BILINGUAL PUBLISHING GROUP

Journal of Environmental & Earth Sciences

https://journals.bilpubgroup.com/index.php/jees

REVIEW

AI-Enabled Digital Twin Framework for Real-Time Water Resource Management Using Multi-Source Remote Sensing Data

Yunyan Chang

Chongqing Water Resources and Electric Power Vocational and Technical College, Chongqing 402160, China

ABSTRACT

Traditional water resource management relies on statically configured models and sparse in-situ networks, creating critical gaps in situational awareness that lead to operational failures during floods, droughts, and water quality incidents. This review synthesizes advancements in AI-enabled digital twins—constantly updated, stateful digital representations that synchronize with physical water systems through continuous assimilation of multi-source remote sensing data. Unlike conventional modeling workflows, these closed-loop systems integrate heterogeneous observations (optical, Synthetic Aperture Radar (SAR), thermal, microwave, and altimetry) with in-situ Internet of Things (IoT) measurements to maintain real-time alignment with evolving conditions. We propose a unified reference architecture spanning data ingestion, AI-driven downscaling and retrieval, quality-aware multi-modal fusion, state synchronization via data assimilation, probabilistic forecasting using physics-AI hybrids, and decision support with continuous monitoring. The framework explicitly addresses operational constraints, including latency, missing data, and non-stationarity, while prioritizing uncertainty calibration over point accuracy. Our synthesis evaluates design trade-offs across flood response, reservoir operations, drought monitoring, irrigation management, and water quality applications. We conclude by identifying research priorities: standardized state schemas and uncertainty metrics, interoperable application programming interfaces (APIs), robust domain adaptation, and governance frameworks incorporating human-in-the-loop safeguards. This review provides a roadmap for transforming heterogeneous remote sensing streams into reliable, actionable intelligence for real-time water resource management.

Keywords: Digital Twin; Water Resource Management; Multi-Source Remote Sensing; Data Fusion; Uncertainty Quantification

*CORRESPONDING AUTHOR:

Yunyan Chang, Chongqing Water Resources and Electric Power Vocational and Technical College, Chongqing 402160, China;
Email: yychang651@163.com

ARTICLE INFO

Received: 2 January 2026 | Revised: 25 February 2026 | Accepted: 1 March 2026 | Published Online: 20 April 2026
DOI: https://doi.org/10.30564/jees.v8i4.12943

CITATION

Chang, Y., 2026. AI-Enabled Digital Twin Framework for Real-Time Water Resource Management Using Multi-Source Remote Sensing Data. Journal of Environmental & Earth Sciences. 8(4): 203–229. DOI: https://doi.org/10.30564/jees.v8i4.12943

1. Introduction

Water resource systems are moving into a period where the average conditions are less important, and the swift oscillations, compound extremes, and cascading effects are increasingly important. What may begin as a small-scale high-intensity precipitation develops into a rapid urban flood in hours, and what may start as a drought in weeks and months may suddenly go through limits that lead to curtailments, ecological stress, or water-quality disasters [1]. Meanwhile, the water agency operating environment continues to grow more complex: multi-purpose reservoir objectives are premised on a requirement to balance flood management and water resources, hydropower and flood mitigation, as well as environmental flows; multi-purpose upstream irrigation water districts are faced with a greater need to operate at field-scale responsiveness to weather variability and crop demand, and managers of lakes, rivers, and estuaries find that episodic contamination, salinity intrusion, and harmful algal bloom are becoming increasing challenges. Within these settings, one common constraint is not merely the unavailability of models or the unavailability of data, alone, but the unavailability of situational awareness at the system level that could be constantly transformed into uncertain, actionable choices [2,3].

The conventional way of water management employs in situ monitoring networks that consist of stream gauges, meteorological stations, groundwater wells, and water-quality sondes, among other things, in conjunction with the hydrologic and hydraulic models that are based on physics. This paradigm is interpretable and has a well-implemented workflow, but is still not without challenges in how to operate in real time. Monitoring networks are controlled unevenly and are susceptible to outages, especially in remote or resource-limited areas. Cases of model skill degradation may include when there is uncertainty in the meteorological forcings, when parameters vary, or when land surface and channel conditions vary with development or vegetation dynamics, wildfire, or sedimentation. Handoffs between monitoring and analysis, and action are often done by hand and cause delays and diminished flexibility to change operations as conditions change [4]. Simultaneously, the past decade has experienced an accelerated growth in Earth observation potential, which provides unprecedented coverage and repetitive measurements that are directly applicable to the water cycle. Nevertheless, remote sensing, in itself, does not often provide decision-ready information at the spatial and temporal resolution needed by operational management. Products differ in scale, latency, uncertainty, and even physical meaning, and their usefulness might be limited by clouds, revisit gaps, signal saturation, or indirect dependence on hydrologic states. Consequently, the problem with a lot of agencies is that the system has become data-rich, and the information ready to be used in decisions is hard to compile in a quick, consistent, and transparent way [5].

A new answer to this paradox is the creation of digital twins of water resources, which is facilitated more and more by artificial intelligence and multi-source remote sensors. The main concept is to have a constantly updated, stateful digital model of a water system that keeps abreast of reality by incorporating streaming observations and can provide both short- to medium-term predictions and actionable conditions. Within this perspective, remote sensing streams are not just dashboard layers; they would provide active inputs to change the internal condition of the twin, restrict future solutions, and decision-making in the form of a closed-loop system which interacts with observation, guesswork, anticipation, action, and learning. The closed-loop view is especially applicable to operational contexts since it also focuses on predictive accuracy in addition to responsiveness, resilience to missing data, explicit uncertainty management, and the traceability required to make accountable decisions [6–8].

The closed-loop logic of how an operational water digital twin observes, synchronizes, predicts, decides/acts, and learns is summarized in **Figure 1**.

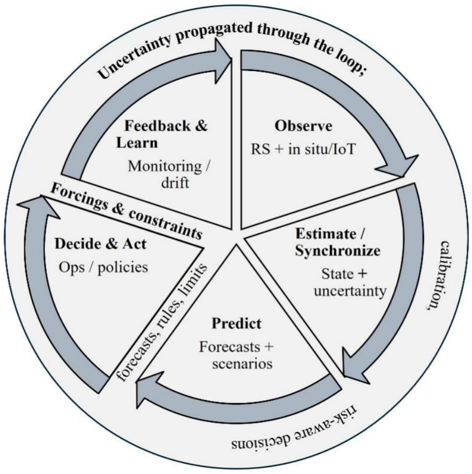

Figure 1. Conceptual closed-loop water digital twin.

The digital twin has inconsistent use in the literature, where it is applied to both a static digital replica and dynamic systems with continuous synchronization and control that are cyber-physical. In water resource management, a useful definition should lay stress on the state development and the applicability of operations. In this review, a digital twin of a water-resource is assumed to be a computational system maintaining an explicit evolving state of the water system (like the storage volume, water flowing, soil moisture conditions, to the point of inundation, or even water-quality indicators), it periodically updates that state with incoming observations to keep it consistent with reality, it provides forecasts or counterfactual proposals to aid planning and risk evaluation and even produces decision-supporting quantifiable and/or auditable uncertainty. This view separates the category of a digital twin and a visualization platform, which reports observations without systematic state estimation, an independent forecasting model that runs every now and then, and an offline calibrated model from which adaptation to variable conditions is impossible. It is also used to differentiate between the twin and single-purpose machine learning predictors that can yield the correct results on a certain variable but do not reflect system-wide conditions or have the ability to maintain feedback between the decision and subsequent updates. To clarify the scope and differentiate water-resource digital twins from related systems, **Table 1** compares digital twins with dashboards, standalone predictive models, and offline calibrated models[9–11].

Table 1. Core concepts and boundary conditions for a water-resource digital twin.

Concept/System	Primary Purpose	Key Characteristics	What It Typically Lacks vs. a Digital Twin	Typical Outputs
Digital twin (water resources)	Real-time, closed-loop estimation–forecast–decision support	Stateful representation; continuous synchronization; uncertainty-aware; auditability; feedback/learning	—	Current-state estimates + probabilistic forecasts + decision products
Monitoring dashboard/Geographic Information System (GIS) viewer	Situational awareness and visualization	Map layers; static or periodic updates	State estimation, forecasting, closed-loop updating, decision optimization	Maps, time series displays
Standalone forecasting model	Predict future states/flows	Periodic runs; physics or ML	Continuous synchronization; multi-source fusion; operational guardrails	Deterministic or probabilistic forecasts
Offline calibrated model	Scenario analysis, planning	Calibrated parameters; slower updates	Real-time ingestion, adaptation to drift, latency guarantees	Planning simulations
Single-task AI model	Predict one target variable	Fast inference; data-driven	System-wide state representation; explicit constraints; feedback loop	Soil moisture/ET/inundation predictions

The key component of this digital twin vision is multi-source remote sensing, which is able to increase observability in areas that would not be possible with ground sensors. The water cycle can be handled in a direct response to sensed modalities using the modern Earth observation systems. In favorable conditions, optical imagery can outline surface water and can facilitate the inferences about turbidity or algal blooms, synthetic aperture radar can be used to detect inundation during cloud cover and at night, passive microwave and radar-derived products can provide details about precipitation and soil moisture, thermal observations can be used to estimate the land surface temperature, and altimetry can be used to monitor water levels in lakes and rivers, and gravimetry can reveal anomalies in large-scale storage. But it is their very diversity that poses an integration challenge[12]. The individual sensors possess different revisit frequencies, spatial resolutions, latencies, and error properties, and the connection between what is measured by a sensor and the management-relevant state variable is usually indirect. As an example, water, including the products of a precipitation event, can be biased in a storm-dependent and topographically-dependent way. Soil moisture retrieval can be contaminated by plant biomass and mountainous roughness, optical QoI water observation may fail in the clouds and gloom, and radar water mapping can be frustrated by novel vegetation or mountain-bumped surfaces. The most fundamental challenge in a wide range of real-time situations is the fact that no single product will give full, consistent, and up-to-date coverage. The task of operations is thus to fuse techniques that can combine partial and biased observations into consistent, uncertainty-sensitive estimates [13].

