertainty-aware approaches that illustrate the potential direction of the field, but these remain methodologically heterogeneous and are not yet reported in standardised forms that support cross-study quantitative synthesis. Examples include: (i) geostatistical mapping of spatially explicit retrieval error surfaces [86]; (ii) Mixture Density Networks and ensemble machine-learning models that

generate probabilistic outputs with embedded uncertainty [38]; and (iii) uncertainty-aware deep-learning architectures that attempt to separate epistemic and aleatoric uncertainty, yielding calibrated reliability estimates relevant for real-time and risk-sensitive applications [46,68]. While these approaches demonstrate clear methodological promise, their limited adoption and inconsistent reporting practices, combined with the lack of explicit treatment of upstream sources of uncertainty, explain why accuracy-only (U0) remains the dominant pattern observed in the synthesised evidence (Table 13). Addressing this gap requires moving beyond prediction-level uncertainty toward integrated, end-to-end uncertainty characterisation that explicitly accounts for acquisition conditions, preprocessing choices, atmospheric correction strategies, and validation design within operational monitoring frameworks.

4.1.4. Standardisation and Infrastructure

Another persistent bottleneck is the lack of standardised preprocessing protocols, including atmospheric correction and radiometric calibration, as well as limited cross-sensor harmonisation and open benchmarks. Comparative studies highlight substantial biases across atmospheric correction products [48,126], underscoring the need for community-wide standards. Importantly, these inconsistencies should not be viewed solely as technical implementation differences but as major sources of uncertainty propagation that directly influence retrieval accuracy, cross-sensor comparability, and the reliability of water quality products. Variability in atmospheric correction schemes, preprocessing workflows, and harmonisation strategies can introduce systematic biases that propagate throughout the entire retrieval pipeline and are rarely explicitly quantified in current practice. Operational frameworks must therefore incorporate automated, cloud-based pipelines capable of:

1. Near-real-time (≤ 48 h) data ingestion and QA/QC;
2. Cross-sensor harmonisation [38,119];
3. Open APIs and FAIR-compliant data services for dissemination.

In this context, standardisation is not only a requirement for scalability but also a prerequisite for consistent characterisation of uncertainty across datasets, sensors, and processing chains. Without harmonised preprocessing and transparent documentation of atmospheric correction choices, uncertainty arising from upstream processing stages remains untraceable and may be misinterpreted as model error. Collectively, these infrastructure- and standardisation-related gaps, situated within the broader systemic barriers summarised in Figure 11, explain why many promising approaches remain research prototypes rather than operational tools. Addressing them requires coordinated global initiatives: the development of open multi-sensor benchmark datasets [96], advancement of interpretable and uncertainty-aware ML/DL frameworks [62,68], and investment in standardised, automated infrastructures [48]. These steps are prerequisites for achieving trusted, scalable, and policy-relevant operationalisation of water quality remote sensing.

Taken together, the interrelated limitations discussed across Sections 4.1.1–4.1.4 highlight the need for an integrated, system-level response rather than isolated methodological improvements. To explicitly connect the identified systemic barriers with the methodological solutions synthesised in this review, Figure 12 presents a barrier-to-solution roadmap that links these challenges to the core components of a transferable, physics-informed, and uncertainty-aware framework.

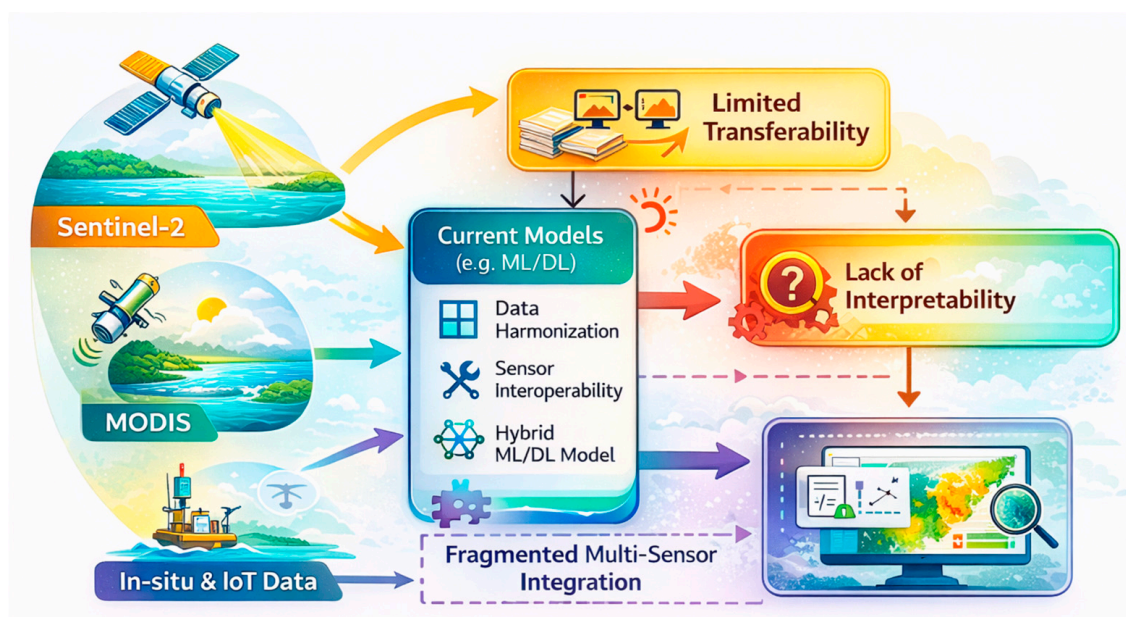


Figure 11. Conceptual illustration of key systemic barriers limiting the transition from research-oriented satellite-based water quality models to operational monitoring systems, including limited transferability, weak interpretability, insufficient uncertainty quantification, and fragmented multi-sensor integration. These interrelated constraints collectively reduce the model’s robustness, scalability, and applicability for decision support. The visual layout of this figure was developed with the assistance of a generative artificial intelligence tool for graphical design; all scientific content and interpretations were defined and validated by the authors, and AI use was limited strictly to visualisation support.

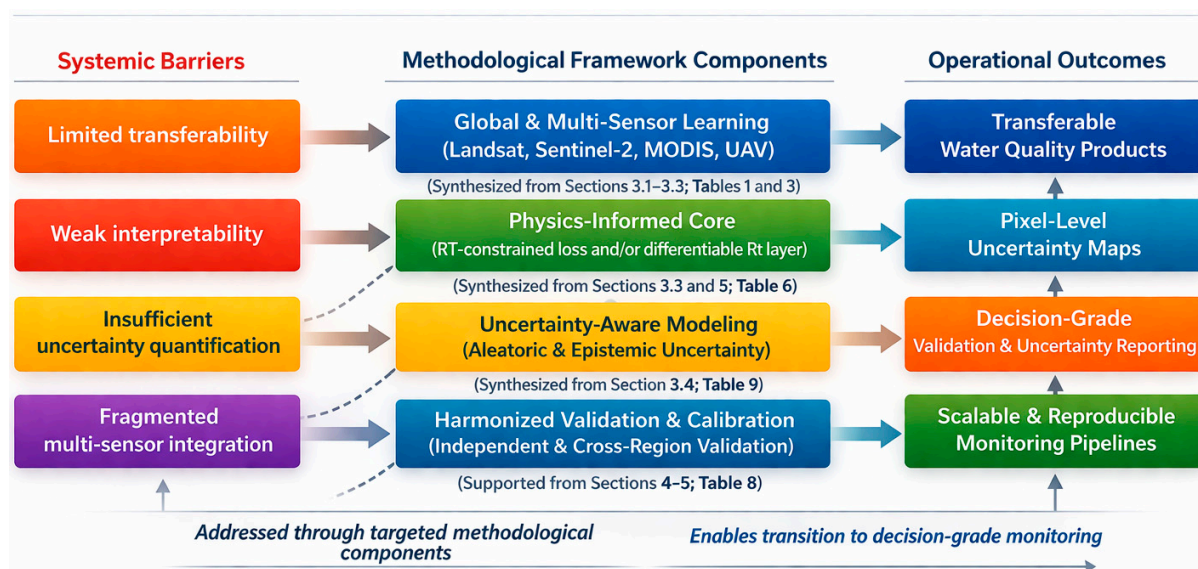


Figure 12. Barrier-to-solution roadmap illustrating how key systemic limitations identified in this review (limited transferability, weak interpretability, insufficient uncertainty quantification, and fragmented multi-sensor integration)—are addressed through an integrated set of methodological components. These include global multi-sensor learning, a physics-informed modelling core, uncertainty-aware modelling (capturing both aleatoric and epistemic uncertainty), and harmonised validation strategies. Arrows indicate the directional relationships between systemic barriers, methodological components, and operational outcomes, showing how each limitation is systematically mitigated and translated into transferable, decision-grade, and operational satellite-based water quality monitoring products.

4.2. Comparative Insights: A Quantitative Meta-Summary of Reported Model Performance

To complement the preceding synthesis of methodological and structural gaps, we conducted a quantitative, distribution-aware meta-summary of reported validation performance across the reviewed literature. The aim of this synthesis is not to estimate pooled or average performance, but to characterise the empirical distribution of reported skill levels across heterogeneous observational, ecological, and modelling contexts. Before interpreting absolute performance levels, it is essential to emphasise that reported accuracy is strongly conditioned on validation design. As demonstrated in Section 3.4, models evaluated using random or spatially non-independent cross-validation consistently report inflated performance relative to those assessed under spatially or temporally independent validation. Accordingly, all quantitative comparisons presented below are interpreted primarily through the lens of validation robustness, rather than peak reported accuracy. To mitigate within-study dependence arising from frequent evaluation of multiple models or hyperparameter configurations, the primary synthesis retained a single performance estimate per study. Following established practice in large-scale evidence mapping, we selected the best-performing model configuration as reported by the original authors. We explicitly acknowledge that this aggregation rule is susceptible to upward bias, reflecting selective reporting of peak performance rather than expected operational behaviour. Consequently, this choice is treated as a pragmatic device to enable cross-study comparability, not as an estimator of deployable performance. The sensitivity of the comparative insights to this selection strategy was therefore explicitly evaluated using alternative within-study aggregation rules, as described below.

Validation performance was primarily reported using coefficients of determination (R^2), complemented where available by error-based metrics such as RMSE and NSE. Given the pronounced heterogeneity in sensors, spatial scales, ecosystems, and validation protocols, a distribution-aware synthesis was adopted to preserve cross-study variability and avoid the interpretive distortions inherent in pooled or mean-based summaries. Figure 13 summarises the resulting comparative evidence by presenting empirical distributions of reported R^2 values for empirical and physics-based approaches relative to machine learning (ML) and deep learning (DL) models across major water quality parameters. Violin-box representations jointly depict central tendency and dispersion (median and interquartile range), enabling robust cross-study comparison while retaining information on variability, skewness, and validation-dependent performance spread.

Across the reviewed literature, chlorophyll-a (Chl-a) emerges as the most consistently and accurately retrieved parameter. Deep learning architectures, particularly convolutional neural networks (CNNs), routinely report R^2 values exceeding 0.90 for Sentinel-2-based applications [35,64], while global CNN frameworks demonstrate encouraging cross-ecosystem transferability [67]. Total suspended matter (TSM) is likewise retrieved with high accuracy: physics-based radiative transfer models commonly achieve R^2 values of approximately 0.85–0.92 [41,99], and recent advances in quantisation-aware CNNs report near-operational performance (e.g., NSE $\approx$ 0.83 with RMSE $\approx$ 3.6 mg L⁻¹), enabling edge-ready deployments [68]. In parallel, regional hybrid models that integrate satellite observations with meteorological and land-use drivers further enhance explanatory power for turbidity dynamics [75]. Turbidity retrievals, whether based on empirical ratios or ML frameworks, frequently achieve R^2 values above 0.9 [45,60]; however, empirical approaches remain highly site-specific and exhibit limited robustness beyond their calibration domains. In contrast, non-optical parameters display substantially greater dispersion in reported performance. For total phosphorus (TP), ensemble learning methods such as random forests and gradient-boosting models can reach R^2 values of approximately 0.7–0.8 under favourable conditions [62,78], yet predictive skill remains inconsistent across ecosystems and hydro-

logical regimes. Total nitrogen (TN) retrievals exhibit pronounced seasonal dependence, with strong performance reported during wet periods (R^2 up to 0.92) but sharp degradation under dry conditions [42]. Similarly, colored dissolved organic matter (CDOM) retrievals achieve moderate-to-high accuracy ($R^2 \approx 0.8$) using both physics-based and ML approaches [91,144], although transferability is frequently constrained by covariance with Chl-a and TSM and by sensitivity to atmospheric correction choices.