This synthesis is possible through the use of artificial intelligence, which offers ways of information extraction, downscaling, fusion, and quick prediction. Machine learning used on the observation side can assist in the retrieval of hydrologic variables using raw remote sensing signals in a way that can be robust to noise and heterogeneity compared to purely empirical or purely physics-based retrieval pipes under the conditions that sufficient training data and constraints are at their disposal. Another way that AI can carry out quality control is by detecting artifacts and screening anomalies, which is crucial when streaming pipelines, and it is not possible to manually screen them. Between the scale levels of spatiotemporal aspects in management requirements and coarse measurements of operational resolution (fine units), coarse measurements in time (temporal sharpening), and operational domains with translation. cross-sensors (learned gap-filling) may be used to fill lapses and gaps. On the modeling side, with surrogates that are faster than expensive simulators, AI can speed up computation, facilitate faster ensembles, exploration in scarce scenarios, and increase updates. Data-driven and physics-based models. Hybrid models, which combine elements of data-driven and physics-based modeling, can lessen systematic biases, data consistency with constraints, and the generalization weaknesses of purely data-driven models when they become ill-posed due to regime changes. To estimate the state, AI may be used to augment data assimilation by means of learning observation operators, approximating costly computational steps, or giving representations that are more useful for fusing multi-modal data. Lastly, AI can be applied in decision support to optimize and control by learning a policy or value function, especially when a decision space is extensive or conflicting goals are put in place, as long as safety factors and accountability are ensured [14–17].

Simultaneously, the sphere of operational water management is a high-stakes one, and AI presents certain risks that have to be discussed thoroughly within the context of a digital twin. The models are prone to overconfidence when they are out of the distribution of their training, particularly when they are near extreme values that are seldom found in historical data, but are at the center of concern in risk management. The performance may decline due to land-use transition, sensing drift, baselines of changing climates, or unexpected infrastructure activities. Many potent learning models are hard to decipher, and they make it hard to communicate with the operators and regulators. These are not peripheral concerns, but they define what it is to make a digital twin deployable and not just demonstrative. This means that the incorporation of AI into the water digital twins will need to be completed with uncertainty quantification, drift monitoring, strong fallback measures, and human-in-the-loop mechanisms that will ensure the operators can interpret, question, and ignore the algorithmic Advisories where necessary.

It is against this context that the literature on water digital twins and integration of AI-enabled remote sensing is quickly growing but fragmented in the fields of hydrology, remote sensing, machine learning, systems engineering, and control. Various communities have different terms and assumptions, and research tends to consider single components, e.g., flood inundation mapping, soil moisture retrieval, evapotranspiration estimation, or streamflow prediction, without explaining how such individual components are brought to work together to achieve the operation reliability requirements of a closed-loop digital twin. Evaluation practices are also rather divergent: some studies focus on pixel-level performance, whereas others focus on hydrologic skill, and only a subset of studies have reported latency, uncertainty calibration, performance in the face of missing data, and decision-level performance. The discontinuity renders the translation of research developments into feasible advice towards the development and implementation of a system capable of providing reliable real-time management to support systems [18,19].

This review thus generalizes the area in terms of an AI-driven digital twin framework through multi-source remote sensing. The idea is to summarize what can be done already, explain architectural patterns that are repeated in the various successful applications, and determine the circumstances when various modeling and fusion approaches are suitable. A series of reiterating questions, including what would make a robust reference architecture behind real-time water digital twins, how should multi-source remote sensing products be aligned, fused, and assimilated to keep synchronized under latency constraints, where AI would add the most value compared to physics-based modeling, and where it poses unacceptable risk, frames most of the review, along with how uncertainty may be reported as well as propagated between observations and forecasts into decisions [20].

Formulating the evidence based on these questions, the article is supposed to deliver a handy roadmap for both researchers and practitioners. It also points out a taxonomy of methods, by management goals, data sources, strategies of coupling, operating loop, and it condenses a reference frame traversing ingestion and retrieval and fusion and state estimation and prediction and decision support, to applications on risk-conscious decision making, and evaluation principles that are based on missing data graceful degradation, and transparency of organizing uncertainty. Finally, according to the review, the community should not merely enhance models or data, but rather come up with standardized, uncertainty-conscious, and interoperable system designs that may transform remote sensing streams into reliable, actionable intelligence to support real-time water resource management [21].

2. Foundations: Data, States, and Operational Constraints

Digital water resource twins are at the crossroads of monitoring the environment, predictive analysis, and integrated decision-making. In contrast to traditional analysis pipelines, a real-time twin is evaluated by its capability to maintain alignment with a dynamical physical system whilst sustaining action-specified outputs with severe time, reliability, and governance demands. The section forms the theoretical and methodological underpinnings required to draw meaning from the literature on AI-enhanced digital twins of water at the behest of multi-source remote sensing (RS). We concentrate on (i) what is real-time water management as a decision making, (ii) how the twin state can be modeled and defined at scales, (iii) what are the latency, robustness and reliability requirements that make a difference in choosing an architecture, and (iv) what are the governance and the system-engineering requirements that make the deployment of the twin's candidate and not the research prototype [22].

2.1. Real-Time Water Management as a Closed-Loop Decision Process

Water resource management is a control problem by definition: repeated decision-making based on incomplete information in an effort to meet goals, which would be limited by physical, legal, or operational constraints. Practically, this control dilemma is shared by several actors (operators, forecasters, regulators, emergency managers, irrigation schedulers) and the decisions are made at varying cadences. As an illustration, response to flash floods can need news of the order in minutes to hours, operations of reservoirs can be on hourly down to daily schedules, and drought allocation and seasonal planning can be on weekly or a month-to-month basis. A digital twin to be used in a so-called real-time environment has to be integrated into a designated decision loop instead of being perceived as a generic monitoring platform [23].

One of the useful abstractions is the observe, estimate, predict, decide act, learn. Observations are both of RS products as well as ground-based measurements, estimation synchronizes the internal state of the twin with available evidence that is best known, prediction projects the state forward, and gets the physical system, and learn provides performance feedback and detects failure or drift. Implementing a digital twin is not characterized by a particular aspect but the wholeness of the loop, as the system needs to change in reaction to new information and the effects of the previous actions, and not to generate one-off predictions out of operational feedback [24].

This framing also explains the reason why the accuracy is not enough. An operational model can be correct in average form but operationally incorrect in that it becomes overconfident when the data is at extremes, and fails silently when one of the sensor streams fails, or it does not give results within the decision-making time. In contrast, a system with a small average skill can also be useful when it is adequately calibrated, timely, and resilient to missing data as well as transparent on the uncertainty in a manner that enables risk-taking. This change of focus from predictive assessment to operational preparedness is increasingly represented in the literature, particularly in situations where decision-making is safety-critical or legally accountable.

2.2. Defining the Digital Twin State: What Must Be Synchronized?

An essential design decision is how to describe the internal state of the twin, i.e., the variables that represent the system at any particular time and are updated with new observations. In water resources, the state in question will rely on the purpose of management, although it will normally be owned by multiple storage and transport systems, as well as quality processes in several compartments. In flood-centric twins, state variables can be the soil moisture in the catchment, channel flow, water levels and the extent of flooding in the flood plains. In the case of reservoir operations, the state may also have reservoir level/storage, inflows, outflows, upstream snow and soil moisture conditions (as an antecedent to inflow), and state such things as rule curves/turbine capacities. The state can focus on root-zone moisture proxies, crop water requirement, evapotranspiration (ET), and the state of a conveyance system to irrigate the land and manage agriculture. With water-quality applications, the state may go to temperature, proxies of turbidity, nutrient indicators, dynamics of dissolved oxygen, or to an index of bloom susceptibility, depending on observables available and the requirements of regulation [25].

Due to the spatial organization of water systems, the state representation is also a geometrical option. Examples of common representations are a regular grid (easy to incorporate with an RS), river networks (natural to model routing and hydraulic models), and node-link graphs (reservoirs, diversions, canals, and control structures). In practical applications, hybrid representations are often used: a grid over land-surface states is used as input to a routing model on a network; reservoir nodes are modeled to have operation constrained by operational bounds, urban drainage also potentially receiving high-resolution hydraulic representations in critical regions, whilst the non-urban basin is simulated at a lower level. This multi-resolution property does not just make verification easier using computers; it is the result of it being a fact that certain decisions are localized (e.g., gate operations, urban inundation hotspots) and other decisions are basin-scale (e.g., allocation policies, storage balancing).

The definition of a state also defines how RS data may be exploited. There are those state variables that are directly related to the RS product (e.g., surface water extent as an indicator of inundation) and indirect (e.g., land surface temperature as an indicator of ET and the energy balance, and gravimetry as a storage anomaly at a coarse scale). An overly narrow state can behave by paying little to no attention to the major drivers and provide volatile forecasts; an overly broad state can be unobservable based on the data streams available, and as a result, provide unidentifiable updates and spuriously confident forecasts. Practical literature tends to reduce to a minimal, observable, operational state which is predictive enough to make relevant forecasts to guide decisions, and other variables are regarded as diagnostics or scenario inputs [26–28].

2.3. Real-Time Constraints: Latency, Cadence, and Compute Budgets

The operationalization of the meaning of real-time has to be in terms of latency budget: the largest allowable interval between data availability and decision-ready output. The different causes of latency include satellite acquisition times and product delivery times, time spent in preprocessing to harmonize, mask clouds, calibrate, or tile; inference and fusion time; assimilation or ensemble forecasting run time; organizational functions, including validation and authorization. The policy or risk tolerance may fix the decision window in most working situations (e.g., reservoir releases are updated at a fixed time, a warning is issued at a certain threshold, irrigation schedules operate in cycles). The twin has to be designed, then, to match the slowest component of the end-to-end chain, rather than to optimize any one component of the model [29].

This results in typical architectural trade-offs. High-fidelity physics models can give detailed, interpretable physics dynamics, but can be unusable due to their slowness in their update rate or in large ensembles. AI surrogates can be faster at making predictions, and can quickly explore new scenarios, but they are fragile to distribution shift and need to be accompanied by uncertainty estimates and monitoring. Assimilation of data enhances the synchronization, but it may be computationally intensive; high-frequency operation may require approximations, reduced-order models, or AI-accelerated operators. Multi-source fusion can also enhance completeness, but with the extra overhead of alignment and quality control, which needs to be automatable to be used in real-time.

Mismatch of cadences of observation streams is a related problem. There are those RS inputs that are near-real-time with high revisit (depending on sensor type and area) and others that can be updated daily or less. On-site sensors could be continuous and sparse and could be affected by outages. Meteorological forecasts automatically update themselves. A digital twin has to coordinate these various cadences with either an asynchronous update, an event-based update, or periodic assimilation. One typical approach is to model the twin as a streaming system capable of accepting any incoming data when it is available, opportunistically updating any pertinent state components, and having an integrated best current estimate with associated uncertainty and provenance [30].

The location of the computer also counts. Scalability and heavy workloads (e.g., ensembles, large mosaics) are supported by centralized cloud execution, whereas edge or near-edge inference can be used to minimize latency and enhance resilience in bandwidth-constrained environments. To take a review view, the most important fact is that operational constraints are not an issue that comes after; they determine the methods that can be realised and the research findings that can be made into deployable systems [31].

2.4. Robustness and Reliability: Missing Data, Drift, and Extremes

The operational water systems have to work in the conditions that the data and models are most likely to fail: storms that cause sensors to become inoperable, wildfire smoke that lowers optical retrievals, cloud cover that continues during the monsoon seasons, infrastructure adjustments that invalidate calibrated parameters, or extreme extremes never seen before and exceeding the training distributions. The requirement of robustness thus becomes a first-class requirement. System-wise, robustness involves the capability of (i) noticing the absence or unreliability of data, (ii) gracefully degrading by switching to other sources or fallback models, (iii) sustaining calibrated uncertainty that indicates less information, and (iv) recovering rapidly when streams restart [32].

Multi-source RS provides redundancy, which is only effective when fusion algorithms are capable of dealing with partial modalities without hallucinating structure. As an illustration, with no optical imagery available because of clouds, radar-inundation mapping can be used to some extent, although vegetation and roughness of the surface are problematic. Where the soil moisture retrieval is crude or slow, in situ moisture proxies and precipitation forecasts could offer partial substitutes. To enable a strong twin, a clear missingness treatment (e.g., masking, imputation with uncertainty, modality dropout training), quality-aware fusion (weighting sources by indicators of reliability) is required.

Model drift is another operational hazard. Drift may be caused by sensor changes (recalibrations, algorithm updates), changes in the environment (land use, channel morphology), or non-stationary climatic conditions. Drift in AI components can take the form of varying error distributions and overconfident predictions; drift in physics components can be in the form of persistent bias or poor event timing. A functional twin must thus encompass monitoring layers to monitor performance indicators, anomalies and cause recalibration or retraining. Notably, drift management is a challenge that needs both technical and governance infrastructure: version control of data products and models, reproducible pipelines and processes to verify updates before moving to production [33].

Special attention should be given to extreme events as they are decision-critical, as well as underrepresented in training data. Research prototypes also tend to report mean performance measures which may mask tail failure. In terms of review, a key methodological shortcoming of the literature is inconsistent reporting of extreme-specific competence, resilience to stress tests (e.g., sensor dropout), and uncertainty calibration during events. The framework of a strong digital twin assumes extremes as a major target of evaluation rather than a corner case [34,35].

2.5. Data Governance, Provenance, and Interoperability for Deployable Twins

The deployed digital twins have to meet the needs that are somewhat underrepresented in academic research: provenance, traceability, and interoperability. The provenance can be defined as the possibility to trace the origin of every piece of information, such as sensor type, version of processing algorithm, timestamps, and quality flags. Traceability applies to decisions: an operator must be able to trace which data and model versions created a specific forecast and recommendation. These attributes are critical towards accountability within the controlled settings and also diagnosing failures during post-event analysis [36].

Interoperability is also significant since water agencies hardly work in greenfield settings. Information should be connected to the preexisting hydrology databases, SCADA systems, forecast centers, and GIS systems. State is then supposed to be presented as stable interfaces (APIs) and with well-defined metadata. The control of access and cybersecurity in a multi-stakeholder environment is also essential: ingestion pipelines should protect against data flows with corruption, unauthorized changes, and accidental disclosure of sensitive infrastructure information. Although these issues might seem engineering-friendly, they have a direct effect on methodology, like whether or not to employ black-box models whose validation may not be explicitly available, or whether uncertainty quantities are in forms accessible to downstream decision systems [37].

2.6. Implications for How the Literature Should Be Read and Compared

The above foundations mean that the comparison of digital twins studies is not achieved by comparing model architectures only. Such a meaningful comparison should take into consideration the context of the decision being made (task and cadence), what is being synchronized (state definition), the list of modalities and latencies available to make the observation, the fusion/assimilation strategy (and treatment of missing data), and such functional properties as runtime, reliability, and uncertainty calibration, and monitoring. Numerous published papers add some useful elements: retrieval models, fusion networks, surrogate simulators, but they fail to specify the interaction of these elements to form a closed loop of operation in real time and the overall behavior of this loop in the presence of outage and drift. This reason explains why subsequent parts of this review focus on taxonomies matching studies based on a combination strategy and operational preparedness, an assessment direction that improves latency-conscious measures, robustness examinations, and decision-level results in relation to well-known accuracy designs [38].

To conclude, an AI-enabled water digital twin can be viewed as a system that needs to align a carefully selected state that is subjected to heterogeneous and imperfect observations governed by tough operational conditions. Multi-source remote sensing increases observability and brings about fusion and government, AI creates potent mechanisms of retrieval, downscaling, fusion, and acceleration without negating the dispute of robustness and trust. It is these foundations that give the perspective upon which the following sections will package together techniques and practice, and which encourage why the next improvements in the field will be due as much to uncertainty-conscious system design, and to operational integrations, as to evolutionary advances in predictive ability [6,39–41].

3. Multi-Source Remote Sensing and AI-Enabled Information Extraction

Multi-source remote sensing (RS) is the most feasible path with the highest scaling ability of exploring the dynamics of water systems above the in-situ network limitations. RS is no longer a default in the case of an AI-enabled digital twin; it is one of the key proponents of system observability, in which case the twin can reconcile states across large regions, like ungauged basins and areas with dense monitoring infrastructure, or at times unavailable or prohibitively costly. Its operational issue is that the heterogeneity in spatial resolution, revisit time, latency, viewing geometry, and error structure of RS data is not observed, but rather a set of variables that are of water interest is predicted by retrieval algorithms. To this end, the success of RS-based digital twins depends on robust information extraction pipelines that synchronize multi-sensor information, quantify unpredictability, and offer variants that are operationally relevant at the frequency [42,43]. **Table 2** sums up some typical RS modalities, water variables that they informed, and their common strengths and RS limitations, and examples of how they apply to a digital twin.

Table 2. Multi-source remote sensing inputs for water digital twins.

RS Modality/Data Type	Water-Relevant Variables (Examples)	Strengths for Real-Time Twins	Common Limitations	Typical Twin Uses
Optical multispectral	Surface water extent; land cover; water-quality proxies (when feasible)	High spatial detail; intuitive interpretation	Clouds/haze/smoke; illumination dependence	Water extent updates; shoreline dynamics; land-cover constraints

Table 2. *Cont.*

RS Modality/Data Type	Water-Relevant Variables (Examples)	Strengths for Real-Time Twins	Common Limitations	Typical Twin Uses
SAR (Synthetic Aperture Radar)	Inundation extent; wetland dynamics; surface roughness cues	All-weather, day–night capability	Vegetation/buildings/terrain effects; speckle; geometric distortion	Flood mapping under clouds; rapid event updates
Thermal infrared (IR)	Land/water surface temperature; ET-related inputs/proxies	Drought/irrigation relevance; stress indicators	Atmospheric correction; indirect link to ET; mixed pixels	ET estimation support; drought stress monitoring
Microwave (passive/active products)	Soil moisture; precipitation-related fields (product-dependent)	Frequent coverage in many regions; weather resilience	Often coarse resolution; retrieval ambiguity	Antecedent wetness constraints; forcing correction
Altimetry	River/lake/reservoir water level (where observable)	Direct relevance to storage/levels	Sampling limits; narrow channels challenging	Storage/level constraints; calibration/validation
Gravimetry-scale storage	Total water storage anomalies (coarse scale)	Basin-scale drought/storage context	Very coarse spatial support; not local	Large-scale storage constraint; trend diagnostics
In situ / IoT (complementary)	Stage/flow, rainfall, water quality, telemetry	High-frequency, direct measurement	Sparse; outages; maintenance	Ground truth anchors; assimilation; drift detection

3.1. Remote Sensing Modalities and What They Contribute to Water-System Observability

One characteristic of modern observation of the Earth is the diversity of modality. Each modality gets a separate view of the water cycle, but with its own modality-generated constraints that constrain models and assimilation at the downstream. **Figure 2** shows the role of the complementary observations of optical, SAR, thermal, microwave, altimetry/gravimetry, and in situ/IoT by summarizing their mapping to the water-relevant variable.

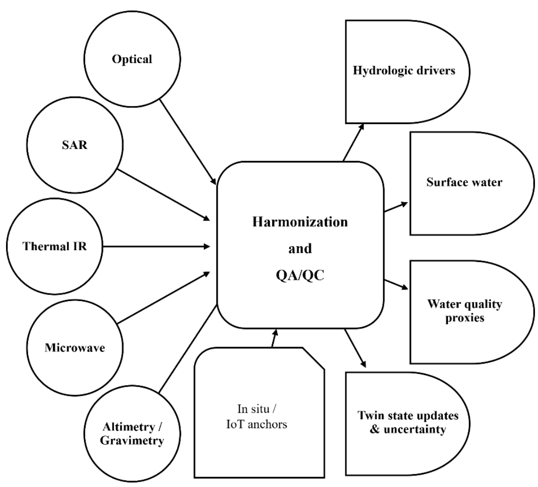

Figure 2. Multi-source remote sensing observation stack for water variables.

The optical multispectral imagery is extensively applied in surface water cover mapping, shoreline and land cover modeling, and in deriving proxies of interest to water quality (e.g., turbidity and algal-bloom indicators) when atmospheric and optical quality conditions are good. Optical data may offer great spatial resolution and local scale surveillance, but is limited by cloud cover, haze, and smoke, as well as conditions of illumination. These limitations are, in many cases, the worst in the same periods of most hydrologic concern, including storm seasons, in the operational setting. Synthetic aperture radar (SAR) is an enhancement to optical imagery and has all-weather, day-night operations as well as well-sensitized to surface water and inundation, which is especially important in flood mapping and wetland dynamics. SAR can also be used in cloudy conditions, although the geometric distortions caused by vegetation, water surface roughness caused by the wind, built environments, and terrains can make the interpretation of the image challenging. The problems are what render the preprocessing and classification strategies central to the trusted SAR-based inundation products[44].

Thermal infrared measurements can be used to estimate land surface temperature, and can be used to infer evapotranspiration (ET) along with meteorological data and equations of surface energy balance. Thermal signals are instructive indicators of drought and irrigation due to their relevance to the land-atmosphere exchange and plant water stress proxies, but they are indirect and deviant to admissible atmospheric correction error, emissivity modeling, and mixed pixels.