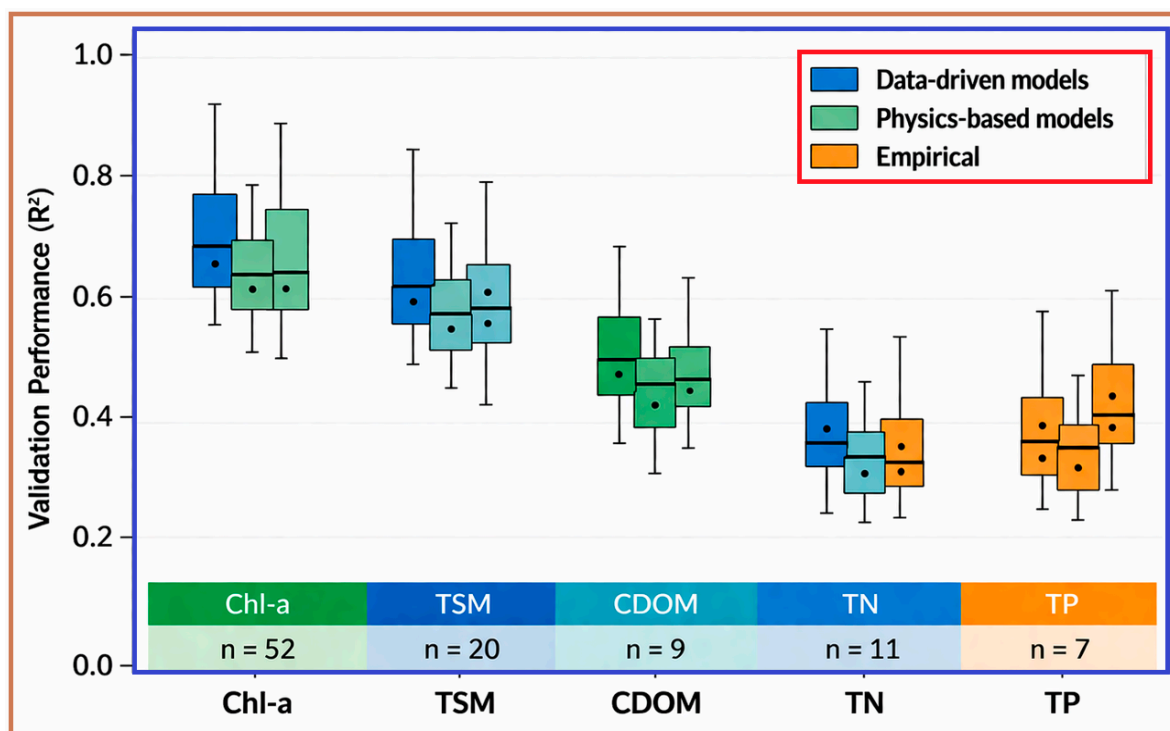


Figure 13. Distribution-aware comparison of model performance across major water quality parameters. Violin/box representations summarise the distribution of reported R^2 values for machine learning and deep learning (ML/DL) models versus physics-based and empirical approaches across key water quality parameters. Central tendency and dispersion (median and interquartile range) are shown to capture variability across studies, providing a robust, distribution-aware quantitative synthesis of evidence that goes beyond mean-based summaries.

Long-term validation efforts further indicate that atmospheric correction (AC) products exert a non-negligible influence on retrieval reliability, with systematic biases observed across AC schemes and sensors [48,126]. Collectively, these findings reinforce a fundamental trade-off among model families. Empirical approaches are computationally efficient and can achieve high local accuracy but lack robustness and transferability [96]. Physics-based models offer physical interpretability grounded in first principles, yet their reliance on detailed characterisation of inherent optical properties limits scalability and operational deployment [30]. Learning and deep learning frameworks consistently deliver the strongest reported predictive performance, particularly when multi-sensor observations and contextual drivers are incorporated [135,137]; however, their limited physical transparency and inconsistent reporting of predictive uncertainty continue to constrain trust and adoption in policy and regulatory contexts [96,127].

Sensitivity to Model-Selection Strategy

To evaluate the sensitivity of the quantitative synthesis to the choice of within-study model selection, we conducted an alternative aggregation in which the median reported performance per study was retained when multiple model configurations were available.

This sensitivity analysis systematically reduced absolute performance across all model families, confirming selection-related inflation in the primary synthesis. Importantly, however, the relative ranking of model families and the cross-parameter patterns observed in Figure 13 remained qualitatively consistent. These results indicate that while absolute R^2 values are sensitive to model-selection strategy, the comparative insights and structural conclusions drawn from the distribution-aware synthesis are robust to reasonable alternative aggregation rules. Table 17 synthesises these comparative insights by summarising the reported ranges of R^2 and RMSE across empirical, physics-based, and learning-based approaches. Together, the synthesis highlights domains of methodological maturity (Chl-a, TSM, turbidity) alongside parameters for which fundamental challenges remain unresolved (TP, TN, CDOM), underscoring the need for transferable, physically informed, and uncertainty-aware modelling strategies.

Table 17. Comparative synthesis of reported model performance for key water quality parameters. The table summarises study-level ranges of reported coefficients of determination (R^2) and root-mean-square error (RMSE) values across empirical, physics-based, and learning-based approaches. Performance ranges reflect one representative performance estimate per study, following the dependency-control and model-selection rules described in Section 4.2. R^2 is used as the primary comparative metric, while RMSE values are reported where available to provide complementary context. The synthesis highlights systematic differences in accuracy, transferability, and robustness between optically active and non-optical water quality variables.

Parameter	Illustrative Studies	Dominant Model Types	Reported R^2 Range	Reported RMSE Range	RMSE Units	Comparative Insight
Chlorophyll-a (Chl-a)	[64,67]	CNNs; global DL frameworks	0.90–0.95	2–5	$\mu\text{g L}^{-1}$	Highest reported accuracy across sensors; global CNNs show improved generalisation, though shallow and optically complex waters remain challenging.
Total Suspended Matter (TSM)	[41,68,75]	Physics-based RTMs; quantised CNNs; RF/XGB with drivers	0.83–0.92	3–10	mg L^{-1}	Physics-based models remain robust; quantisation-aware DL enables edge-ready deployments; regional performance is enhanced by land-use and climate drivers.
Turbidity	[45,61]	Empirical ratios; ML	0.88–0.93	3–8	NTU	Empirical methods perform strongly in local settings; ML approaches improve cross-site generalisation.
Total Phosphorus (TP)	[62,78]	RF; Extra Trees; XGB + SHAP	0.70–0.80	0.04–0.08	mg L^{-1}	Retrieval skill remains variable; contextual drivers (e.g., land use, precipitation) are critical for improved performance.
Total Nitrogen (TN)	[42]	RF; ML	0.50–0.92	—	—	Strong seasonal dependence; high performance during wet periods (R^2 up to 0.92) but substantial degradation under dry conditions.
Colored Dissolved Organic Matter (CDOM)	[91,121]	Physics-based; ML	0.80–0.85	0.05–0.10	m^{-1}	Moderate-to-high accuracy; transferability limited by covariance with Chl-a and TSM and sensitivity to atmospheric correction.
Dissolved Oxygen/Secchi Depth	[126]	RF with AC products	0.30–0.40	—	—	Weak predictive skill; highlights challenges related to atmospheric correction and shallow-water effects.
Long-term trends (multi-product)	[48]	Multi-product AC intercomparison	≤ 0.88 (565 nm)	—	Reflectance	Reliable interannual trends observed; absolute performance and bias depend strongly on the atmospheric correction scheme.

4.3. Policy and Scientific Implications

The evidence synthesised in this review indicates that satellite-based water quality monitoring has reached a critical inflexion point: methodological sophistication has advanced rapidly, yet structural barriers continue to impede systematic operationalisation. Figure 14 conceptualises this transition pathway by linking prevailing research paradigms to systemic limitations and, subsequently, to the enabling conditions required for deployment within decision-grade monitoring systems. The left panel reflects the dominant research configuration identified throughout this synthesis: sensor-specific algorithms, geographically localised calibration, and models frequently evaluated under non-independent validation. The central panel summarises the structural constraints repeatedly observed across Sections 3 and 4.2: limited transferability across optical regimes and regions, weak interpretability of complex ML/DL architectures, insufficient and non-standardised uncertainty quantification, and fragmented multi-sensor processing infrastructure. The right panel outlines the necessary transition conditions: cross-sensor-transferable models, interpretable and uncertainty-aware learning frameworks, and harmonised, automated, cloud-based pipelines capable of sustained operational delivery.

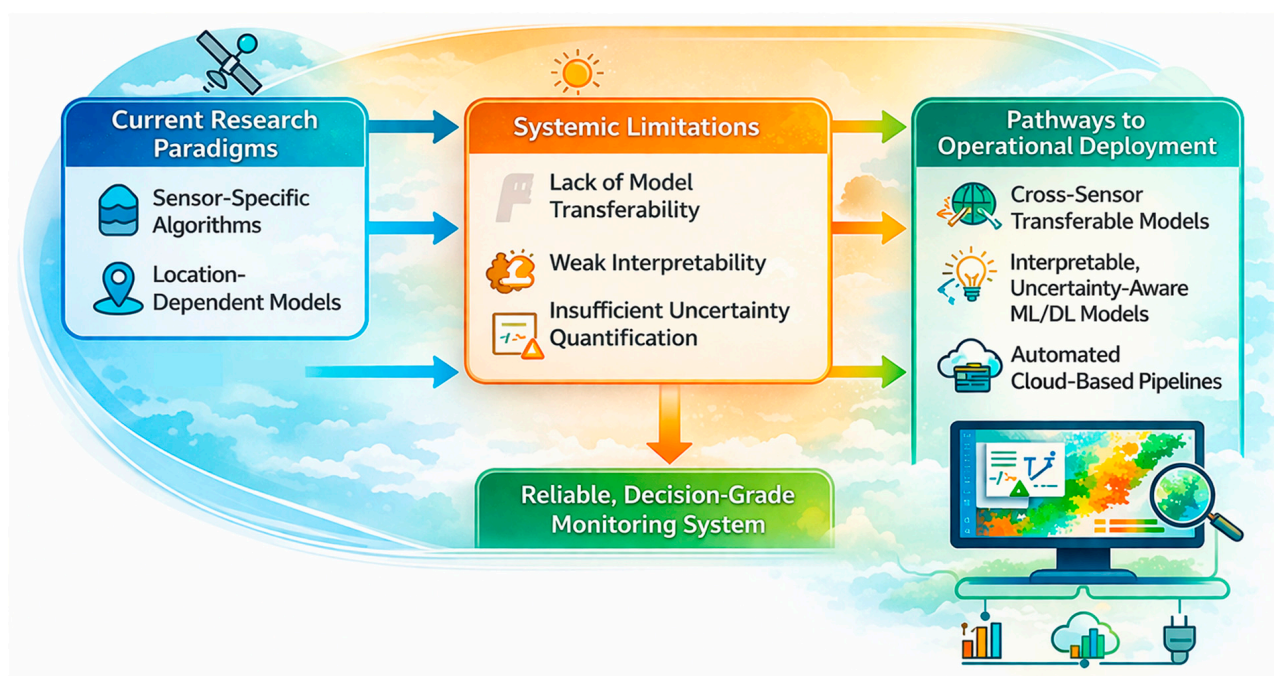


Figure 14. The conceptual framework illustrates the transition from research-oriented to operational, decision-grade, satellite-based water quality monitoring systems. The diagram links current sensor-specific and location-dependent approaches to key limitations, limited transferability, weak interpretability, and insufficient uncertainty quantification and outlines pathways toward deployment through transferable models, interpretable and uncertainty-aware ML/DL frameworks, and automated cloud-based processing pipelines. It emphasises that operationalisation requires coordinated advances across methodological, infrastructural, and governance dimensions. The visual layout was developed with the assistance of a generative artificial intelligence tool for graphical design; all scientific content and interpretations were defined and validated by the authors, with AI use limited strictly to visualisation support.

From a policy perspective, the implications are substantial. Satellite-derived water quality products are increasingly positioned to support regulatory compliance, basin-scale nutrient management, and early-warning systems for bloom events and sediment pulses [145,146]. However, the distribution-based performance synthesis presented in Section 4.2 demonstrates that reported predictive accuracy alone is insufficient to guarantee

regulatory credibility. Transferability across heterogeneous optical regimes and transparent communication of uncertainty emerge as preconditions for policy integration. As shown in Figure 14, operationalisation therefore requires coordinated progress not only in model performance, but in validation rigour, preprocessing transparency, and infrastructure interoperability. The documented sensitivity of retrieval outcomes to atmospheric correction schemes [48,126] further reinforces that preprocessing standardisation is not a peripheral technical detail but a foundational requirement for regulatory trust.

Beyond governance implications, synthesis also redefines priorities for research design. Table 18 consolidates the primary scientific directions emerging from this review, translating systemic barriers into structured research imperatives. Geographically distributed validation and cross-regime benchmarking are essential for mitigating calibration-domain bias and improving transferability [38,78]. Multi-source data integration linking satellite observations within situ networks, IoT sensors, hydrological models, and meteorological drivers consistently enhances contextual robustness and explanatory interpretability [38,147]. Equally critical is the institutionalisation of uncertainty-aware modelling practices. While Bayesian, ensemble, and probabilistic deep-learning approaches demonstrate technical feasibility [38,68,119], the dominance of accuracy-only reporting identified in earlier sections indicates that uncertainty discipline remains unevenly operationalised. Emerging sensor modalities, including hyperspectral missions such as SDGSAT-1 and exploratory microwave applications [93,148], expand observational capacity but simultaneously intensify the need for rigorous cross-sensor harmonisation. Without standardised interoperability frameworks, increasing sensor diversity risks amplifying fragmentation rather than strengthening monitoring capability. Consequently, the pathway from research to operations must be conceptualised as a system-level transformation that integrates methodological innovation, computational infrastructure, validation harmonisation, and institutional alignment. Taken together, Figure 14 and Table 18 articulate a coherent roadmap: achieving reliable, decision-grade water quality monitoring requires embedding transferability, interpretability, and uncertainty quantification directly within model development, evaluation, and reporting practices. The transition is therefore neither purely technical nor purely administrative; it is a coordinated evolution across scientific design, computational architecture, and governance integration. Only under such an integrated paradigm can satellite remote sensing fulfil its potential as a scalable, trustworthy component of global water quality management systems.

Table 18. High-level synthesis of policy-relevant and scientific implications for advancing satellite-based water quality monitoring, highlighting priority directions in operational integration, multi-source data fusion, uncertainty-aware modelling, transferability, and emerging sensor development.