Products based on passive microwave and radar also provide information on precipitation and soil moisture, which have a fairly high temporal availability in many locations, but at relatively low spatial resolution. This results in a repetitive scale mismatch between what microwave sensors can safely provide and what local managers need (e.g., field-scale irrigation scheduling or runoff response to sub-catchment), etc. One of the essential forces of AI integration is the resulting necessity to downscale.

Altimetry is used in estimating water levels in rivers, lakes, and reservoirs and offers direct applicability to the management of storage and the flow. Orbit geometry and target size can also constrain coverage and temporal sampling, and in narrow channels, irregular topography, or vegetated areas, quality can be poor. Gravimetry makes basin-scale water storage anomaly data, which may have useful methods in large-scale drought assessment and large-scale groundwater-related research, but is generally too coarse to be useful in independent operational control, and is typically an instrument of constraint or diagnostic in large scales.

Practically, the digital twins have the greatest advantages when these modalities are not substitutes but complementary. A combination of thorough inundation maps can be enhanced by optical and SAR; microwave data on soil moisture can be narrowed to basin-level constraints on antecedent wetness, and altimetry or water level can be used to control storage calculation on a union with bathymetry or empirical area-volume correlations. This complementary philosophy is being replicated in the literature, though the use of single sensor products is replaced by multi-sensor observation stacks, which are designed to operate in the state variables of the twin [45–47].

3.2. Water-Relevant Variables from Remote Sensing: From Observables to Decision-Ready Quantities

In the case of water resource management, RS-derived variables that are classified as either (i) reflect decision-relevant states directly, or (ii) highly constrain state drivers are the most valuable ones. Examples common to it are precipitation, soil moisture, snow cover and melting, ET, surface water coverage, water level, and water quality proxies. However, RS does not often determine these variables per se, but measures radiance or backscatter or brightness temperature or phase differences, which must be converted into hydrologic quantities through retrieval models. This translation also adds structural uncertainty, which should be taken into account with regard to the integration of products into a digital twin.

One common difference in the literature is that primary RS observables (measurement of a sensor) and derived RS products (retrieved variables). Derived products can be more readily included in hydrologic processes, yet might hide sources of uncertainty as well as frustrate flexibility. On the other hand, integrating primary observables may be a theoretically attractive approach, but it needs correct observation operators and requires a large computational and modeling infrastructure. In most working digital twin applications, derived products are the realistic option, assuming that the characterization of uncertainty is known and that the behavior of products and bias is monitored over time [48].

The other practical problem is that the quantities are often in the form of decision-ready formats, which have to be transformed after retrieving them. The increases and decreases in the operation of the reservoir might demand storage instead of the water level; irrigation scheduling might need crop ET or root-zone moisture instead of the half of the surface temperature; flood response might demand depth images and hazard maps instead of the inundation depth itself. Such changes are generally conditional on supporting data (e.g., bathymetry, land cover, soil properties) and introduce an extra uncertainty. Thus, an advanced digital twin pipeline would consider information extraction a multi-stage process: it has to be retrieved, harmonized, and turned into management-relevant indicators without any uncertainty or provenance loss in between [49].

3.3. Harmonization and Preprocessing for Multi-Source Integration

Until AI models can be confident in learning on multi-source RS, downstream fusion and assimilation can be operated predictably. RS data should be moved into a shared spatiotemporal and radiometric framework. Although harmonization is usually not given much focus in methodological papers, it is the core of operational robustness since minor misalignments and undocumented changes in processing can produce systematic errors that cause hydrologic signals [50].

Spatial harmonization is usually associated with georeferencing, reprojection, and resampling to some standard grid, or collection of management units (sub-basins, river reaches, reservoirs, irrigation districts). Temporal harmonization-oriented data to the update frequency of the twin, such as hourly, daily, or events, with an aggregation strategy, interpolation strategy, or windowed composing strategy. The trade-offs depend on each option mentioned above. Temporal aggregation may produce noise reduction, but it may also blur the peaks; compositing is also able to minimize cloud contamination, but can create delays; interpolation may also bridge gaps, but may hallucinate dynamics without treating the uncertainty appropriately.

Radiometric, as well as modality-specific, preprocessing is of equal significance. Optical workflows need to be corrected against atmospheric variations, cloud and shadow masking, and in certain cases, they need bidirectional reflectance distribution function (BRDF) normalization. SAR has to undergo radiometric calibration, use of speckle filters, terrain correction, and considerations of incidence angle effects. Thermal products fail without atmospheric correction as well as emissivity assumptions. In all modalities, cross-sensor calibration and bias correction are crucial when crossing products across platforms or time, and in particular, as operational twins are expected to be stable across sensor upgrades, as well as across algorithm version changes [51].

Metadata and quality indicators also need to be maintained in a harmonized manner in the context of a digital twin. Quality flags, geometry of acquisition, processing lineage, as well as uncertainty estimates, are an optional extra and are required by quality-aware fusion, missing-data processing, and drift monitoring. The commonest type of best practice in the literature is to consider such indicators as being first-class features that are pushed into AI models and fusion layers instead of being lost during preprocessing.

3.4. AI Methods for Retrieval, Downscaling, and Spatiotemporal Reconstruction

AI participates in RS-driven water digital twins in three forms most clearly, namely in retrieval (map of observables to hydrologic variables), downscaling (resolution disparity), and spatiotemporal reconstruction (closure and coherence enforcement over time).

To be able to retrieve, machine learning methods include classical regression and tree-based methods, as well as deep learning unitary structures using spatial context. The application of deep convolutional networks, encoder-decoder models, and attention-based networks is steadily growing, whereby spatial patterns have hydrologic interpretations (e.g., inundation limits, the influence of land cover heterogeneities on surface temperature). The performance of retrieval models can be high in the empirical domain, but the value of the operation of these models is reliant on the issue of generalization; a model that is trained on one basin or season may not work in another one because of vegetation, soil, topography, infrastructure, or atmospheric conditions. In this spirit, physics-inspired constraints are progressively included in studies, such as the existence of relationships (e.g., has to be monotonic) and the existence of mass/energy balance (e.g., should not exceed a physical limit). Multi-task learning of related variables jointly to enhance stability is included in studies [52].

A downscaling is an especially critical example of an AI application since most existing products in the world (particularly microwave-based precipitation and soil moisture) are not as fine as they should be. They are commonly investigated using super-resolution techniques, cross-modality guiding (where a higher-resolution optical/thermal signal is used to refine a low-resolution signal), and using statistical/deep hybrids. In digital twin applications, downscaling has to be determined by not just the similarity to the reference datasets at the pixel level, but also by the ability of the process to preserve hydrologically useful structure (e.g., gradients across the landscape, storm footprints) and by whether the process has calibrated uncertainty. Very sharp downscaled fields may form false accuracy that disrupts assimilation or makes decisions that are too aggressive [53].

Spatiotemporal reconstruction fills in gaps in data due to clouds, revisit gaps, and sensor outages. Gap filling with the aid of AI can also take advantage of seasonal structure and persistence, using models of time (e.g., recurrent networks, temporal convolutions, attention mechanisms), possibly augmented with meteorological covariates. Gap filling is, however, dangerous in a real-time decision environment: it has to explicitly model uncertainty and should not introduce artifacts that are subsequently confused with actual hydrologic change. A combination of uncertainty estimation, masking with modality, and fallback logic that is conservative in weak evidence is the most potent one [54].

In retrieval and reconstruction processes, domain adaptation and transfer learning have been brought into the spotlight. To enhance the portability across basins, techniques such as fine-tuning, feature alignment, self-supervised pre-training, and region-aware modeling can be applied. This portability is not an arbitrary academic aspect of digital twins; it defines whether a technique can be implemented in several jurisdictions of operational activities and be resilient to changing conditions.

3.5. Quality Control and Uncertainty Characterization as Operational Necessities

An automated, continuous quality control is a major distinction between research pipelines and operational digital twins. The causes of degradation in RS products can be clouds, haze, sensor noise, geometric distortions, or changes in upstream processing. In the event that only manual detection is implemented, these failures are likely to be passed on silently through the twin and present inaccurate state updates and forecasts. Definitions Quality control of literature includes rule-based checks (range tests, consistency checks, cloud masks) and learning based anomaly detection that may either warn about unusual patterns in comparison to past distributions or physical constraints. Quality control (QC) most effectively works in a twin environment, which results in actionable outputs; that is, reliability scores, or details gauging uncertainty inflation, or coefficients of individual weights of sources, which can be utilized by downstream fusion and assimilation layers [55].

Characterization of uncertainty is also grounded. Operational choices must include, but not be limited to, quality estimates that have awareness of observational, notably retrieval, ambiguity, scale between the model and the point of measurement, and scale of models. Uncertainty in RS-driven pipelines is frequently like that encountered in aleatoric measurements (noise and variability due to the measurement instrument), as well as those that are epistemic (model uncertainty, domain shift). Practical strategies would involve ensembles, Bayesian approximations, and calibration strategies that match predicted uncertainty with apparent errors. Notably, the uncertainty should be transmitted instead of being reinstated: in case downscaling or gap filling is done, the uncertainty should tend to widen, with the exception that other independent evidence justifies confidence. In the case of uncertainty that is well modeled, a digital twin can be conservative, e.g., generating risk-sensitive warnings or keeping buffers safe in the reservoir, as opposed to making fragile decisions on the basis of overconfident point estimates [56–58].

3.6. Summary: Implications for Digital Twin Design

Central, multi-source RS makes a variety of water-system states and drivers more visible; however, it also presents heterogeneity and uncertainty, which will need an explicit management approach. There is an apparent change in the literature to multi-modal observation stacks, as well as motionless use of commodities to AI-assisted retrieval, downscaling, as well as rebuilding, which has the capacity to decide more spread and decision-adjusted inputs. Simultaneously, operations preparedness rests on the much more mundane elements: harmonization pipelines, provenance tracking, quality control, and due process. These factors conclude whether RS-obtained inputs can stabilize the twin and make decisions better, or provide some concealed biases and false accuracy that worsen the performance in situations when management with real-time response is most important [59].

These base preconditions the following part, in which the emphasis is shifted to the end-to-end digital twin architecture: the fusing, assimilating, and coupling of multi-source RS products and in situ data with physics-based and AI models to coordinate state, make predictions, and assist in making sound decisions within the time constraints of real-time environments.

4. AI-Enabled Digital Twin Framework: Fusion, Synchronization, and Prediction

The operationally meaningful digital twin of the water resource is possible only once the heterogeneous observations are transformable into coherent system states, which are updated in time and projected into the future to make decisions. Section 3 explained how information can be extracted using multi-source remote sensing (RS) and AI; the point of this section is how such streams of information will be incorporated into a complete end-to-end system that ensures a synchronized process and generates actionable forecasts. Literature is united by a collection of repeated architectural units, these being data ingestion and harmonization, multi-source fusion and state estimation, forecasting and scenario generation, uncertainty quantification (UQ), and system feedback, along with monitoring and governance. Nonetheless, there is a wide divergence in implementations in terms of coupling strategies (AI-only, physics only, hybrid), and the way they respond to asynchronous arrival of data, and whether they are fidelity or performance-bounded. In this section, these patterns are synthesized and emphasize the trade-offs associated with the methodology that makes or breaks the digital twin to be reliable in real-time [60].

A helpful point of view is to view the twin as a dynamical estimator-predictor system, whereby the state of the system is constantly corrected by evidence. The state can be taken to mean the wetness of the land surface, snow, the hydraulics of the channels, the storage capacity of the reservoirs, or indicators of water quality, depending on which management task needs to be addressed. The twin then has to solve two interrelated issues (i) discretization that creates the current most accurate approximation to the state when we are given observations and model structure, and (ii) prediction, which provides the dynamics of the state with future forcings and control actions consistent with physically realistic dynamics, and uncertainty. Multi-source RS can be used to achieve better observability, at the cost of scale, noise, and speed; AI can be used to provide potent mapping, fusion, and acceleration, but is vulnerable to distribution change. These tensions should be resolved by the framework through direct provisions on how to ensure quality-conscious fusion with purposive uncertainty and fallback [61].

4.1. Reference Architecture for an AI-Enabled Water Digital Twin

Digital twins are conceptualized in a variety of areas as layered cyber-physical systems where data, models, and operations are combined to simulate and monitor real-world systems. A realistic and operational reference architecture can be depicted in five interrelated layers in the context of water resources, with a cross-cutting operations layer. This organization is such that all the required elements, such as data collection to decision support, operate in harmony to deliver a full, real-time picture of the water system. To relate methods with pipeline functions, **Table 3** assigns significant AI roles, retrieval, downscaling, fusion, synchronization, forecasting, and decision support to common inputs, outputs, strengths, and operating risk [62].

Table 3. Taxonomy of AI roles and methods across the digital twin pipeline.

Twin Pipeline Function	Typical AI Methods (Examples)	Inputs	Outputs	Strengths	Risks/Safeguards
Retrieval (RS → hydrologic variables)	Convolutional neural network (CNN)/U-Net/Transformers; tree-based ML; physics-guided learning	Radiance/backscatter/brightness temperature (BT) + metadata	Soil moisture/ET/inundation/indices + uncertainty	Learns complex nonlinear mappings	Domain shift → add calibration, constraints, QC
Downscaling/super-resolution	SR networks; cross-modality guidance; spatiotemporal models	Coarse products + high-res covariates	Fine-scale fields + uncertainty	Bridges scale mismatch	False precision → uncertainty inflation, conservative smoothing
Gap filling/reconstruction	Temporal attention; diffusion/sequence models; masked modeling	Multi-time observations + covariates	Completed time series/fields + reliability	Improves continuity under missing data	Hallucination → masks, confidence gating, fallback rules
Multi-source fusion	Cross-attention fusion; gating; graph neural nets	Multi-modal features + quality flags	Fused state proxies + uncertainty	Robustness to partial modalities (if designed)	Brittle missingness → modality dropout training, masks
State estimation/filtering	AI-accelerated data assimilation (DA); learned observation operators; learned filters	Observations + model prior	Posterior state distribution	Faster updates; better use of heterogeneous data	Filter divergence → QC, conservative priors, monitoring
Forecasting/emulation	Surrogates; neural operators; hybrid residual learning	State + forcings + controls	Probabilistic forecasts	Speed, ensemble scalability	Nonstationarity → drift detection, retraining triggers
Decision support	Risk-aware optimization; constrained reinforcement learning (RL)/model predictive control (MPC) hybrids	Forecast distributions + constraints	Recommended actions/risk metrics	Decision-aligned outputs	Safety/legality → constraints, human-in-loop, audit trails

The first layer will deal with data ingestion and cataloging. It presupposes the aggregation of a significant diversity of data, such as remote sensing products, in situ and Internet of Things (IoT) streams, meteorological predictions, and infrastructure telemetry (e.g., gate positions, release records). There are also static geospatial data, including land cover, soils, bathymetry proxy, and river network. This is a layer that handles metadata, time stamps, lineage, and quality flags such that streams of data, typically event-driven or continuous, are well managed. Special consideration of intermittent data feeds is done in real-time systems, and in such systems, streaming capabilities with strength are needed [63].

Then, the data processing and uncertainty tagging play important roles in making sure that the data obtained is consistent with the state variables and the update rate of the digital twin. This layer deals with retrieving, harmonizing, and quality assuring the data. It also generates reliability scores and uncertainty estimates that are vital in assessing the correctness of the data that is further used to create the layers, especially in the decision-making process. The content and reliability of the data, and its ambiguity, are vital to the dependable functioning and prediction. The main element that coordinates the data and models is the fusion and state estimation layer. Through a combination of multiple sources of observations and the application of sophisticated models, including data assimilation or learned fusion models, this layer provides an update to the internal state of the digital twin and has an approximation of the present state of the system and a representation of uncertainty associated with it. Such synchronization creates a digital twin that is a realistic reflection of the real-life system, which can be forecasted and analyzed in a meaningful way [64].

Forecasting and scenario generation are based upon synchronization of state and extrapolating it under the meteorological and operational environments. This layer is able to use physics-based simulators, AI surrogates, or hybrid models, and generate probabilistic predictions that can be used to make decisions with a range of lead times. Such predictions are crucial in the future system behavior, which is crucial in managing the resources, eliminating the risks, and making interventions. Lastly, the decision support and interaction layer converts the forecast distributions and the existing state into actionable insights. This involves the creation of alerts, establishment of risk limits, and recommendations for the reservoir release, irrigation timetable, or resource allocation. An important component of this layer is the human in the loop review, which can undergo further elaboration, auditing, and validation of the recommendations, which will ensure that decisions are made confidently and transparently [65].

The operations layer that cuts across the whole system aims at sustaining the performance of the digital twin and making sure that it is robust and up-to-date. Tracking model drift, version control of both the model and the data products, initiations of recalibration or retraining of the model, and access control are also part of this layer. This layer also has safety mechanisms to ensure reliability in extreme conditions; thus, this layer is used in high-consequence applications, which include fallback modes, override mechanisms, and strict change-management processes.

The main peculiarity of this architecture is that it explicitly reflects the presence of uncertainty and makes monitoring and governance mechanisms one of the main components of the system. In contrast with most research-based pipelines, which can look at uncertainty and governance as secondary concerns, this architecture makes them central to its functioning. The structure provides a stable system of categorizing research contributions: some deal with the processing of observations, some with state estimation or prediction, and some with decision support. A fully realized digital twin combines these elements in a seamless way, and thus, they ought to be integrated to generate dependable, actionable results. In addition, the added advantage is that there are reliability mechanisms in its entirety [66].

4.2. Multi-Source Fusion Strategies: Aligning Modalities, Scales, and Reliability

The fundamental technical issue in creating an RS-based twin is that information is received through a variety of modalities at varying spatial support, temporal cadence, and error structure. The literature generally recognizes fusion strategies as data-level (early) fusion, feature-level (intermediate) fusion, and decision-level (late) fusion with different implications for interpretability and robustness.

In early fusion, observations are combined into a single piece of data before modeling—e.g., the subsampling of several products of RS onto one common grid, and the application of the variables. This method is simple, has a tendency to be effective in predictive modeling, and it may conceal sensor-induced uncertainties, and it may infer performance when resampling induced artifacts. At the feature level, modality-specific encoders (e.g., obtaining networks to represent SAR and optical) are maintained, and the learned representations are combined into one via attention, gating, or cross-modal interaction. This strategy is more resistant to missing modalities, and it is able to include reliability weights. Late fusion takes the results of two or more individual models (e.g., two independent inundation maps or independent estimates of soil moisture), inputting ensemble rules or learned mixture weights, which is not as complex as cross-modal deep features and provides modularity and more readily debuggable combinations, though it may occasionally miss the cross-modal learners.

Water-use Feature-level and late fusion is becoming popular where robustness of operation is needed, such as during a cloud cover or partial sensor outage. Reliability-conscious fusion is one such motif; fusion modules apply quality flags as well as observation geometry together with uncertainty estimates, to weight the sources adaptively. Considering the example, the twin can give the SAR a higher priority in detecting the extent of floods during cloudy storms and optical signals during clear conditions to narrow down the boundaries and obtain water-quality proxies. Equally, the impact of coarse microwave soil moisture can offer constraints on the scale of the basin, but can be mitigated using downscaling uncertainty or situations where the land cover implies retrieval instabilities [67,68].

Structural alignment with hydrologic topology is another dimension of fusion that is important. Naturally represented as graphs are river networks, reservoir reach systems, and irrigation canals, and spatial consistency can be enforced along flow paths and connectivity of infrastructure by using graphs. This is specifically applicable to twins who leave the pixel maps behind and enter the realm of system-level state estimation, where the idea is not to find out where water is but how water moves and can be stored in managed infrastructure.

4.3. Synchronization through Data Assimilation and Learned State Estimation

The process through which a digital twin keeps in line with reality is through synchronization. In hydrology and hydraulics, the concept of synchronization can be followed by the term data assimilation (DA), whereby model states, or parameters, are reconciled by observations. DA provides a principled framework for integrating model predictions and observations in the presence of uncertainty; however, classical DA may be computationally expensive and may fail when observation operators are complicated, errors are non-Gaussian, and state-observation maps are very nonlinear [43].

In water twins due to RS, DA has two questions in practice, namely, what to assimilate and how to state uncertainty. It is often important to assimilate derived products of RS (e.g., products that are symptomatic and independent of hydrologic variables, e.g., soil moisture estimate), as it makes products easier to map to hydrologic variables, though bias in products and errors often go together. Assimilation is able to store information and minimize the use of upstream retrieval assumptions, but needs correct observation operators and large domain models. A hybrid is commonly used in many operationally oriented studies, with derived products integrated and biased by the correction of uncertainties and by inflating the uncertainty where a particular operator of observation may be well understood or can be learnt.