Implication	Rationale/Key Finding	Supporting References
Operational Policy Integration	Satellite-derived products can be embedded in regulatory frameworks and near-real-time management systems; they require QA/QC and AC transparency	[19,48,145,147]
Crisis Management	Rapid detection of algal blooms, floods, and pollution events supports emergency response and mitigation	[14,146,149]
Long-Term Trend Analysis	Multi-decadal archives (Landsat, Sentinel, GCOM-C) enable robust detection of interannual trends and drivers of water quality change	[44,48,137]
Scientific Direction: Data Integration	Multi-source integration (satellite + IoT + hydrology + meteorology) improves robustness and interpretability	[75,133,138]
Scientific Direction: Uncertainty	UQ must be standardised, with probabilistic outputs for risk-sensitive decisions	[38,62,68,119]
Scientific Direction: Transferability	Models must be trained on diverse datasets; global CNNs and transfer learning approaches offer pathways	[38,42,67,78]
Emerging Sensors	Exploration of novel modalities (e.g., microwave, SDGSAT-1 hyperspectral) is a frontier for next-gen monitoring	[47,148]

5. A Proposed Transferable Framework

The synthesis of the reviewed literature highlights a clear need to move beyond prevailing retrieval paradigms toward next-generation frameworks that explicitly address four persistent and interrelated limitations: restricted transferability across sensors and regions, limited physical interpretability, insufficient treatment of predictive uncertainty, and fragmented multi-sensor integration and infrastructure [36,45]. Building directly on the methodological and empirical gaps identified through the quantitative and qualitative analyses presented in Sections 3 and 4, this section proposes an integrated framework that synthesises existing advances while delineating directions that remain conceptual or exploratory. To avoid overinterpretation, it is important to distinguish clearly between framework components that have been demonstrated in the literature and those that currently represent aspirational research directions. Accordingly, the elements described below are explicitly contextualised based on their current levels of empirical validation and operational maturity, ranging from prototype-scale demonstrations to conceptual extensions that require further investigation.

As illustrated in Figure 15, the proposed framework adopts an end-to-end, modular design linking multi-source data ingestion, physics-informed learning, and uncertainty-aware outputs. Data harmonisation and integration demonstrated in numerous regional and multi-sensor studies constitute a mature foundational layer, enabling the alignment of harmonised satellite observations (e.g., Sentinel-2, Landsat-8/9, MODIS, and complementary commercial platforms) with in situ measurements and emerging IoT sensor streams. While such harmonisation strategies are increasingly well established at regional and continental scales, their seamless extension to fully global, continuously updated infrastructures remains an active area of development. At the core of the framework, physics-informed neural networks (PINNs) demonstrated at prototype and regional scales in recent studies provide a principled mechanism for embedding physical consistency directly within data-driven learning processes. By constraining model behaviour through physical laws or process-based regularisation, these approaches offer a pathway toward improved generalisation beyond purely empirical formulations. However, their deployment in large-scale, operational water quality monitoring systems remains limited and heterogeneous across parameters and sensing platforms. An explicit uncertainty quantification module is a central yet variably mature component of the proposed framework. Bayesian and ensemble-based uncertainty-aware formulations have been demonstrated in controlled experiments and selected regional applications, enabling partial propagation of input and model uncertainty to predicted water quality products. Nevertheless, comprehensive uncertainty propagation across multi-sensor, multi-parameter workflows remains inconsistently implemented and largely absent from operational monitoring systems, underscoring the need for further methodological consolidation. The workflow culminates in spatially explicit water quality products accompanied by corresponding uncertainty layers. While uncertainty-aware outputs have been demonstrated in research and pilot contexts, their routine integration into decision-grade monitoring and management workflows remains predominantly aspirational. Consequently, rather than presenting an immediately deployable solution, the proposed framework outlines a structured pathway for progressively translating research-oriented retrieval models into physically consistent, transferable, and uncertainty-aware monitoring systems, directly addressing the limitations identified in the preceding sections.

5.1. Transferable Architecture

A persistent challenge in satellite-based water quality monitoring is the limited generalizability of retrieval models across diverse aquatic systems, optical regimes, and observational contexts. To address this limitation, the proposed framework adopts a globally

trained, multi-sensor learning architecture that harmonises observations from platforms such as Sentinel-2, Landsat-8/9, MODIS, and Planet Scope [150,151]. By leveraging shared spectral-optical characteristics across sensors and water bodies, this approach aims to capture transferable relationships among reflectance constituents rather than site-specific correlations, thereby reducing the need for repeated local recalibration [152]. Recent advances support the feasibility of such architecture. Globally trained convolutional neural networks have demonstrated robust performance for chlorophyll-a retrieval across heterogeneous inland and coastal waters [67], while quantisation-aware and edge-ready implementations indicate that scalable models can be deployed efficiently across both cloud-based and resource-constrained environments [68]. These developments position multi-sensor learning as a practical foundation for operational-scale monitoring. Importantly, transferable architecture need not rely on a single global model. Given the diversity of optical regimes, sensor characteristics, and target parameters, model performance is inherently context-dependent, and no single modelling approach consistently performs optimally across all conditions. In this context, multi-model fusion and ensemble-based architecture provide a complementary strategy for enhancing robustness and generalisation. Such approaches may include mixture-of-experts frameworks, stacked ensembles, or context-conditioned routing mechanisms that dynamically select or weight empirical, physics-based, and ML/DL models based on input characteristics (e.g., OWT class, sensor type, or environmental conditions). By combining complementary inductive biases and strengths across model families, fusion-based architectures can reduce model-specific errors, improve adaptability across heterogeneous domains, and provide more stable performance under cross-region and cross-sensor deployment scenarios.

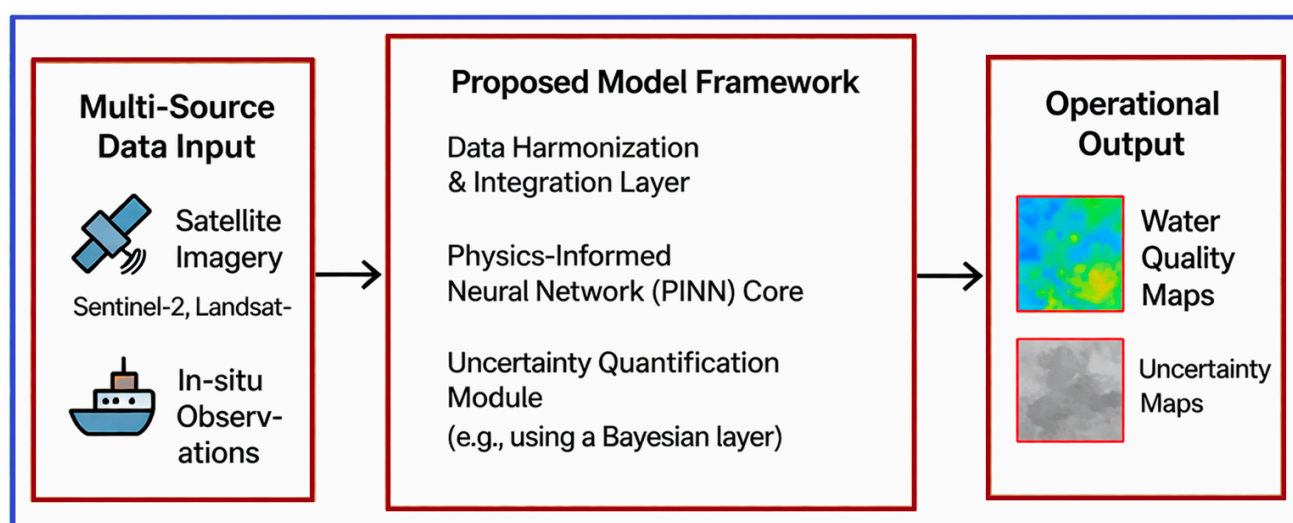


Figure 15. Conceptual diagram of the proposed transferable, physics-informed, and uncertainty-aware framework for satellite-based water quality monitoring. The schematic illustrates the end-to-end workflow from multi-source inputs (satellite and in situ data), through a physics-informed modelling framework incorporating data harmonisation, a PINN-based core, and uncertainty quantification, to operational outputs including water quality and uncertainty maps. Color coding is used for visual grouping of major components (inputs, modelling framework, and outputs) and does not represent quantitative differences or additional data dimensions.

Within this architectural paradigm, Optical Water Type (OWT) information serves as a key conditioning variable to improve transferability and interpretability. OWTs can be incorporated as conditional inputs, stratification layers, or validation axes, enabling regime-aware learning and evaluation. Importantly, the specific form of OWT integration is not

uniform, but depends on the classification paradigm (e.g., spectral clustering, component-dominant, or rule-based), sensor spectral capability, and retrieval model architecture. This context-dependent alignment determines how effectively optical-regime heterogeneity is represented during learning. In this framework, OWT sensor model alignment should be treated as a design variable rather than a post hoc adjustment. For example, spectral clustering-based OWTs are naturally compatible with multisensory ML/DL architecture through conditional learning or stratified training, whereas component-dominant OWTs are more aligned with physics-informed or hybrid models that explicitly leverage bio-optical relationships. Rule-based OWTs, while operationally simple, are better suited to constrained or stable environments and are typically integrated into empirical or semi-analytical workflows. Collectively, these considerations indicate that transferability is not solely a function of model complexity or data volume, but emerges from the coordinated design of OWT representation, sensor characteristics, modelling strategy, and, where appropriate, fusion across complementary model families. Embedding this alignment within the architecture provides a principled pathway toward scalable, transferable, and uncertainty-aware satellite-based water quality monitoring systems.

5.2. Physics-Informed Core

Within the proposed transferable framework, physical consistency is enforced through an explicit physics-informed learning formulation rather than being treated as a purely conceptual or post hoc regularisation. Consistent with the general paradigm of physics-informed neural networks [115], model training is expressed as the minimisation of a composite objective function that balances empirical data fidelity with physically grounded constraints:

$$L = L_{data} + \lambda L_{phys},$$

where L_{data} represents the empirical data-misfit term, typically formulated using mean squared error or related loss functions, quantifying discrepancies between predicted and observed water quality parameters derived from near-synchronous in situ matchups. The second term L_{RT} (also referred to as the physics-consistency loss) constrains the solution space to remain compatible with established radiative-transfer theory and bio-optical relationships governing light–water interactions [116,153]. The weighting parameter λ controls the relative influence of physical constraints and data fidelity and is typically selected via cross-validation or sensitivity analysis to ensure physical consistency while preserving predictive performance. In the context of satellite-based water quality retrieval, a representative formulation of the physics-based constraint can be expressed as a radiative-transfer consistency loss:

$$LRT = || \hat{R}_{rs} - F_{RT}(\hat{p}) ||^2,$$

where $\hat{R}_{rs}$ denotes the predicted remote-sensing reflectance, $\hat{p}$ represents the estimated set of water quality constituents, and F_{RT} is a forward radiative-transfer operator derived from physically based bio-optical models. This formulation reflects the core principle underlying radiative-transfer-based inversion methods, in which spectral residuals between modelled and observed reflectance are minimised to enforce physical plausibility [54,105,153]. Operationally, the radiative-transfer constraint may be implemented in multiple forms depending on computational requirements and deployment objectives. These include (i) differentiable radiative-transfer layers embedded directly within the learning architecture, as commonly adopted in physics-informed and hybrid neural networks [115,116], or (ii) lookup-table-based constraints derived from simplified forward bio-optical models, which penalise departures from physically admissible reflectance constituent relationships while preserving computational efficiency [153,154]. For optically active constituents such as chlorophyll-a

and total suspended matter, the physics-based loss directly constrains predictions through well-established absorption and scattering relationships. In contrast, optically inactive parameters, including total nitrogen and total phosphorus, are constrained indirectly via their learned coupling with optically active proxies and contextual hydrological or biogeochemical drivers. By embedding physical structure directly into the learning objective, the proposed framework suppresses spurious correlations, enhances interpretability, and establishes traceable links between model outputs and governing mechanisms properties increasingly recognised as prerequisites for regulatory acceptance, cross-site transferability, and operational trust in satellite-derived water quality products [37,60].

5.3. Uncertainty-Aware Outputs

Deterministic point estimates inherently limit the suitability of satellite-derived water quality products for operational and regulatory decision-making, as they fail to convey confidence, risk, or conditions under which predictions may degrade. To overcome this limitation, the proposed framework explicitly integrates uncertainty-aware modelling components, including Bayesian learning formulations and Monte Carlo dropout, to generate calibrated, pixel-level predictive intervals alongside mean estimates [119]. This probabilistic treatment enables a principled decomposition of predictive uncertainty into aleatoric uncertainty, driven by noise measurement, atmospheric variability, and environmental heterogeneity, and epistemic uncertainty, associated with model structure, limited or biased training data, and cross-domain or cross-sensor transfer. Importantly, uncertainty-aware outputs should not be interpreted solely as a property of model predictions, but as the outcome of the full observation-to-inference pipeline. In addition to model-derived uncertainty, reliable water quality products should account for upstream sources of uncertainty arising from sensor characteristics and data acquisition, match-up construction, preprocessing and cross-sensor harmonisation, and atmospheric correction (AC). Variability in AC schemes, input data quality, and harmonisation strategies can introduce systematic biases and uncertainty that propagate through the retrieval process, yet are rarely explicitly quantified in current implementations. As such, uncertainty-aware outputs should ideally incorporate, or at a minimum transparently document, these upstream contributions to enable consistent interpretation across sensors, regions, and temporal conditions. Recent empirical evidence highlights the practical relevance of this integrated perspective in dynamic aquatic systems. For instance, strong seasonal sensitivity of total nitrogen retrievals to rainfall regimes demonstrates that uncertainty inflation, rather than point accuracy alone, is often the primary indicator of model reliability under hydrologically variable conditions [42]. Complementary to probabilistic uncertainty quantification, post hoc explainability techniques such as SHAP attribution have been shown to enhance interpretability by revealing the relative influence of contextual drivers, including precipitation and land-use patterns, on nutrient variability [62]. Importantly, such explainability analyses do not constitute uncertainty quantification; rather, they provide diagnostic insight into model behaviour that can be jointly interpreted with uncertainty estimates. Taken together, these advances underscore the necessity of integrated frameworks that combine explicit uncertainty characterisation with transparent model interpretation and pipeline-aware uncertainty documentation to support confidence-aware, decision-grade water quality monitoring.

5.4. Illustrative Prototype/Workflow

To demonstrate how the proposed framework can be instantiated beyond a purely conceptual level, we outline an illustrative prototype workflow representative of realistic cross-sensor and cross-domain deployment scenarios in satellite-based water quality

monitoring. In this workflow, harmonised multispectral satellite observations, specifically Sentinel-2 MSI for model training and Landsat-8/9 OLI for deployment, are ingested alongside sparse in situ measurements and auxiliary hydrometeorological drivers. A dedicated data harmonisation layer enforces cross-sensor radiometric consistency and aligns spatial and temporal resolutions prior to model inference, thereby minimising systematic discrepancies arising from sensor-specific characteristics. The harmonised inputs are subsequently processed by a physics-informed learning core, in which physical constraints derived from radiative transfer theory and inherent optical properties are embedded in the learning objective. These constraints remain fixed during deployment, ensuring physically plausible predictions across diverse optical regimes. Model transferability is explicitly assessed by applying a globally trained configuration calibrated using Sentinel-2 data to water bodies, seasons, or sensor domains not represented in the calibration dataset, thereby testing generalisation beyond site-specific conditions. Uncertainty-aware inference is then performed to propagate input uncertainty, model uncertainty, and observation-related variability through to the final outputs. This enables explicit tracking of uncertainty shifts during cross-sensor transfer, with particular emphasis on epistemic uncertainty inflation due to domain mismatch rather than calibration noise. The resulting products consist of spatially explicit water quality parameter maps accompanied by pixel-level uncertainty estimates, supporting confidence-aware interpretation and downstream decision-making. This illustrative prototype does not constitute a full experimental case study but rather provides a reproducible, implementation-oriented blueprint that demonstrates how multi-sensor data integration, physics-informed modelling, and uncertainty-aware outputs can be coherently combined within a single, transferable pipeline. Depending on latency and operational requirements, practical implementations may leverage cloud-based processing environments or edge-optimised models. Together with the architectural overview in Figure 15 and the component-level synthesis in Table 19, this workflow clarifies how the proposed framework can transition from conceptual design to reproducible and operationally relevant applications.

Table 19. Core components of the proposed transferable, physics-informed, and uncertainty-aware framework for satellite-based water quality monitoring. The table delineates the distinct roles of transferability, physical consistency, probabilistic uncertainty quantification, model interpretability, and multi-source data integration, highlighting how their coordinated implementation supports reproducible and decision-grade operational monitoring.

Component	Rationale & Functionality	Supporting Evidence	Recent Advances/Next Steps
Transferable Architecture	Global multi-sensor training improves cross-site and cross-sensor generalizability while reducing dependence on repeated local recalibration.	[38,132,150,155]	Global CNNs for Chl-a retrieval [67]; edge-ready and quantised CNNs enabling near-real-time deployment [68].
Physics-Informed Core	Embeds radiative-transfer relationships and optical constraints directly into learning objectives to enforce physical consistency and improve interpretability across optical regimes.	[13,33,37,60,105]	Physics-informed neural networks with improved physical transparency under active evaluation in aquatic optics
Uncertainty-Aware Output	Provides calibrated prediction intervals and explicitly separates aleatoric and epistemic uncertainty to support risk-aware interpretation and decision-grade applications.	[38,103,119,137]	Advances in pixel-level uncertainty estimation and calibration; analysis of model sensitivity under seasonal hydrological variability as a complementary context [42].
Model Interpretability (Complementary)	Enhances transparency by identifying contextual and environmental drivers that influence model predictions, without constituting uncertainty quantification.	[62,67]	SHAP-based attribution revealing effects of precipitation and land use on nutrient variability; global benchmarking to delineate domains of model validity
Multi-Source Integration	Combines satellite, UAV, in situ, IoT, and auxiliary drivers (hydrology, meteorology) to improve robustness across spatial and temporal scales.	[133,138,147]	Socio-hydrological data fusion for turbidity dynamics [75]; cross-product atmospheric-correction benchmarking and bias assessment [19,48]

6. Future Research Directions

Building on the distribution-based evidence synthesis and the transferable, physics-informed, and uncertainty-aware framework proposed in Section 5, future research in satellite-based water quality monitoring should prioritise structural robustness rather than incremental algorithmic refinement. The central challenge is no longer merely improving in-domain accuracy, but ensuring transferability across optical regimes and regions, embedding calibrated uncertainty into operational workflows, and aligning methodological innovation with deployable infrastructures. Progress, therefore, depends on coordinated evolution across sensing capabilities, modelling paradigms, and implementation ecosystems.

6.1. Advancing Sensing and Data Ecosystems

Continued improvements in sensor technology will expand observational capacity, but their value depends on interoperability and harmonisation. Hyperspectral satellite missions with enhanced spectral resolution offer improved discrimination of overlapping absorption and scattering features in optically complex waters, strengthening both parameter retrieval and calibration stability [94]. Early evidence from SDGSAT-1 indicates that improved signal-to-noise ratios and spatial resolution can outperform Sentinel-2 under selected inland and coastal conditions, although cross-mission calibration and atmospheric correction consistency remain unresolved constraints [47].

Beyond passive multispectral sensing, integration of complementary modalities represents a strategic direction. Microwave and LiDAR systems address structural limitations of optical sensors under highly turbid, low-illumination, or ice-covered conditions [148,156,157]. At finer scales, UAV-borne hyperspectral observations and IoT-enabled in situ networks reduce temporal aliasing and strengthen calibration in optically heterogeneous systems [22]. Future sensing strategies should therefore emphasise harmonised multi-source acquisition pipelines, in which satellite, airborne, and ground-based observations are interoperable by design rather than reconciled retrospectively.

6.2. Methodological Priorities: Transferability, Uncertainty, and Temporal Robustness

Methodological research must increasingly focus on robustness under spatial and temporal domain shifts. Generalised, physics-informed learning frameworks remain a promising pathway for reducing calibration-domain bias by embedding radiative-transfer constraints and inherent optical structure directly within the learning objective [105,106,158]. However, empirical validation across diverse optical water types and climatic regimes remains limited, underscoring the need for geographically distributed benchmarking and regime-aware validation designs. In addition, future methodological development should explicitly examine multi-model fusion and ensemble learning strategies to improve transferability and robustness. By combining the complementary strengths of empirical, physics-based, and data-driven models, fusion frameworks can enhance physical consistency, nonlinear predictive capability, and adaptability across heterogeneous optical regimes. Approaches such as mixture-of-experts, stacked ensembles, and context-conditioned model routing (e.g., based on OWT or environmental drivers) offer promising pathways to mitigate domain-specific model limitations and improve cross-region generalisation, although systematic evaluation and standardisation of these methods remain limited. Uncertainty quantification should be institutionalised as a core design principle. Pixel-level probabilistic outputs enable risk-aware interpretation in regulatory contexts, particularly where predictive skill varies systematically across seasons or hydrological regimes. Bayesian ensembling and Monte Carlo dropout offer scalable mechanisms for estimating and decomposing epistemic and aleatoric uncertainty within operational pipelines [119,159]. Seasonal sensitivity analyses, such as those demonstrating rainfall-driven variability in total nitrogen

retrieval [42], illustrate that uncertainty inflation often provides earlier diagnostic insight into model degradation than point accuracy alone. Importantly, uncertainty characterisation should extend beyond model predictions to include upstream sources arising from data acquisition, preprocessing, atmospheric correction, and reference-data construction, which collectively influence the reliability and transferability of retrieval outputs. Temporal generalisation further requires explicit evaluation protocols. Multi-year hold-out validation, drift-aware monitoring using residual and uncertainty diagnostics, and structured model updating strategies can mitigate non-stationarity in climate-sensitive aquatic systems. Rather than treating temporal stability as an assumed property of trained models, future work should conceptualise it as an operational capability supported by adaptive evaluation and recalibration mechanisms.

6.3. Implementation and Infrastructure Feasibility

Translating advanced models into operational systems entails non-trivial infrastructural constraints. Physics-informed and probabilistic frameworks typically increase computational demand, requiring careful trade-offs between latency, spatial resolution, and uncertainty characterisation. Cloud-native processing environments offer scalable solutions for ingesting asynchronous, multi-sensor data streams and executing harmonised pipelines in near real time [133,160]. At the same time, edge-optimised inference strategies may support latency-sensitive applications such as bloom detection and event-driven hazard monitoring [68,149]. Cross-mission harmonisation remains a structural prerequisite for reliable long-term monitoring. Persistent inconsistencies in atmospheric correction schemes, radiometric calibration, and product definitions introduce systematic biases across archives, affecting both absolute retrievals and trend analyses [126,161]. Addressing these constraints requires coordinated data standards, transparent preprocessing documentation, and interoperable processing chains that ensure reproducibility across sensors and time. Equally critical is regulatory credibility. Even high-performing models may face limited adoption if their outputs lack transparent uncertainty quantification, clear communication, or traceable physical grounding. Standardised metadata, uncertainty layers, and decision-ready reporting formats aligned with regulatory frameworks are, therefore, central to embedding satellite-derived indicators within compliance and water-resource management systems [41,150,162].

6.4. Toward a Global, Decision-Grade Monitoring Infrastructure

Realising a globally interoperable monitoring infrastructure requires integration beyond individual missions or models. As synthesised in Table 20, priority directions converge around five interdependent domains: advanced sensor integration, generalised physics-informed modelling, robust uncertainty quantification, automated data harmonisation, and science-to-policy alignment. These elements collectively delineate a scalable pathway toward transferable and decision-grade systems.

Long-term credibility additionally depends on cross-site calibration stability, transparent uncertainty reporting, and bias-aware trend detection, as inter-product discrepancies continue to influence interannual assessments [126,162]. Routine delivery of uncertainty-aware products, including prediction intervals and quality flags, remains essential for risk-informed interpretation across use cases. Figure 16 synthesizes these research priorities into a structured roadmap linking current site-specific paradigms to an interoperable, uncertainty-aware global monitoring system. The roadmap integrates technological innovation, methodological discipline, and infrastructural harmonization into a coherent transition pathway.

Table 20. Roadmap summarising priority research directions for advancing satellite-based water quality monitoring toward transferable, uncertainty-aware, and operationally deployable systems, highlighting scientific rationale, supporting evidence, and emerging implementation pathways.

Direction	Rationale & Contribution	Supporting Evidence	Recent Advances/Next Steps
Advanced Sensor Integration	Enables monitoring in complex environments through multi-modal observations (hyperspectral, UAV, IoT, microwave/LiDAR).	[148,149]	SDGSAT-1 high SNR for turbid waters [47]; UAV–IoT fusion for near-real-time validation [22].
Development of Generalised Models	Reduces site-specificity by embedding physical constraints and training across diverse aquatic systems.	[37]	Global CNN generalisation [67]; integration of PINNs under development
Robust Uncertainty Quantification	Delivers decision-grade prediction intervals and separates epistemic and aleatoric uncertainty.	[38,119]	Seasonal TN sensitivity highlighting uncertainty under hydrologic variability [42]; QAT models with embedded uncertainty for near-real-time deployment [68]
Data Harmonisation & Automation	Supports consistent long-term records and near-real-time product delivery.	[38,133]	Cross-product atmospheric-correction benchmarking [48,126]
Science-to-Policy Translation	Aligns satellite products with regulatory frameworks and management needs.	[5,41,150]	Event-driven hazard analytics [149]; integration into compliance monitoring pilots

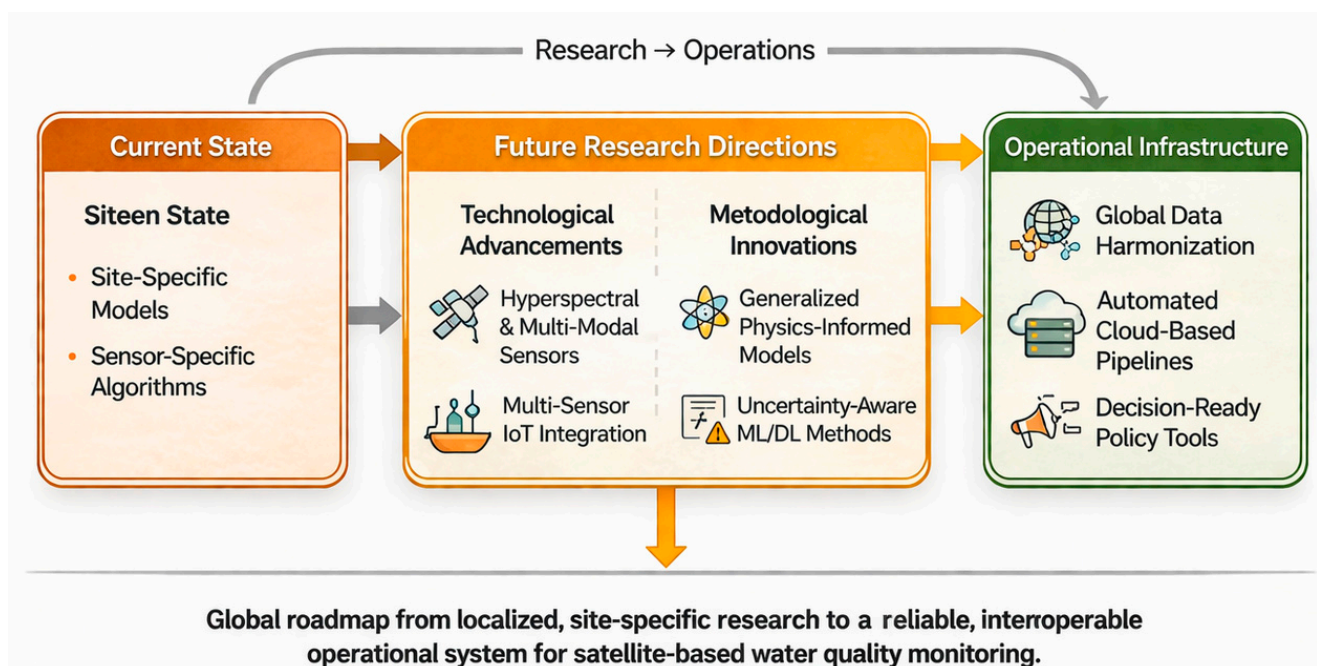


Figure 16. Conceptual roadmap illustrating the transition from site-specific, research-oriented water quality models to scalable and interoperable operational monitoring systems. The framework links the current state of sensor- and site-specific approaches to future research directions, including technological advancements (e.g., hyperspectral and multi-modal sensing, multi-sensor integration) and methodological innovations (e.g., physics-informed and uncertainty-aware ML/DL models), and ultimately to operational infrastructure components such as global data harmonisation and automated cloud-based pipelines. Arrows indicate directional relationships and progression across stages; different arrow colors are used for visual distinction of pathways and do not represent distinct quantitative processes or categories.