AI helps in synchronization in many ways. Operators of learned observation may be used to provide an approximation of the mapping between state variables and predicted sensor observations, so that assimilation may be used in cases where explicit physical operators are not available, or are too slow. AI surrogates offer several advantages compared to mainstream ensemble methods, namely, these surrogates are able to achieve higher propagation rates in ensemble procedures and therefore can implement larger ensembles or more dynamic updates on limited budgets through AI. Representation learning is able to give low-dimensional latent states, which are simpler to update and are able to diminish filter degeneracy in high-dimensional contexts. But learned synchronization gives rise to novel dangers: in case an operator learned is biased or cannot operate in a new domain, assimilation can inadvertently draw the twin in the wrong direction with an undeservedly high level of confidence. As a result, effective methods emphasize conservative uncertainty, intensive quality gating, and periodic validation of independent measurements [69].

An independent, closely associated body of work considers synchronization as a learned filtering problem, in which recurrent or attention-based models consume state estimates as the direct output of the input sequences of observations. These methods may be computationally efficient and may inherently be able to deal with asynchronous inputs, but they may need large quantities of training data that contain representative extremes and may find it hard to extrapolate when system dynamics vary. Purely learned filters can also be non-transparent, where infrastructure operations and rule changes are of interest. Due to these reasons, learned components are often coupled with physics structures or DA frameworks instead of being them with more deployable and entirely different patterns [70].

4.4. Forecasting Engines: Physics Models, AI Surrogates, and Hybrid Approaches

After the synchronization, a digital twin needs to produce predictions and situations that can be used when making decisions. The literature is overtaken by three paradigms of forecasting: physics-based simulation, AI-based forecasting, and hybrid modeling.

The physics-based models maintain their important role as a result of being a representation of conservation laws, routing dynamics, and constriction of infrastructure. They are especially useful in extrapolation to history, as well as in analysis of scenarios under new conditions. Nonetheless, the results of high-fidelity models may be sluggish, responding to uncertain variables, and difficult to support over areas and evolving scenery. Their real-time performance relies on forcing quality (e.g., precipitation forecasts) as well as intense state correction by assimilation.

AI forecasting is fast and can represent nonlinear relationships with complexities, particularly when it is trained on large datasets with multiple regimes. To give an example, deep sequence models can be trained to learn rainfall-runoff dynamics, and convolutional or attention-based models can be trained to learn spatiotemporal dynamics pertaining to inundation development. The key benefit in a twin-based scenario is that AI can be able to offer fast inference, which can be used in real-time updates and massive ensembles. The prime weaknesses include vulnerability to distribution shift and the extreme possibility of overconfidence in predictions.

Hybrid solutions are trying to use the strong points of both. Common techniques are the residual learning (AI is trained to learn the systematic error of a physics model), physics-guided learning (consideration of conservation or monotonic constraints is put on the training losses), and natural hybrids (physics can perform routing when AI is expected to estimate challenging components like effective rainfall, ET, or parameter modifications). Particularly effective is hybridization, in which physics is good at giving the right structural relationships, but is biased in its inputs or simplified models of processes, and AI can correct the biases without necessarily learning the dynamics afresh. In the case of operational twins, hybrids also have a pragmatic governance benefit; the physics core is availed of a clear baseline, and AI modules can be checked and can be limited as a refinement instead of a complete replacement [71–73].

Another required capability is scenario generation. The plausible forcings of the future (meteorological ensembles), the alternative actions of the operation (different release timetables), and the uncertainty of the initial conditions are often involved in operational choices. A digital twin consequently enjoys rapid, probabilistic scenario generation that can be facilitated by AI surrogates that execute simulation more rapidly, or probabilistic forecasting models that generate distributions (not point estimates). As risk-based decisions are usually involved, the quality of such distributions can be even more important than their mean accuracy.

4.5. Uncertainty Quantification and Propagation from Observations to Decisions

The linking agent on an operational digital twin is uncertainty: taking bad measurements and translating them into state approximations, predictions, and finally decisions. Measurement noise, retrieval ambiguity, cloud and sampling gaps, scale mismatch, model structural error, and parameter and forcing uncertainty all bring uncertainty in RS-driven systems. A powerful framework must thus have three levels of uncertainty representation: observation uncertainty, state uncertainty, and forecast uncertainty, with clear propagation throughout the pipeline.

The literature employs a number of working strategies. Ensemble methods are still widely used since they offer a more natural means of modeling uncertainty in DA and forecasting. Epistemic uncertainty in AI modules can be represented by Bayesian approximations and deep ensembles, and aleatoric uncertainty in retrieval and downscaling can be represented by heteroscedastic models. The use of calibration techniques is becoming common to make sure that the expected intervals match the observed frequency of errors, which is essential to the credibility of risk communication. Distribution-free approaches and conformal prediction have also been of interest to produce cover guarantees under assumptions, but have yet to be successfully used in the practice of water [74].

In the digital twin perspective, uncertainty has to be decision-based. Operators require the uncertainty in the form that can be acted upon: the excess likelihood of floods that exceed the facility, or a confidence interval around the forecasts of the inflows in the reservoir, or some risk measure concerning alternative facilities that could be used to avert the event. This also reflects the need not to have false precision in downscaled or gap-filled products. In the case of weak evidence, one should put more distance between the evidence, and decision modules need to be skewed towards conservative decision-making approaches as opposed to acting on slicing point estimates.

4.6. Deployment Patterns: Streaming Updates, Edge–Cloud Execution, and Operational Guardrails

One common deficiency in the demonstrations conducted in research and in actual twins is the absence of unequivocal deployment patterns. In real-life settings, streaming architectures are progressively used in streams with ingestion updates found in shoulder calls. Indicatively, an inundation update and a localized state correction can be caused by a new SAR scene; uncertainty inflation and falling back to model predictions can be caused by a gauge outage; a batch of ensemble runs can be triggered by a new meteorological forecast cycle. Then, event-driven design minimizes unnecessary calculation and increases the responsiveness [75].

Also applicable is edge cloud partitioning. There are inference problems that can be executed nearer to their data sources, such as basic quality assurance/quality control (QA/QC), lightweight segmentation, anomaly detection, and heavy ensemble simulation and multi-sensor mosaicking can be left on the cloud. Irrespective of where it is physically located, operational twins need well-developed versioning and monitoring: data products evolve, models change, and system performance may shift. Its applications, such as model registries, automated testing, performance dashboards, rollback procedures, etc., are mandated by safety-critical/high-stakes applications [76].

Barriers to operation are also crucial. These are conservative fallback modes (e.g., the use of physics-only forecasts when AI uncertainty peaks), explicit steps of human approval of high-impact ones, and audit trails to trace the decisions to their underlying data and model conditions. The degree to which automated control will be acceptable will be dependent on governance needs and institutional risk tolerance in many jurisdictions. As a result, one of the most frequent short-term trends is in decision support and not complete autonomy: the twin generates probabilistic suggestions and explanations that are reviewed and executed according to the current processes by operators.

4.7. Synthesis: Coupling Strategies and When They Are Appropriate

There are three strategies of coupling that can be seen throughout the literature. Artificial intelligence (AI) only twins focus on state estimation and prediction using learning data, and may perform very well when using a history and their runtime is limited, but also, they demand large volumes of training data and drift control. Physics-oriented twins apply RS mainly to assimilation and calibration, which is interpretable and extrapolates better; however, latency and forcing uncertainty can pose a challenge to them. Hybrid twins combine the physical structure with AI modules of retrieval, bias correction, acceleration, or residual learning, and they are likely to provide the best channel towards operational robustness with uncertainty-independent fusion and monitoring.

The correct strategy will be based on the criticality of the decision, the availability of data, and operational constraints. Assimilation-enhanced physics or hybrids can be most appropriate where there are good-quality physics models and telemetry (e.g., managed reservoirs). In model-constrained data-rich settings (e.g., flood mapping in cities), AI can take over observation-to-state inference with physics being applied where it is needed (e.g., routing, plausibility). The RS scale limitation of coarse storage in large drought monitoring can be well served by the multi-scale hybrids, which relate basin-level signals to the finer local indicators using learned mappings with learnt uncertainty [77].

The literature, on the whole, points to a very concise conclusion: it does not matter what particular model people select, but rather how internally coherent the end-to-end system is, quality-aware fusion, Sy vision that leaves uncertainty in check, latency-appropriate forecasts, certain dissimilarity maintains trust under Nihil bis una data: operation guardians. These lessons precondition the following section, which generalizes the way the framework is implemented in key areas of application and how performance is to be measured, compared, and contrasted in a manner that is sensitive to operational realities and is not merely retrospective[76].

5. Applications, Evaluation, and Benchmarking

The usefulness of an AI-powered digital twin depends on its practical implementation in real decisions in real-time, and ultimately, that is the benchmark of its usefulness. This section encapsulates the literature instantiations of the framework in the key applications to water management and, most importantly, the performance measurement process. One consistent conclusion in the body of research is that high skill on models in retrospective tests does not necessarily translate to operational advantage unless latency, robustness, calibration of uncertainty, and impact on decision-making have been measured directly. To this end, this section arranges applications based on the decision task, and then summarizes evaluation practices into a framework of benchmarkable criteria in the entire pipeline, from remote sensing retrieval and fusion all the way to state estimation, forecasting, and decision support. A retrieval performance-to-state synchronization-to-forecasting-skill-to-decision-outcome end-to-end assessment ladder has been summarized in **Figure 3**. **Table 4** gives a brief checklist of what needs to be reported by each evaluation layer, robustness tests, and runtime[60,78].

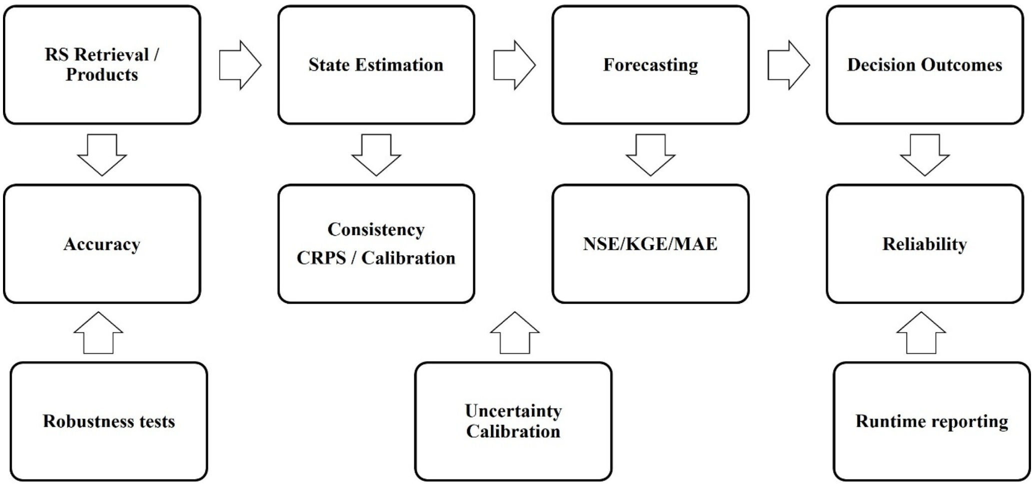

Figure 3. Evaluation and benchmarking framework.