7. Conclusions: Toward Decision-Grade, Operational Water Quality Monitoring

This review synthesised four decades of satellite-based water quality research and demonstrated that methodological sophistication has outpaced operational integration. Although reported predictive accuracies for optically active parameters have reached high levels across diverse sensors and modelling paradigms, the evidence shows that performance metrics alone are insufficient to ensure scalable, decision-grade deployment. Through a systematic, distribution-aware synthesis of 152 peer-reviewed studies, this work identified four persistent and interdependent constraints that continue to limit operational adoption: restricted transferability across optical regimes and sensor domains, limited physical interpretability of increasingly complex learning architectures, inconsistent and often absent uncertainty quantification, and fragmented multi-sensor harmonisation and processing infrastructures. The central insight emerging from this synthesis is that operational readiness is a systems-level property rather than a function of isolated algorithmic performance. Robust monitoring depends on the coordinated alignment of three structural pillars: (i) transferable modelling strategies capable of generalising beyond calibration domains; (ii) physically grounded formulations that constrain predictions within radiative-transfer and bio-optical principles; and (iii) explicit, calibrated uncertainty characterisation that communicates reliability across spatial, temporal, and environmental contexts. Without these elements operating in concert, high in-domain accuracy risks masking fragility under cross-region transfers, seasonal non-stationarity, or sensor-domain shifts.

To translate these insights into actionable guidance, several key recommendations for future research and operational implementation are proposed. First, transferability-oriented model design should become a primary objective rather than a secondary evaluation criterion. This includes the development of multi-domain training strategies, context-aware model conditioning (e.g., based on optical regimes or sensor characteristics), and systematic cross-region validation protocols that explicitly test model robustness beyond calibration datasets. Second, physics-informed and hybrid modelling approaches should be further advanced to bridge the gap between interpretability and predictive performance. Future work should prioritise integrating radiative-transfer constraints, spectral-consistency regularisation, and physically meaningful feature representations into machine-learning and deep-learning frameworks to improve generalisation and scientific interpretability. Third, end-to-end uncertainty characterisation must be standardised and expanded beyond model outputs. In addition to probabilistic prediction intervals, future studies should explicitly quantify and propagate uncertainties arising from data acquisition, atmospheric correction, preprocessing, and reference-data construction. Establishing standardised uncertainty-reporting protocols will be critical for ensuring comparability and decision relevance. Fourth, multi-sensor harmonisation and infrastructure standardisation should be treated as a core research priority. This includes the development of consistent atmospheric correction workflows, cross-sensor calibration frameworks, and interoperable data pipelines that integrate satellite, UAV, and in situ observations. Cloud-native processing environments and open, reproducible workflows will be essential for enabling scalable, near-real-time monitoring. Fifth, enhanced treatment of optically inactive parameters is required through the systematic integration of auxiliary data sources (e.g., meteorological, hydrological, and land-use information) and the explicit modelling of proxy relationships. Future studies should move toward mechanism-informed hybrid approaches that improve the robustness and transferability of indirect retrievals. Finally, multi-model fusion and ensemble strategies should be further explored as a pathway to improve robustness under heterogeneous environmental conditions. By combining the complementary strengths of

empirical, physics-based, and data-driven models, fusion approaches can reduce sensitivity to domain shifts and enhance operational reliability.

Building on these recommendations, the proposed transferable, physics-informed, and uncertainty-aware framework reframes satellite water quality retrieval as an integrated pipeline rather than a standalone modelling exercise. By coupling harmonised multi-sensor inputs with physically constrained learning objectives and probabilistic outputs, the framework establishes a structured pathway from research-oriented retrieval models to operational monitoring systems. Importantly, this pathway is progressive rather than binary: components such as global multi-sensor training, physics-informed regularisation, and pixel-level uncertainty estimation are advancing at different levels of empirical maturity, and their coordinated integration represents the next critical frontier. As environmental variability intensifies and regulatory demands for transparency and accountability increase, satellite-based water quality monitoring must evolve from demonstrating predictive capability to delivering operational credibility. Taken together, this review shifts the evaluative lens of the field: from maximising reported accuracy to establishing reproducible, transferable, and uncertainty-disciplined information systems. In doing so, it provides both an evidence-based diagnosis of current limitations and a structured, actionable roadmap for developing globally consistent, physics-informed, and uncertainty-aware monitoring frameworks that support regulation, early warning, and long-term stewardship of inland and coastal waters amid accelerating environmental change.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/rs18071098/s1>, Supplementary Table S1 (Excel file) contains the complete study-level and configuration-level dataset compiled from all 152 peer-reviewed articles included in this systematic review. Supplementary Table S2 documents the coding variables and stratification scheme used for quantitative performance synthesis. Supplementary Table S3 presents the standardised data-extraction protocol and evidence-based coding rules applied consistently across all reviewed studies. Supplementary Protocol S4 (PDF) provides the finalised systematic review protocol, including the search strategy, screening workflow, and quality appraisal framework. Appendix A (included in the main manuscript) presents the open-source implementation roadmap accompanying this review.

Author Contributions: Conceptualisation, S.P.; data curation, S.P.; formal analysis, S.P., V.G., A.R. and L.A.D.; funding acquisition, S.P. and L.A.D.; investigation, S.P.; methodology, S.P.; supervision, V.G., A.R. and L.A.D.; visualisation, S.P., V.G., A.R. and L.A.D.; writing—original draft, S.P.; writing—review and editing, S.P., V.G., A.R. and L.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received support from the Centre of Studies in Geography and Spatial Planning (CEGOT), funded by national funds through the Foundation for Science and Technology (FCT) under the references UIDB/04084/2025 and UIDP/04084/2025. The first author is supported by a PhD grant from FCT under the reference UI/BD/154881/2023.

Institutional Review Board Statement: This research did not involve human participants, animal subjects, or clinical trials. All field studies and data collection followed applicable local regulations and ethical standards. The relevant authorities granted necessary permissions for fieldwork and data acquisition. No sensitive personal data or protected sites were involved in this study.

Data Availability Statement: All data supporting the findings of this study are available within the article and its Supplementary Materials and are additionally archived in a publicly accessible repository. The complete study-level and configuration-level dataset compiled from all 152 peer-reviewed articles included in this systematic review is provided in Supplementary Table S1 (Excel format). This dataset reports, for each extracted model configuration, the satellite or airborne sensor(s) used, target water quality parameters, modelling and retrieval approaches, validation

strategies, and reported performance metrics, thereby enabling full transparency, independent verification, and secondary analysis. Detailed documentation of the coding variables and stratification scheme used for the quantitative performance synthesis is provided in Supplementary Table S2, which includes operational definitions, categorical levels, evidence sources, and the analytical roles of each coding dimension. The standardised data-extraction protocol, unit-of-analysis decisions, aggregation logic, and quality-control procedures applied consistently across all reviewed studies are documented in Supplementary Table S3. The finalised systematic review protocol, including the literature search strategy, screening workflow, inclusion and exclusion criteria, and quality appraisal framework, is provided in Supplementary Protocol S4 (PDF). An archived version of all Supplementary Datasets and review materials has been deposited in Zenodo and is accessible at: <https://doi.org/10.5281/zenodo.18432052> (accessed on 1 January 2025). This archive provides a citable, versioned record of the extracted data and methodological materials, ensuring long-term preservation, reproducibility, and reuse in accordance with open-science best practices.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Open-Source Implementation and Reproducibility

To promote transparency, reproducibility, and long-term community reuse, the transferable, physics-informed, and uncertainty-aware framework synthesised and proposed in this review is accompanied by a publicly accessible, open-source implementation roadmap. The objective of this implementation is not to deliver a finalised, production-ready software system at the present stage, but rather to provide a structured, modular, and reproducible scaffold that faithfully reflects the conceptual architecture articulated throughout the manuscript. The open-source implementation roadmap associated with this review is publicly available at: <https://github.com/omid2red-lgtm/WQ-Physics-UQ-Framework/tree/main> (accessed on 30 March 2026).

The repository is intentionally designed to support extensibility across sensor types, water quality parameters, spatial and temporal scales, and modelling paradigms, and follows a modular organisation aligned with the analytical structure and synthesis presented in this review. In particular, the repository comprises the following core components:

- **Data/:** Modules and utilities for ingesting, preprocessing, harmonising, and managing satellite-based (e.g., multispectral, hyperspectral, and UAV platforms) and in situ observations, including metadata handling, quality control, and cross-sensor alignment.
- **Physics/:** Physics-based components encompassing radiative transfer representations, lookup tables, forward models, and physics-informed constraints applicable to optically active constituents, as well as indirect physical regularisation strategies for non-optically active water quality variables.
- **Models/:** Machine learning and deep learning model architectures, including empirical approaches, physics-informed neural networks (PINNs), and hybrid configurations consistent with the model classes synthesised and discussed in the framework sections of this review.
- **Uncertainty/:** Uncertainty quantification modules supporting probabilistic prediction and uncertainty-aware inference, including Bayesian neural networks, ensemble-based methods, Monte Carlo techniques, and probabilistic output layers.
- **Fusion/:** Multi-source and multi-sensor data fusion components addressing spatial-temporal harmonization, cross-sensor transferability, and domain adaptation across heterogeneous observation platforms.
- **Evaluation/:** Model evaluation utilities for performance assessment, uncertainty calibration, and systematic analysis of transferability across sites, sensors, and environmental conditions.

- Examples/: Illustrative and minimal working workflows demonstrating representative methodological use cases such as cross-sensor transfer, site-to-site generalisation, and uncertainty-aware prediction pipelines intended for conceptual and methodological illustration rather than operational deployment.

At its current stage, the repository primarily serves as a reproducible and extensible scaffold aligned with the conceptual framework proposed in this review, rather than as a fully implemented operational system. No benchmark datasets, standardised training pipelines, or production-ready models are distributed at this stage. Future releases are intended to incorporate reference implementations, benchmark experiments, and fully reproducible case studies developed in follow-up methodological and applied research.

By providing this open-source implementation roadmap, the present review aims to bridge the gap between systematic synthesis and practical reproducibility, facilitating transparent translation of the proposed framework into future empirical studies and, ultimately, operational water quality monitoring workflows.

References

1. O'Reilly, C.M.; Sharma, S.; Gray, D.K.; Hampton, S.E.; Read, J.S.; Rowley, R.J.; Schneider, P.; Lenters, J.D.; McIntyre, P.B.; Kraemer, B.M.; et al. Rapid and highly variable warming of lake surface waters around the globe. *Geophys. Res. Lett.* **2015**, *42*, 10773–10781. [CrossRef]
2. Paerl, H.W.; Huisman, J. Blooms like it hot. *Science* **2008**, *320*, 57–58. [CrossRef]
3. Michalak, A.M.; Anderson, E.J.; Beletsky, D.; Boland, S.; Bosch, N.S.; Bridgeman, T.B.; Chaffin, J.D.; Cho, K.; Confesor, R.; Daloglu, I.; et al. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 6448–6452. [CrossRef]
4. American Public Health Association (APHA); American Water Works Association (AWWA); Water Environment Federation (WEF). *Standard Methods for the Examination of Water and Wastewater*, 23rd ed.; American Public Health Association: Washington, DC, USA, 2017.
5. IOCCG. *Earth Observations in Support of Global Water Quality Monitoring*; Greb, S., Dekker, A.G., Binding, C., Eds.; IOCCG Report Series No. 17; International Ocean-Colour Coordinating Group: Dartmouth, UK, 2018. [CrossRef]
6. Palmer, S.C.J.; Kutser, T.; Hunter, P.D. Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sens. Environ.* **2015**, *157*, 1–8. [CrossRef]
7. Chen, L.; Liu, L.; Liu, S.; Shi, Z.; Shi, C. The application of remote sensing technology in inland water quality monitoring and water environment science: Recent progress and perspectives. *Remote Sens.* **2025**, *17*, 667. [CrossRef]
8. Mohan, S.; Kumar, B.; Nejadhashemi, A.P. Integration of machine learning and remote sensing for water quality monitoring and prediction: A review. *Sustainability* **2025**, *17*, 998. [CrossRef]
9. Adjovu, G.E.; Stephen, H.; James, D.; Ahmad, S. Overview of the application of remote sensing in effective monitoring of water quality parameters. *Remote Sens.* **2023**, *15*, 1938. [CrossRef]
10. Jaywant, S.A.; Arif, K.M. Remote sensing techniques for water quality monitoring: A review. *Sensors* **2024**, *24*, 8041. [CrossRef]
11. Braga, C.Z.F.; Setzer, A.W.; de Lacerda, L.D. Water quality assessment with simultaneous Landsat-5 TM data at Guanabara Bay, Rio de Janeiro, Brazil. *Remote Sens. Environ.* **1993**, *45*, 95–106.
12. Matthews, M.W. A current review of empirical algorithms for remote sensing of chlorophyll-a in inland and coastal waters. *Remote Sens. Environ.* **2011**, *115*, 3250–3263. [CrossRef]
13. Olmanson, L.G.; Bauer, M.E.; Brezonik, P.L. A 20-year Landsat water clarity census of Minnesota's 10,000 lakes. *Remote Sens. Environ.* **2008**, *112*, 4086–4097. [CrossRef]
14. Hou, X.; Feng, L.; Duan, H.; Chen, X.; Sun, D.; Shi, K. Fifteen-year monitoring of the turbidity dynamics in large lakes and reservoirs in the middle and lower basin of the Yangtze River, China. *Remote Sens. Environ.* **2017**, *190*, 107–121. [CrossRef]
15. Binding, C.E.; Greenberg, T.A.; Bukata, R.P. An analysis of MODIS-derived algal and mineral turbidity in Lake Erie. *J. Great Lakes Res.* **2012**, *38*, 107–116. Available online: <https://www.sciencedirect.com/science/article/pii/S0380133011002565> (accessed on 30 March 2026). [CrossRef]
16. Binding, C.E.; Greenberg, T.A.; McCullough, G.; Watson, S.B.; Page, E. An analysis of satellite-derived chlorophyll and algal bloom indices on Lake Winnipeg. *Remote Sens. Environ.* **2018**, *209*, 808–821. [CrossRef]
17. Toming, K.; Kutser, T.; Laas, A.; Sepp, M.; Paavel, B.; Nöges, T. First experiences of Sentinel-2 MSI in monitoring lake water quality. *Remote Sens.* **2016**, *8*, 640. [CrossRef]

18. Buma, W.G.; Lee, S.-I. Evaluation of Sentinel-2 and Landsat 8 images for estimating chlorophyll-a concentrations in Lake Chad, Africa. *Remote Sens.* **2020**, *12*, 2437. [\[CrossRef\]](#)
19. Cheng, K.H.; Chan, S.N.; Lee, J.H.W. Remote sensing of coastal algal blooms using unmanned aerial vehicles (UAVs). *Mar. Pollut. Bull.* **2020**, *152*, 110889. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Cillero Castro, C.; Domínguez Gómez, J.A.; Delgado Martín, J.; Hinojo Sánchez, B.A.; Cereijo Arango, J.L.; Cheda Tuya, F.A.; Díaz-Varela, R. An UAV and satellite multispectral data approach to monitor water quality in small reservoirs. *Remote Sens.* **2020**, *12*, 1514. [\[CrossRef\]](#)
21. Cui, M.; Sun, Y.; Huang, C.; Li, M. Water turbidity retrieval based on UAV hyperspectral remote sensing. *Water* **2022**, *14*, 128. [\[CrossRef\]](#)
22. Wasehun, E.T.; Hashemi Beni, L.; Di Vittorio, C.A. UAV and satellite remote sensing for inland water quality assessments: A literature review. *Environ. Monit. Assess.* **2024**, *196*, 277. [\[CrossRef\]](#)
23. Coffey, M.M.; Schaeffer, B.A.; Salls, W.B.; Minucci, J.M.; Cronin-Golomb, O. Recommendations for temporal aggregation of water quality data from multi-platform satellite constellations. *Int. J. Remote Sens.* **2026**, *47*, 177–199. [\[CrossRef\]](#)
24. Jiang, B.; Fan, D.; Huang, Q.; Li, X.; Ve, N.D.; Ren, F.; Yu, J.; Boss, E. Evaluation of different approaches for assessing water quality using Sentinel-2/MSI: A case study in coastal Ningde. *J. Mar. Sci. Eng.* **2026**, *14*, 267. [\[CrossRef\]](#)
25. Neves, V.H.; Sánchez-Pérez, L.; Antunes, S.C.; Pace, G.; Sòria-Perpinyà, X.; Delegido, J. Development and calibration of Sentinel-2 spectral indices for water quality parameter estimation in Alqueva Reservoir, Southern Portugal. *Remote Sens.* **2026**, *18*, 469. [\[CrossRef\]](#)
26. Basirian, S.; Najafzadeh, M.; Demir, I. Water quality monitoring for coastal hypoxia: Integration of satellite imagery and machine learning models. *Mar. Pollut. Bull.* **2026**, *222*, 118735. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Qin, H.; Fang, C.; Liu, G.; Song, K.; Li, Z.; Li, S.; Tao, H.; Yan, Z. Temperature is a key factor affecting total phosphorus and total nitrogen concentrations in northeastern lakes based on Sentinel-2 images and machine learning methods. *Remote Sens.* **2025**, *17*, 267. [\[CrossRef\]](#)
28. Deng, Y.; Zhang, Y.; Pan, D.; Yang, S.X.; Gharabaghi, B. Review of recent advances in remote sensing and machine learning methods for lake water quality management. *Remote Sens.* **2024**, *16*, 4196. [\[CrossRef\]](#)
29. Pan, D.; Deng, Y.; Yang, S.X.; Gharabaghi, B. Recent advances in remote sensing and artificial intelligence for river water quality forecasting: A review. *Environments* **2025**, *12*, 158. [\[CrossRef\]](#)
30. Gitelson, A.A.; Schalles, J.F.; Hladik, C.M. Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake Bay case study. *Remote Sens. Environ.* **2007**, *109*, 464–472. [\[CrossRef\]](#)
31. Mishra, S.; Mishra, D.R. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sens. Environ.* **2012**, *117*, 394–406. [\[CrossRef\]](#)
32. Odermatt, D.; Gitelson, A.; Brando, V.E.; Schaepman, M. Review of constituent retrieval in optically complex waters from space. *Remote Sens. Environ.* **2012**, *118*, 116–126. [\[CrossRef\]](#)
33. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geosci. Front.* **2016**, *7*, 3–10. [\[CrossRef\]](#)
34. Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [\[CrossRef\]](#)
35. Guo, H.; Tian, S.; Huang, J.; Zhu, X.; Wang, B.; Zhang, Z. Performance of deep learning in mapping water quality of Lake Simcoe with long-term Landsat archive. *ISPRS J. Photogramm. Remote Sens.* **2022**, *183*, 451–469. [\[CrossRef\]](#)
36. Zhang, H.; Xue, B.; Wang, G.; Zhang, X.; Zhang, Q. Deep learning-based water quality retrieval in an impounded lake using Landsat 8 imagery: An application in Dongping Lake. *Remote Sens.* **2022**, *14*, 4505. [\[CrossRef\]](#)
37. He, Y.; Jin, S.; Shang, W. Water quality variability and related factors along the Yangtze River using Landsat-8. *Remote Sens.* **2021**, *13*, 2241. [\[CrossRef\]](#)
38. Pahlevan, N.; Schott, J.R. Characterizing the relative calibration of Landsat-7 (ETM+) visible bands with Terra (MODIS) over clear waters: The implications for monitoring water resources. *Remote Sens. Environ.* **2012**, *125*, 167–180. [\[CrossRef\]](#)
39. Casal, G. Assessment of Sentinel-2 to monitor highly dynamic small water bodies: The case of Louro lagoon (Galicia, NW Spain). *Oceanologia* **2022**, *64*, 88–102. [\[CrossRef\]](#)
40. Potes, M.; Rodrigues, G.; Penha, A.M.; Novais, M.H.; Costa, M.J.; Salgado, R.; Morais, M.M. Use of Sentinel-2 MSI for water quality monitoring at Alqueva Reservoir, Portugal. *Proc. Int. Assoc. Hydrol. Sci.* **2018**, *380*, 73–79. [\[CrossRef\]](#)
41. Xie, Y.; Zhou, Y.; Tao, Z.; Shao, W.; Yang, M. Remote sensing inversion of the total suspended matter concentration in the Nanyi Lake based on Sentinel-3 OLCI imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 10380–10389. [\[CrossRef\]](#)
42. Muhoyi, H.; Gumindoga, W.; Mhizha, A.; Misi, S.N.; Nondo, N. Remote sensing application in complement to in-situ monitoring of water quality: Lower Manyame Sub-catchment, Zimbabwe. *Sci. Afr.* **2025**, *27*, e02551. [\[CrossRef\]](#)

43. Katlane, R.; El Kilani, B.; Dhaoui, O.; Kateb, F.; Chehata, N. Monitoring of sea surface temperature, chlorophyll, and turbidity in Tunisian waters from 2005 to 2020 using MODIS imagery and the Google Earth Engine. *Reg. Stud. Mar. Sci.* **2023**, *66*, 103143. [\[CrossRef\]](#)
44. Xia, K.; Wu, T.; Li, X.; Wang, S.; Shen, Q. A new method for accurate inversion of Forel-Ule index using MODIS images—Revealing the water color evolution in China's large lakes and reservoirs over the past two decades. *Water Res.* **2024**, *255*, 121560. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Ayana, E.K.; Worqlul, A.W.; Steenhuis, T.S. Evaluation of stream water quality data generated from MODIS images in modeling total suspended solid emission to a freshwater lake. *Sci. Total Environ.* **2015**, *523*, 170–177. [\[CrossRef\]](#)
46. Zhang, Y.; Wu, L. Surveillance of urban river environment by quantifying distributions of water quality parameters using hyperspectral remote sensing-based ripple propagation graph network. *Environ. Pollut.* **2025**, *384*, 126875. [\[CrossRef\]](#)
47. Li, P.; Naz, I.; Aslam, R.W.; Liaquat, M.A.; Said, Y. Groundwater quality assessment for rangeland dynamic: Integration of multicriteria decision analysis with remote sensing data. *Rangel. Ecol. Manag.* **2025**, *102*, 110–127. [\[CrossRef\]](#)
48. Salem, S.I.; Toratani, M.; Higa, H.; Son, S.; Siswanto, E.; Ishizaka, J. Long-term evaluation of GCOM-C/SGLI reflectance and water quality products: Variability among JAXA G-Portal and JASMES. *Remote Sens.* **2025**, *17*, 221. [\[CrossRef\]](#)
49. Li, Y.; Fu, Y.; Lang, Z.; Cai, F. A high-frequency and real-time ground remote sensing system for obtaining water quality based on a micro hyper-spectrometer. *Sensors* **2024**, *24*, 1833. [\[CrossRef\]](#)
50. Schaeffer, B.A.; Schaeffer, K.G.; Keith, D.; Lunetta, R.S.; Conmy, R.; Gould, R.W. Barriers to adopting satellite remote sensing for water quality management. *Int. J. Remote Sens.* **2013**, *34*, 7534–7544. [\[CrossRef\]](#)
51. Rahat, S.H.; Steissberg, T.; Chang, W.; Chen, X.; Mandavya, G.; Tracy, J.; Wasti, A.; Atreya, G.; Saki, S.; Bhuiyan, M.A.E.; et al. Remote sensing-enabled machine learning for river water quality modeling under multidimensional uncertainty. *Sci. Total Environ.* **2023**, *898*, 165504. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Bertone, E.; Peters Hughes, S. Probabilistic prediction of satellite-derived water quality for a drinking water reservoir. *Sustainability* **2023**, *15*, 11302. [\[CrossRef\]](#)
53. Uudeberg, K.; Aavaste, A.; Kõks, K.-L.; Ansper, A.; Uusõue, M.; Kangro, K.; Ansko, I.; Ligi, M.; Toming, K.; Reinart, A. Optical water type guided approach to estimate optical water quality parameters. *Remote Sens.* **2020**, *12*, 931. [\[CrossRef\]](#)
54. Sagan, V.; Peterson, K.T.; Maimaitijiang, M.; Sidike, P.; Sloan, J.; Greeling, B.A.; Maalouf, S.; Adams, C. Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Sci. Rev.* **2020**, *205*, 103187. [\[CrossRef\]](#)
55. Vakili, T.; Amanollahi, J. Determination of optically inactive water quality variables using Landsat 8 data: A case study in Geshlagh Reservoir. *J. Clean. Prod.* **2020**, *247*, 119134. [\[CrossRef\]](#)
56. Assunção, A.; Silva, T.F.G.; de Carvalho, L.A.S.; Vinçon-Leite, B. Assessing water quality restoration measures in Lake Pampulha (Brazil) through remote sensing imagery. *Environ. Sci. Pollut. Res.* **2025**, *32*, 3838–3868. [\[CrossRef\]](#)
57. Brivio, P.A.; Giardino, C.; Zilioli, E. Determination of chlorophyll concentration changes in Lake Garda using an image-based radiative transfer code for Landsat TM images. *Int. J. Remote Sens.* **2001**, *22*, 487–502. [\[CrossRef\]](#)
58. Niroumand-Jadidi, M.; Bovolo, F.; Bresciani, M.; Gege, P.; Giardino, C. Water quality retrieval from Landsat-9 (OLI-2) imagery and comparison to Sentinel-2. *Remote Sens.* **2022**, *14*, 4596. [\[CrossRef\]](#)
59. Gu, K.; Zhang, Y.; Qiao, J. Random forest ensemble for river turbidity measurement from space remote sensing data. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9028–9036. [\[CrossRef\]](#)
60. Xiong, J.; Lin, C.; Cao, Z.; Hu, M.; Xue, K.; Chen, X.; Ma, R. Development of remote sensing algorithm for total phosphorus concentration in eutrophic lakes: Conventional or machine learning? *Water Res.* **2022**, *215*, 118213. [\[CrossRef\]](#)
61. Quang, N.H.; Hang, L.T.T.; Karamuz, E.; Nones, M. Modelling sea and brackish water quality of Ha Long City (Vietnam) using machine learning and remote sensing techniques. *Adv. Space Res.* **2025**, *75*, 4575–4587. [\[CrossRef\]](#)
62. Liang, Y.; Ding, F.; Liu, L.; Yin, F.; Hao, M.; Kang, T.; Zhao, C.; Wang, Z.; Jiang, D. Monitoring water quality parameters in urban rivers using multi-source data and machine learning approach. *J. Hydrol.* **2025**, *648*, 132394. [\[CrossRef\]](#)
63. Kandasamy, L.; Mahendran, A.; Sangaraju, S.H.V.; Mathur, P.; Faldu, S.V.; Mazzara, M. Enhanced remote sensing and deep learning aided water quality detection in the Ganges River, India supporting monitoring of aquatic environments. *Results Eng.* **2025**, *25*, 103604. [\[CrossRef\]](#)
64. Meng, H.; Zhang, J.; Zheng, Z.; Song, Y.; Lai, Y. Classification of inland lake water quality levels based on Sentinel-2 images using convolutional neural networks and spatiotemporal variation and driving factors of algal bloom. *Ecol. Inform.* **2024**, *80*, 102549. [\[CrossRef\]](#)
65. Shukla, B.K.; Gupta, L.; Sharma, P.K.; Tyagi, K.; Yadav, H.; Singh, S.; Yadav, Y. Advancements in remote sensing for water quality assessment: A comprehensive exploration. In *Geo-Data Revolution: Advances in Spatial Analysis and Natural Hazard Mapping*; Pathak, S., Shukla, A.K., Sharma, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2025. [\[CrossRef\]](#)