Table 4. End-to-end evaluation and benchmarking checklist.

Evaluation Layer	What to Report	Example Metrics/Tests	Why It Matters Operationally
Observation products	Accuracy + availability + latency	Root mean square error (RMSE)/mean absolute error (MAE/bias); intersection over union (IoU)/F1 (extent); uptime; delivery delay	Inputs drive all downstream performance
Synchronization (state estimation)	Consistency + calibration	Innovation stats; continuous ranked probability score (CRPS); spread–skill; independent validation	Prevents drift; supports trustworthy current state
Forecasting	Lead-time skill + extremes	Nash–Sutcliffe efficiency (NSE)/Kling–Gupta efficiency (KGE)/MAE; peak timing error; tail exceedance skill	Decisions depend on *when* and *how bad* extremes are
Uncertainty	Calibration + sharpness	Coverage tests; reliability diagrams; CRPS	Avoids overconfident actions; enables risk-based control
Robustness	Missing data + domain shift	Modality dropout; cloud/outage stress tests; cross-basin transfer	Real-time systems must degrade gracefully
Runtime/scalability	End-to-end latency + throughput	Wall-clock time per update; compute cost per km^2/basin	Determines feasibility for real-time operation
Decision impact	Outcome-based performance	False alarm/miss rates; constraint violations; reliability; regret	Measures real operational value
Reproducibility	Provenance + versions + splits	Product versions; preprocessing details; code/data availability	Enables fair comparison and safe deployment

5.1. Application Patterns and Digital Twin Instantiations

Even though implementations differ, the majority of published water digital twin systems can be categorized into a few archetypes of applications with different prevalent state variables, dependence on observation, lead times required, and risk acceptance.

One of the most popular digital twin applications is flood monitoring and early warning, which directly benefits since near-real-time RS benefits inundation, whereas fast predictions are of use to support alerts. This archetype involves the twin, utilizing the paleo transport, sharing precipitation observations and forecasts with hydrology, hydraulic routing, and interface RS (optical/SAR), to share and reconcile inundation areas, correct hydraulic states, or audit risk maps. In the urban setting, including complicated hydraulics and prohibitive computing, AI-friendly inundation mapping and speedy surrogates to forecast rapid hazard products are frequently employed, and physics is employed to proffer checks of consistency or course routing. One critical operation needed is sensitivity to events: the system needs to possess the capability to deal with cloud cover, promptly move the water level, and nonstationary environments like litter creation, blockages, or infrastructure breakdowns [79].

A second archetype is reservoir and multi-reservoir operations, where the state is highly dependent on infrastructure restrictions and management efforts. In this case, RS helps by limiting upstream snow and soil moisture (as inflow precursors), monitoring reservoir level or surface area, and giving information on drought stress and evapotranspiration demand. The application of digital twins tends to focus on physics models that are augmented with assimilation and release decisions that are optimized. AI can be introduced as an inflow forecasting part, as a substitute used to find quick ensemble assessments, or as a part that receives a manner of learning biases in contrast to physics-based forecasts. The outputs in most decisions are usually multi-objective, balancing flood control, reliability of supply of floods, hydropower, and environmental flow, where there are constrained and defined limits and domination margin [80].

A third archetype is drought monitoring and water allocation, of which the spatial scale is large, and the states of interest are soil moisture anomalies, deficits in evapotranspiration, and storage anomalies. Multi-source RS can be central often as in situ networks are skinny at basin scales, and that signals of drought are spatially anisotropic. In this area, the focus is often on digital twins, enhanced estimations, and uncertainty signs over high-frequency control. The lead times are longer (weeks to months), but the reliability and consistency are paramount since the decisions may have an economic and regulatory impact. Storage limits at the grit level of gravimetry, ET products, and land surface anomaly of temperature are often amalgamated with ground information wherever possible to generate drought status and indicators that are allocation relevant.

Fourth archetype: Irrigation and agricultural water management. This requires fine spatial resolution and field-level relevance. In this case, the core states include ET estimation, crop stress proxies, and soil moisture dynamics. Multisource RS can be applied to monitor large areas of agriculture, whereas AI may be utilized in downscaling, gap filling, and model crop-type-aware ET. In this area, digital twins were often used as decision support systems, which suggest when and how much to irrigate, and they may combine farm-level telemetry where that is available. One problem is that of mismatch of scales: operational choices are made at the farm or field scale, whereas most RS products and models are made at finer resolutions and with models whose latency may not be trivial [76].

Another archetype that is nearly a newly developed one is water quality monitoring and incident response, in which RS is used to give proxies of turbidity, chlorophyll-related indicators, or surface temperature, or other signals, which can be correlated with water-quality risks under appropriate conditions. In this case, digital twins can be a combination of RS-generated surface indicators and hydrodynamic models, or AI anomaly classifiers. Water-quality twins, in comparison to those based on quantities, are associated with complex relationships between observations and state and greater reliance on local calibration and in situ validation. Operational value is associated with early detection, false alarm management, and explainability, especially in times when decisions related to the population health messages or treatment process need to be made.

Through easing these archetypes, the literature immunizes that the most effective digital twin deployments are those that (i) characterize a minimal, apparent operational state congruent to choices (ii) information sources multi-source RS with explicit quality control and uncertainty, and (iii) produce outputs launchable by the type of operators that may act on (potentialities, risk classes, constraint-sensitive promotion)[81,82].

5.2. Evaluation across the Pipeline: What Should Be Measured and Why

The big problem in the field is that evaluation has been subject to isolation: papers in the RS retrieval literature declare pixel-based metrics, papers in hydrologic forecasting declare discharge skill, papers in decision-support declare optimization goals, but few studies have bridged the gap between these metrics into an end-to-end evaluation. In the case of an SCI review, there must be a coherent evaluation framework that grades four levels, including observation products, state estimation, forecasting, and decision outcomes.

At the observation-product level, the evaluation is assessed in retrieval accuracy and spatial detection skill. RMSE/MAE/bias (continuous): use on continuous responses (e.g., soil moisture, ET, proxies of water levels, etc.) and intersects-over-union (IoU), F1-score, or similar measures of success at segmentation (surface water extent map, flood inundation maps). Nevertheless, in a twin framework, error structure is also important to consider, including spatial correlation, heteroscedasticity between land cover, and systematic biases by season or region, since such characteristics directly influence fusion and assimilation behavior. To achieve operational preparedness, performance metrics should be reported like latency, rates of data availability, and failure conditions (e.g., what fraction of the cloud cover is required to have optical products to become unreliable)[83].

The question on the state-estimation level to be answered is: Is the twin sufficient to deliver a credible and consistent estimate of the state using imperfect and asynchronous measurements? For DA-based methods, diagnostics include innovation statistics (observation-minus-forecast), the consistency of ensemble spread and observed errors, and probabilistic scores like the continuous ranked probability score (CRPS). Amongst learned state estimators, the analogous assessment can be carried out by comparing estimated states to independent reference observations, and by checking calibration: an estimate that is operationally preferred (albeit slightly biased) but severely calibrated may be useful in comparison with the apparently accurate estimated states, which tend to become overconfident in the case of data gaps.

Evaluation needs to be stabilized at the forecasting level with respect to lead time and decision relevance. Nash Sutcliffe efficiency (NSE), Kling Gupta efficiency (KGE), bias, peak timing error, and threshold exceedance detection are common measures of streamflow forecasting and water level and inundation forecasting might favor event detection and measures of spatial overlap. In the case of reservoir inflow prognostication, high inflow exceedance probabilities (e.g., at critical quantiles) tend to be more critical than average error. One of the main suggestions that has been seen in literature is the assessment of extremes directly: the magnitude of the peak, the time of its appearance, and the probability of the tails are to be reported independently of bulk performance. Besides, the evaluation of the forecasts ought to include meteorological forcing uncertainty, initial-condition uncertainty, particularly in cases of risk-based actions of the twin[84].

At the decision and outcome level, the level of evaluation should encompass the enhancement of operational goals under constraints by the twin. In the case of flood warning, this can be a combination of false alarm rate, missed event rate, and lead time enhancement at a certain hazard threshold. In the case of reservoir operations, trade-off metrics can be the reduction of flood risks, supply reliability, hydropower generation, ecological flow compliance, and constraint violations, and are commonly used in multi-objective trade-offs. In the case of irrigation, measures can be the yield proxies, water use efficiency metrics, and observation of the system capacity limitations. The most significant thing is that the value of a digital twin cannot be determined with predictive skill per se, but it should be measured by its adaptability in terms of altering actions and whether those actions enhance performance under realistic limitations[85].

5.3. Benchmarking Protocols: Beyond Retrospective Accuracy

Digital twin benchmarking needs protocols that are indicative of operational reality. The literature is becoming more aware of a number of protocol elements that are required to make a credible comparison. To begin with, there should be cross-region and cross-regime generalization tests. The fact that different basins have different land cover, climate, infrastructure, or sensor conditions may cause the models to fail. Benchmarking needs to incorporate, therefore, transfer tests (train in one set of basins, test in others) and seasonal shift tests (train in normal seasons, test in non-normal conditions). In the case of RS retrieval and downscaling, the benchmarks are supposed to stratify performance with respect to cloud conditions, vegetation density, terrain complexity, and acquisition geometry since these variables are the most influential in error behavior [76].

Second, missing data and asynchronous sensing stress tests are essential. Digital twins should operate in cases where optical imagery is lost over a prolonged time, SAR coverage is not continuous, or in situ sensors are broken down. Organized modality dropout experiments ought to be incorporated into benchmark protocols and ought to report the expansion of uncertainty and change to decision outputs. A system with kept uncertainty and graceful degradation is operationally better than a system with sharp and unreliable output.