66. Shu, P.; Aslam, R.W.; Naz, I.; Ghaffar, B.; Kucher, D.E.; Qudoods, A.; Raza, D.; Abdullah-Al-Wadud, M.; Zulqarnain, R.M. Deep learning-based super-resolution of remote sensing images for enhanced groundwater quality assessment and environmental monitoring in urban areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 7933–7949. [CrossRef]
67. Giménez, J.G.; González, M.; Martínez-España, R.; Cecilia, J.M.; López-Espín, J.J. Enhancing shallow water quality monitoring efficiency with deep learning and remote sensing: A case study in Mar Menor. *J. Ambient. Intell. Smart Environ.* **2025**, *17*, 182–197. [CrossRef]
68. Moon, J.; Jung, S.; Suh, S.; Pyo, J. Development of deep learning quantization framework for remote sensing edge device to estimate inland water quality in South Korea. *Water Res.* **2025**, *283*, 123760. [CrossRef] [PubMed]
69. Niroumand-Jadidi, M.; Gege, P. WASI-AI: Synergistic integration of AI and physics for retrieving water quality and benthic parameters from multi- and hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 22832–22846. [CrossRef]
70. Chen, W.; Nguyen, T.T.N.; Pein, J.; Jourdin, F.; Fablet, R.; Staneva, J. Physics-informed neural data assimilation for high-resolution coastal suspended particulate matter reconstruction. *Appl. Ocean. Res.* **2025**, *165*, 10487. [CrossRef]
71. Nikoo, M.R.; Zamani, M.G.; Zadeh, M.M.; Al-Rawas, G.; Al-Wardy, M.; Gandomi, A. Mapping reservoir water quality using Bayesian maximum entropy with Sentinel-2. *Sci. Rep.* **2024**, *14*, 16438. Available online: <https://www.nature.com/articles/s41598-024-66699-2> (accessed on 30 March 2026). [CrossRef]
72. Yang, H.; Du, Y.; Zhao, H.; Chen, F. Water Quality Chl-a Inversion Based on Spatio-Temporal Fusion and Convolutional Neural Network. *Remote Sens.* **2022**, *14*, 1267. [CrossRef]
73. Wei, H.; Jia, K.; Wang, Q.; Cao, B.; Qi, J.; Zhao, W.; Yan, K. A remote sensing index for the detection of multi-type water quality anomalies in complex geographical environments. *Int. J. Digit. Earth* **2024**, *17*, 2313695. [CrossRef]
74. Qiao, H.; Lee, Z.; Wang, D.; Zheng, Z.; Ye, X.; Dou, C. One-step retrieval of water quality parameters from satellite top-of-atmosphere measurements. *Remote Sens. Environ.* **2025**, *323*, 114709. [CrossRef]
75. Steinbach, S.; Bartels, A.; Rienow, A.; Kuria, B.T.; Zwart, S.J.; Nelson, A. Predicting turbidity dynamics in small reservoirs in central Kenya using remote sensing and machine learning. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *136*, 104390. [CrossRef]
76. IOCCG. *Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications*; Reports of the International Ocean-Colour Coordinating Group (IOCCG); IOCCG: Dartmouth, NS, Canada, 2006. Available online: <http://www.ioccg.org/reports/report5.pdf> (accessed on 30 March 2026).
77. Chen, X.; Liu, L.; Zhang, X.; Li, J.; Wang, S.; Liu, D.; Duan, H.; Song, K. An assessment of water color for inland water in China using a Landsat-8-derived Forel–Ule Index and the Google Earth Engine platform. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7770–7784. [CrossRef]
78. Khan, R.M.; Salehi, B.; Niroumand-Jadidi, M.; Mahdianpari, M. Global vs. local random forest model for water quality monitoring: Assessment in Finger Lakes using Sentinel-2 imagery and GLORIA dataset. In *Proceedings of the IGARSS 2024—IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024*; IEEE: New York, NY, USA, 2024; pp. 4389–4392.
79. Trevisiol, F.; Mandanici, E.; Pagliarani, A.; Bitelli, G. Evaluation of Landsat-9 interoperability with Sentinel-2 and Landsat-8 over Europe and local comparison with field surveys. *ISPRS J. Photogramm. Remote Sens.* **2024**, *210*, 55–68. [CrossRef]
80. Gonzalez-Marquez, L.C.; Torres-Bejarano, F.M.; Torregroza-Espinosa, A.C.; Hansen-Rodríguez, I.R.; Rodríguez-Gallegos, H.B. Use of Landsat 8 images for depth and water quality assessment of El Guájaro Reservoir, Colombia. *J. South Am. Earth Sci.* **2018**, *82*, 231–238. [CrossRef]
81. Maimouni, S.; Moufkari, A.A.; Daghor, L.; Fekri, A.; Oubraim, S.; Lhissou, R. Spatiotemporal monitoring of low water turbidity in Moroccan coastal lagoon using Sentinel-2 data. *Remote Sens. Appl. Soc. Environ.* **2022**, *26*, 100772. [CrossRef]
82. Sudduth, K.A.; Jang, G.-S.; Lerch, R.N.; Sadler, E.J. Long-term agroecosystem research in the Central Mississippi River Basin: Hyperspectral remote sensing of reservoir water quality. *J. Environ. Qual.* **2015**, *44*, 71–83. [CrossRef]
83. Chu, H.-J.; He, Y.-C. Remote sensing water quality inversion using sparse representation: Chlorophyll-a retrieval from Sentinel-2 MSI data. *Remote Sens. Appl. Soc. Environ.* **2023**, *31*, 101006. [CrossRef]
84. Kutser, T. Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. *Int. J. Remote Sens.* **2009**, *30*, 4401–4425. [CrossRef]
85. Kutser, T. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. *Limnol. Oceanogr.* **2004**, *49*, 2179–2189. [CrossRef]
86. Hajigholizadeh, M.; Moncada, A.; Kent, S.; Melesse, A.M. Land–lake linkage and remote sensing application in water quality monitoring in Lake Okeechobee, Florida, USA. *Land* **2021**, *10*, 147. [CrossRef]
87. Singh, A.; Jakubowski, A.R.; Chidister, I.; Townsend, P.A. A MODIS approach to predicting stream water quality in Wisconsin. *Remote Sens. Environ.* **2013**, *128*, 74–86. [CrossRef]
88. Wynne, T.T.; Stumpf, R.P.; Tomlinson, M.C.; Warner, R.A.; Tester, P.A.; Dyble, J.; Fahnenstiel, G.L. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. *Int. J. Remote Sens.* **2008**, *29*, 3665–3672. [CrossRef]
89. Poddar, S.; Chacko, N.; Swain, D. Estimation of chlorophyll-a in northern coastal Bay of Bengal using Landsat-8 OLI and Sentinel-2 MSI sensors. *Front. Mar. Sci.* **2019**, *6*, 598. [CrossRef]

90. Sanjoto, T.B.; Elwafa, A.H.; Tjahjono, H.; Sidiq, W.A.B.N. Study of total suspended solid concentration based on Doxaran algorithm using Landsat-8 image in coastal water between Bodri River estuary up to east flood canal Semarang City. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *561*, 012053. [\[CrossRef\]](#)
91. Zhao, D.; Huang, J.; Li, Z.; Yu, G.; Shen, H. Dynamic monitoring and analysis of chlorophyll-a concentrations in global lakes using Sentinel-2 images in Google Earth Engine. *Sci. Total Environ.* **2024**, *912*, 169152. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Arıman, S. Determination of inactive water quality variables by MODIS data: A case study in the Kızılırmak Delta–Balık Lake, Turkey. *Estuar. Coast. Shelf Sci.* **2021**, *260*, 107505. [\[CrossRef\]](#)
93. Li, W.; Tang, S.; Tian, L.; Zhao, H.; Ye, H.; Zheng, W.; Liu, Y.; Sun, L. Assessing on-orbit radiometric performance of SDGSAT-1 MII for turbid water remote sensing. *Remote Sens. Environ.* **2025**, *321*, 114683. [\[CrossRef\]](#)
94. Flores-Anderson, A.I.; Griffin, R.; Dix, M.; Romero-Oliva, C.S.; Ochaeta, G.; Skinner-Alvarado, J.; Ramirez Moran, M.V.; Hernandez, B.; Cherrington, E.; Page, B.; et al. Hyperspectral satellite remote sensing of water quality in Lake Atitlán, Guatemala. *Front. Environ. Sci.* **2020**, *8*, 7. [\[CrossRef\]](#)
95. Kim, Y.H.; Son, S.; Kim, H.C.; Kim, B.; Park, Y.G.; Nam, J.; Ryu, J. Application of satellite remote sensing in monitoring dissolved oxygen variabilities: A case study for coastal waters in Korea. *Environ. Int.* **2020**, *134*, 105301. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Ness, E.; Fatima, A.; Maktabdar-Oghaz, M.; Luca, C. An investigation into water quality monitoring models using remote sensing. *Int. J. Remote Sens.* **2025**, *46*, 1742–1772. [\[CrossRef\]](#)
97. Kim, H.-C.; Son, S.; Kim, Y.-H.; Khim, J.-S.; Nam, J.; Chang, W.-K.; Lee, J.-H.; Lee, C.-H.; Ryu, J. Remote sensing and water quality indicators in the Korean West Coast: Spatio-temporal structures of MODIS-derived chlorophyll-a and total suspended solids. *Mar. Pollut. Bull.* **2017**, *121*, 425–434. [\[CrossRef\]](#)
98. Karaoui, I.; Arioua, A.; Boudhar, A.; Hssaisoune, M.; El Mouatassime, S.; Ait Ouhamchich, K.; Elhamdouni, D.; Idrissi, A.E.A.; Nouaim, W. Evaluating the potential of Sentinel-2 satellite images for water quality characterization of artificial reservoirs: The Bin El Ouidane Reservoir case study (Morocco). *Meteorol. Hydrol. Water Manag.* **2019**, *7*, 31–39. [\[CrossRef\]](#)
99. Dekker, A.G.; Vos, R.J.; Peters, S.W.M. Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data. *Int. J. Remote Sens.* **2002**, *23*, 15–35. [\[CrossRef\]](#)
100. Feng, L.; Hu, C.; Li, J. Can MODIS land reflectance products be used for estuarine and inland waters? *Water Resour. Res.* **2018**, *54*, 3583–3601. [\[CrossRef\]](#)
101. Joshi, I.; D'Sa, E.J. Seasonal variation of colored dissolved organic matter in Barataria Bay, Louisiana, using combined Landsat and field data. *Remote Sens.* **2015**, *7*, 12478–12502. [\[CrossRef\]](#)
102. Viridis, S.G.P.; Xue, W.; Winijkul, E.; Nitivattananon, V.; Punpukdee, P. Remote sensing of tropical riverine water quality using Sentinel-2 MSI and field observations. *Ecol. Indic.* **2022**, *144*, 109472. [\[CrossRef\]](#)
103. Zhou, X.; Liu, X.; Wang, X.; He, G.; Zhang, Y.; Wang, G.; Zhang, Z. Evaluation of surface reflectance products based on optimized 6S model using synchronous in situ measurements. *Remote Sens.* **2022**, *14*, 83. [\[CrossRef\]](#)
104. Svircevic, Z.; Simeunovic, J.; Subakov-Simic, G.; Krstic, S.; Pantelic, D.; Dulic, T. Cyanobacterial blooms and their toxicity in Vojvodina lakes, Serbia. *Int. J. Environ. Res.* **2013**, *7*, 745–758. [\[CrossRef\]](#)
105. Sun, D.; Su, X.; Qiu, Z.; Wang, S.; Mao, Z.; He, Y. Remote sensing estimation of sea surface salinity from GOCI measurements in the southern Yellow Sea. *Remote Sens.* **2019**, *11*, 775. [\[CrossRef\]](#)
106. Hu, C. A novel ocean color index to detect floating algae in the global oceans. *Remote Sens. Environ.* **2009**, *113*, 2118–2129. [\[CrossRef\]](#)
107. Song, K.; Wang, Z.; Blackwell, J.; Zhang, B.; Li, F.; Zhang, Y.; Jiang, G. Water quality monitoring using Landsat Thematic Mapper data with empirical algorithms in Chagan Lake, China. *J. Appl. Remote Sens.* **2011**, *5*, 053506. [\[CrossRef\]](#)
108. Gitelson, A.A.; Dall'Olmo, G.; Moses, W.; Rundquist, D.C.; Barrow, T.; Fisher, T.R. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sens. Environ.* **2008**, *112*, 3582–3593. [\[CrossRef\]](#)
109. Moses, W.J.; Gitelson, A.A.; Berdnikov, S.; Povazhnyy, V. Satellite-based estimation of chlorophyll-a concentration using the red and NIR bands: Application to lakes of Russia and the USA. *Remote Sens. Environ.* **2012**, *122*, 118–129. [\[CrossRef\]](#)
110. Dall'Olmo, G.; Gitelson, A.A. Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: Experimental results. *Appl. Opt.* **2005**, *44*, 412–422. [\[CrossRef\]](#)
111. Doerffer, R.; Schiller, H. The MERIS Case 2 Water Algorithm. *Int. J. Remote Sens.* **2007**, *28*, 517–535. [\[CrossRef\]](#)
112. Shamloo, A.; Sima, S. Investigating the potential of remote sensing-based machine-learning algorithms to model Secchi-disk depth, total phosphorus, and chlorophyll-a in Lake Urmia. *J. Great Lakes Res.* **2024**, *50*, 102370. [\[CrossRef\]](#)
113. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [\[CrossRef\]](#)
114. Yang, H.; Kong, J.; Hu, H.; Du, Y.; Gao, M.; Chen, F. A review of remote sensing for water quality retrieval: Progress and challenges. *Remote Sens.* **2022**, *14*, 1770. [\[CrossRef\]](#)
115. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [\[CrossRef\]](#)

116. Daw, A.; Karpatne, A.; Watkins, W.; Read, J.; Kumar, V. Physics-guided Neural Networks (PGNN). *arXiv* **2017**, arXiv:1710.11431.
117. Cao, L.; Zhen, Z.; Chen, S.; Yin, T. A physically based differentiable radiative transfer model (DRTM) for land surface optical and biochemical parameters retrieval. *Remote Sens. Environ.* **2025**, *325*, 114764. [[CrossRef](#)]
118. Pulliainen, J.; Kallio, K.; Eloheimo, K.; Koponen, S.; Servomaa, H.; Hannonen, T.; Tauriainen, S.; Hallikainen, M. A semi-operative approach to lake water quality retrieval from remote sensing data. *Sci. Total Environ.* **2001**, *268*, 79–93. [[CrossRef](#)]
119. Ndou, N. Geostatistical inference of Sentinel-2 spectral reflectance patterns to water quality indicators in the Setumo Dam, South Africa. *Remote Sens. Appl. Soc. Environ.* **2023**, *30*, 100945. [[CrossRef](#)]
120. Salmaso, N.; Mosello, R. Limnological research in the deep southern subalpine lakes: Synthesis, directions and perspectives. *Adv. Oceanogr. Limnol.* **2010**, *1*, 29–66. [[CrossRef](#)]
121. Cao, Z.; Ma, R.; Duan, H.; Pahlevan, N.; Melack, J.; Shen, M.; Xue, K.; Ma, M. A machine learning approach to estimate chlorophyll-a in highly turbid inland waters of China using Sentinel-2 MSI imagery. *Remote Sens. Environ.* **2022**, *269*, 112799. [[CrossRef](#)]
122. Downing, J.A.; Prairie, Y.T.; Cole, J.J.; Duarte, C.M.; Tranvik, L.J.; Striegl, R.G.; McDowell, W.H.; Kortelainen, P.; Caraco, N.F.; Melack, J.M.; et al. The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol. Oceanogr.* **2006**, *51*, 2388–2397. [[CrossRef](#)]
123. Markogianni, V.; Kalivas, D.; Petropoulos, G.P.; Dimitriou, E. An appraisal of the potential of Landsat-8 in estimating chlorophyll-a, ammonium concentrations and other water quality indicators. *Remote Sens.* **2018**, *10*, 1018. [[CrossRef](#)]
124. De Vlaming, V.; DiGiorgio, C.; Fong, S.; Deanovic, L.A.; Carpio-Obeso, M.D.L.P.; Bailey (Miller), J.L.; Miller, M.A.; Richard, N. Irrigation runoff insecticide pollution of rivers in the Imperial Valley, California (USA). *Environ. Pollut.* **2004**, *132*, 213–229. [[CrossRef](#)]
125. Dornhofer, K.; Oppelt, N. Remote sensing for lake research and monitoring Recent advances. *Ecol. Indic.* **2016**, *64*, 105–122. [[CrossRef](#)]
126. Synan, H.E.; Howes, B.L.; Sampieri, S.; Lohrenz, S.E. Water quality monitoring using Landsat 8 OLI in Pleasant Bay, Massachusetts, USA. *Remote Sens.* **2025**, *17*, 638. [[CrossRef](#)]
127. Liao, K.; Song, Y.; Nie, X.; Liu, L.; Qi, S. Suspended sediment concentrate estimation from Landsat imagery and hydrological station in Poyang Lake using machine learning. *IEEE Access* **2024**, *12*, 85411–85422. [[CrossRef](#)]
128. Pisanti, A.; Magri, S.; Ferrando, I.; Federici, B. Sea water turbidity analysis from Sentinel-2 images: Atmospheric correction and bands correlation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *48*, 371–378. [[CrossRef](#)]
129. De Visser, M.H.; Messina, J.P. Exploration of sensor comparability: A case study of composite MODIS Aqua and Terra data. *Remote Sens. Lett.* **2013**, *4*, 599–608. [[CrossRef](#)]
130. Kang, E.; Park, S.; Kim, M.; Yoo, C.; Im, J.; Song, C.K. Direct aerosol optical depth retrievals using MODIS reflectance data and machine learning over East Asia. *Atmos. Environ.* **2023**, *309*, 119951. [[CrossRef](#)]
131. Chao Rodríguez, Y.; el Anjoumi, A.; Domínguez-Gómez, J.A.; Rodríguez-Pérez, D.; Rico, E. Using Landsat image time series to study a small water body in Northern Spain. *Environ. Monit. Assess.* **2014**, *186*, 3511–3522. [[CrossRef](#)]
132. Mansaray, A.S.; Dzialowski, A.R.; Martin, M.E.; Wagner, K.L.; Gholizadeh, H.; Stoodley, S.H. Comparing PlanetScope to Landsat-8 and Sentinel-2 for sensing water quality in reservoirs in agricultural watersheds. *Remote Sens.* **2021**, *13*, 1847. [[CrossRef](#)]
133. Tian, D.; Zhao, X.; Gao, L.; Liang, Z.; Yang, Z.; Zhang, P.; Wu, Q.; Ren, K.; Li, R.; Yang, C.; et al. Estimation of water quality variables based on machine learning model and cluster analysis-based empirical model using multi-source remote sensing data in inland reservoirs, South China. *Environ. Pollut.* **2024**, *342*, 123104. [[CrossRef](#)] [[PubMed](#)]
134. Soomets, T.; Uudeberg, K.; Jakovels, D.; Brauns, A.; Zagars, M.; Kutser, T. Validation and comparison of water quality products in Baltic lakes using Sentinel-2 MSI and Sentinel-3 OLCI data. *Sensors* **2020**, *20*, 742. [[CrossRef](#)]
135. Bonansea, M.; Ledesma, M.M.; Rodríguez, M.C.; Pinotti, L.P. Using new remote sensing satellites for assessing water quality in a reservoir. *Hydrol. Sci. J.* **2019**, *64*, 34–44. [[CrossRef](#)]
136. Sha, J.; Li, X.; Zhang, M.; Wang, Z.-L. Comparison of Forecasting Models for Real-Time Monitoring of Water Quality Parameters Based on Hybrid Deep Learning Neural Networks. *Water* **2021**, *13*, 1547. [[CrossRef](#)]
137. McCarthy, M.J.; Zhang, Y.; Song, X.; Zhu, G.; Paerl, H.W.; Qin, B. Controlling harmful algal blooms in Lake Taihu, China: Nutrient reductions or hydrological controls? *J. Environ. Manag.* **2018**, *223*, 856–862. [[CrossRef](#)]
138. Rajaveni, S.P.; Muniappan, N.; Nandhu, M.; Madhavan, V.S.; Kumar, T.P. Assessment of surface water quality based on Landsat-9 Operational Land Imager combined with GIS and IoT. *J. Indian Soc. Remote Sens.* **2024**, *52*, 139–151. [[CrossRef](#)]
139. Batur, E.; Maktav, D. Assessment of surface water quality by using satellite images fusion based on PCA method in Lake Gala, Turkey. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2983–2989. [[CrossRef](#)]
140. German, A.; Shimoni, M.; Beltramone, G.; Rodríguez, M.I.; Muchiut, J.; Bonansea, M.; Scavuzzo, C.M.; Ferral, A. Space-time monitoring of water quality in an eutrophic reservoir using Sentinel-2 data: A case study of San Roque, Argentina. *Remote Sens. Appl. Soc. Environ.* **2021**, *24*, 100614. [[CrossRef](#)]

141. Kutser, T.; Metsamaa, L.; Strömbeck, N.; Vahtmäe, E. Monitoring cyanobacterial blooms by satellite remote sensing. *Estuar. Coast. Shelf Sci.* **2006**, *67*, 303–312. [\[CrossRef\]](#)
142. Muscutt, A.D.; Harris, G.L.; Bailey, S.W.; Davies, D.B. Buffer zones to improve water quality: A review of their potential use in UK agriculture. *Agric. Ecosyst. Environ.* **1993**, *45*, 59–77. [\[CrossRef\]](#)
143. Engman, E.T. Remote sensing in hydrology. *Geophys. Monogr. Ser.* **2002**, *108*, 165–177.
144. Cao, Z.; Ma, R.; Pahlevan, N.; Liu, M.; Melack, J.M.; Duan, H.; Xue, K.; Shen, M. Evaluating and optimizing VIIRS retrievals of chlorophyll-a and suspended particulate matter in turbid lakes using a machine learning approach. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4211417. [\[CrossRef\]](#)
145. Muchini, R.; Gumindoga, W.; Togarepi, S.; Masarira, T.P.; Dube, T. Near real-time water quality monitoring of Chivero and Manyame Lakes of Zimbabwe. *Proc. Int. Assoc. Hydrol. Sci.* **2018**, *378*, 85–92. [\[CrossRef\]](#)
146. Coffey, M.M.; Nezhin, N.P.; Bartlett, N.; Pasakarnis, T.; Lewis, T.N.; DiGiacomo, P.M. Satellite imagery as a management tool for monitoring water clarity across freshwater ponds on Cape Cod, Massachusetts. *J. Environ. Manag.* **2024**, *355*, 120334. [\[CrossRef\]](#)
147. Bresciani, M.; Stroppiana, D.; Odermatt, D.; Morabito, G.; Giardino, C. Assessing remotely sensed chlorophyll-a for the implementation of the Water Framework Directive in European perialpine lakes. *Sci. Total Environ.* **2011**, *409*, 3083–3091. [\[CrossRef\]](#) [\[PubMed\]](#)
148. Bernier, P.Y. Microwave remote sensing of snowpack properties: Potential and limitations. *Hydrol. Res.* **1987**, *18*, 1–20. [\[CrossRef\]](#)
149. Caballero, I.; Navarro, G. Monitoring cyanoHABs and water quality in Laguna Lake (Philippines) with Sentinel-2 satellites during the 2020 Pacific typhoon season. *Sci. Total Environ.* **2021**, *788*, 147700. [\[CrossRef\]](#)
150. Quemada, C.; Pérez-Escudero, J.M.; Gonzalo, R.; Ederra, I.; Santesteban, L.G.; Torres, N.; Iriarte, J.C. Remote sensing for plant water content monitoring: A review. *Remote Sens.* **2021**, *13*, 2088. [\[CrossRef\]](#)
151. Zhang, K.; Amineh, R.K.; Dong, Z.; Nadler, D. Microwave sensing of water quality. *IEEE Access* **2019**, *7*, 69481–69493. [\[CrossRef\]](#)
152. Ferdous, J.; Rahman, M.T.U. Developing an empirical model from Landsat data series for monitoring water salinity in coastal Bangladesh. *J. Environ. Manag.* **2020**, *255*, 109861. [\[CrossRef\]](#) [\[PubMed\]](#)
153. Mobley, C.D. *Light and Water: Radiative Transfer in Natural Waters*; Academic Press: Cambridge, MA, USA, 1994.
154. Gege, P. WASI-2D: A software tool for regionally optimized analysis of imaging spectrometer data from deep and shallow waters. *Comput. Geosci.* **2014**, *62*, 208–215. [\[CrossRef\]](#)
155. Zhang, Y.; He, X.; Lian, G.; Bai, Y.; Yang, Y.; Gong, F.; Wang, D.; Zhang, Z.; Li, T.; Jin, X. Monitoring and spatial traceability of river water quality using Sentinel-2 satellite images. *Sci. Total Environ.* **2023**, *894*, 164862. [\[CrossRef\]](#) [\[PubMed\]](#)
156. Zhang, Y.; Kong, X.; Deng, L.; Liu, Y. Monitor water quality through retrieving water quality parameters from hyperspectral images using graph convolution network with superposition of multi-point effect: A case study in Maozhou River. *J. Environ. Manag.* **2023**, *342*, 118283. [\[CrossRef\]](#)
157. Dai, H.; Quddoos, A.; Naz, I.; Batool, A.; Yaseen, A.; Ali, M.; Alzahrani, H. Geospatial decision support system for urban and rural aquifer resilience: Integrating remote sensing-based rangeland analysis with groundwater quality assessment. *Rangel. Ecol. Manag.* **2025**, *99*, 102–118. [\[CrossRef\]](#)
158. Graffeuille, O.; Koh, Y.S.; Wicker, J.; Lehmann, M. Remote sensing for water quality: A multi-task, metadata-driven hypernetwork approach. In Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence AI for Good, Jeju Island, Republic of Korea, 3–9 August 2024. Available online: <https://www.ijcai.org/proceedings/2024/806> (accessed on 30 March 2026).
159. Ogashawara, I.; Mishra, D.R.; Gitelson, A.A. *Remote Sensing of Inland Waters: Background and Current State-of-the-Art*; Elsevier: Amsterdam, The Netherlands, 2017.
160. Friedmann, E.; Gleason, C.J.; Feng, D.; Langhorst, T. Estimating riverine total suspended solids from spatiotemporal satellite sensor fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 15443–15462. [\[CrossRef\]](#)
161. Pahlevan, N.; Smith, B.; Alikas, K.; Anstee, J.; Barbosa, C.; Binding, C.; Brando, V.; Campbell, G.; Casey, B.; Costa, M.; et al. Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. *Remote Sens. Environ.* **2022**, *270*, 112860. [\[CrossRef\]](#)
162. Joshi, N.; Park, J.; Zhao, K.; Londo, A.; Khanal, S. Monitoring Harmful Algal Blooms and Water Quality Using Sentinel-3 OLCI Satellite Imagery with Machine Learning. *Remote Sens.* **2024**, *16*, 2444. [\[CrossRef\]](#)

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.