Third, the official reports should be on latency and compute budgets. An algorithm raising proficiency by a small percentage and taking hours of computation cannot be of use in responding swiftly to a flash flood. On the other hand, a moderately inaccurate process that can be executed in several minutes can present greater operational value. End-to-end runtime, update frequency, and throughput at representative spatial scales should therefore be reported in benchmarking, allowing for preprocessing costs [86].

Fourth, uncertainty calibration must be a benchmark of the first class. Point metrics should be accompanied by reliability diagrams, prediction interval coverage tests, CRPS, and other probabilistic measures. Uncalibrated uncertainty is likely to lead to brittle automation and overconfident decisions by digital twins. Lastly, there is a weakness of reproducibility. Numerous research works are based on in-house data, area-specific calibrations, or unregistered preprocessing. The best benchmarking practice comprises both provenance of data and versioning of RS products, (training and validation) splits, and, when practical, the release of code or executable pipelines [87].

5.4. Common Failure Modes and Mitigation Strategies Observed in Practice

The literature recognizes common failure modes that may compromise end-to-end digital twin performance, even if the performance may be good on the part of individual components. False precision. One typical failure is downscaling or gap-filling. Assimilation and decision modules may believe that the artifacts are valid signals, and when the coarse RS products are down-sampled back to the field or neighborhood scale, the failure to use sufficient uncertainty inflation may result in unstable inputs or excessive confidence in control. Mitigation is contained in uncertain down-smoothing, conservative-smoothing, and explicit reliability gating.

A second failure mode is the bias drift because of the variation in the RS processing chains, sensor recalibration, or a non-stationary environment. Provided that the product versions change without being noticed, then the twin will consider a processing change as a hydrologic trend. Product version monitoring, drift, as well as periodic re-benchmarking against a stable reference, can be placed as a mitigation requirement [88].

The third failure mode is incompetence, when in extreme cases, since the training data is biased against rare events. This is particularly urgent in the case of AI-based flood mapping and the prediction of rainfall-runoff. Some mitigation strategies are event-oriented training, augmentation, physics-guided constraints, and hybrid structures, which do not lose physically plausible dynamics. The fourth failure is discontinuous fusion among absent modalities. Fusion models that are known to operate under the assumption that all the inputs are not absent can fail when an important sensor is not available, and fallback pathways make minimal sense when dropped into robust designs, since modality-aware architectures and dropout training are central to learning [76].

A fifth failure is a poorly matched evaluation objective. The systems that are optimized in pixel-level accuracy might not yield any better decision results, and the ones that are optimized in the mean skill in the decision might fail to detect the critical tails. Mitigation will be to match training and assessment to decision thresholds, tail risks, and constraint violation sanctions [89].

5.5. Toward a Standardized Evaluation Checklist for Water Digital Twins

The research on digital twins in different applications and measurement approaches also draws attention to some of the most significant criteria to evaluate the maturity of such systems. These requirements provide a rich system of criteria to assess the effectiveness and operational preparedness of digital twins in various fields, which offers a place of uniformity in the comparison of various systems. An important aspect of such a framework is decision alignment, which makes sure that the capabilities of the digital twin are cleanly aligned with the tasks at hand, the decision-making frequency, and the lead times to provide effective intervention. Data realism is also of the essence, which concerns reporting of important aspects like the latency of remote sensing (RS) data, availability rates in data streams, effects of cloud cover on data accuracy, and data outages of in situ sensors management. The information is useful in evaluating the system reliability in dynamic environments [90,91].

End-to-end metrics are also evaluated in the digital twin's evaluation since retrieval, state estimation, forecasting, and decision outcomes are jointly evaluated. With this holistic approach, a more comprehensible understanding can be made of how well the system has been functioning at all the stages of the operation of the system. Also, another crucial criterion is the extreme-event performance, which focuses on the fact that metrics involving rare or high-impact events require explicit metrics that are crucial in addition to the analysis of system behavior in these cases.

It is also important to calibrate uncertainty because it makes sure that the probabilistic outputs of the system are strictly tested, and the coverage is properly reported, giving an idea of the confidence in the predictions. Robustness tests, additionally, improve the process of assessment by making the system resilient to a variety of challenging circumstances, including data modalities going dead, domain shifts, or model drift. To know the efficiency of the system, runtime reporting, which also involves monitoring the end-to-end latency and the computing resources that are needed, is required. Another important factor is reproducibility, which necessitates data provenance transparency, product versions, preprocessing procedures, and code availability, where possible [92].

These standards offer an organized method for digital twin systems comparison and the transfer of research information to practical developments to apply in practice. There is, however, a big gap in the field whereby there are no single common shared benchmark data sets and protocols, which cross multi-source remote sensing, state estimation, and decision evaluation. Sealing this gap is necessary both to speed up the creation of digital twins and to guarantee that these systems have a positive contribution to enhancing the management of water systems in the face of reality. Critical assessment of this challenge will result in better standardized and transparent assessments, which will eventually contribute to the development of the field.

The foregoing demonstrates that AI-based, RS-based digital twins are being implemented in floods, reservoirs, drought, irrigation, and water quality, and architectural designs are recurrent, and assessment failures are frequent. The soundest evidence of working worth is the research, which quantifies the uncertainty, specifies latency, tests soundness in the absence of data and domain changes, and measures the effect of the decision, as opposed to relying on ex post accuracy. These lessons drive the final part that summarizes the significant lessons learned and a roadmap to the agreed-upon standards of interoperable and trustworthy digital twins to manage water resources in real-time [3,6,93,94].

6. Conclusions

This review has analyzed the new frontier of AI-enabled digital twins to manage real-time water resources with a specific interest in frameworks that incorporate multi-source remote sensing (RS) with in situ measurements and predictive models. The message shown through the literature is that being a worthwhile and efficient product, a digital twin is not a one-dimensional, concreteness or concrete algorithm, or the data output of a system, but the coherence of an end-to-end system that can (i) continue to ensure a constantly synchronized representation of states in water-systems, (ii) generate timely, uncertainty-aware forecasts and scenarios, and (iii) translate these results into actionable decision support under operational constraints. Multi-source RS is able to increase the observability limits that a ground network can achieve, and AI can offer the ability to retrieve, downscale, fuse, accelerate, and optimize. Nevertheless, real-time water management is a high-consequence area, and those aspects that support performance improvements also bring about novel risks, especially false accuracy, fragile generalization in the face of regime changes, and overconfidence in extremes.

A number of synthesis insights come up. First, multimodal integration is not a luxury: no sensor modality has full, timely, and robust coverage of the variety of states of interest in floods, reservoirs, droughts, irrigation, and water quality. The best practices consider RS as an observation stack whose layers are dynamically weighted by quality and availability, and not as layers. Second, the characteristic feature of an actual digital twin is synchronization. However, accomplished by data assimilation or learned filtering (or a combination of the two), the twin has to refresh states with incoming streaming evidence and must so explicitly represent uncertainty and provenance. Third, deployability is increasingly going the way of hybridization. Model-based on physics offers structural correctness, interpretability and the capability to solve scenarios, whereas AI modules offer flexibility and speed when it comes to data-directed retrieval, bias reduction, emulation and robust fusion, especially when combined with assessment of uncertainty and working safeguards.

Evaluation wise, the field is heading past retrospective accuracy to operational readiness and there are still gaps in this domain. Most of the studies still record the performance at remote steps of the pipeline without relating the uncertainty of observation to the state estimation, forecasting and decision outcomes. Extreme performance, behavior in sensor dropout, robustness and calibration of predictive uncertainty are not yet properly evaluated, although these considerations are the key determinants in practice. Similarly, end-to-end latency and computational demands tend to be understated, even though they lie at the heart of real-time implementation. The following shortcomings will have to be improved with benchmark protocols, including explicit support with missing-data stress tests, with cross-region generalization, event-based evaluation and with metrics of decision-level volumes, reliability, risk reduction, and clear reporting of versions of the data products, cross-border preprocessing, and command execution.

Going forward, there are a number of priorities that can help to hasten the pace of reliable and interoperable twins in operation. First would be standardization: shared reference architectures, common state/metadata schema, and standard guidelines on reporting uncertainty, latency, and robustness would help bring comparability and would allow transfer of technology. The second priority is credible AI for water systems, based on calibrated uncertainty, drift watching, defensive failover manners, and in the human-in-the-loop interfaces offering suitable drivers and audit logs that can be comprehended. The third priority is multi-scale modeling and fusion, which would allow coarse-scale storage signals and high-frequency, low-resolution observations on a single scale of local hidden states without bringing in a spurious sense of precision. Lastly, deployment engineering should be included in the research agenda: streaming pipelines, edge cloud-partitioning, and the ModelOps practices (versioning, validation, rollback) can be deemed key to sustainable performance when operating in changing operational situational contexts.

While this review synthesizes a rapidly advancing field, several critical limitations must be acknowledged to chart a path forward. First, the literature is dominated by region-specific case studies, leading to a significant gap in cross-region validation. The performance of retrieval, fusion, and forecasting models is rarely tested across diverse hydroclimatic regimes, geographies, and infrastructures, leaving their true generalization capabilities—and thus their operational reliability in new settings—largely unknown.

Second, the standardization of evaluation, particularly for uncertainty, remains nascent. To address this, the community must move beyond ad-hoc reporting and adopt concrete next steps: 1) Adopt common probabilistic metrics like the Continuous Ranked Probability Score (CRPS) and prediction interval coverage tests as mandatory complements to deterministic scores like RMSE and NSE. 2) Develop shared benchmark datasets that are specifically designed for digital twin evaluation, including paired multi-source RS data, in-situ truth, and pre-defined stress tests for missing data and domain shift. 3) Mandate calibration reporting in all studies, requiring reliability diagrams alongside point forecasts to make uncertainty claims verifiable.

Finally, for AI-driven systems to be truly deployable, governance frameworks must evolve. Future research and pilot projects should focus on developing dynamic governance models that move beyond static approvals. This includes protocols for continuous AI monitoring (detecting drift in real-time), establishing pre-defined fallback protocols (e.g., reverting to a physics-based model when AI uncertainty exceeds a threshold), implementing mandatory human-in-the-loop workflows for high-consequence decisions, and creating transparent audit trails that link every recommendation to its source data and model version. These steps are essential to build the institutional trust required for operational adoption.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No new data were created or generated in this study. As this is a review, it is based on data and information from previously published sources, which are cited in the reference list.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Leonard, M., Westra, S., Phatak, A., et al., 2014. A Compound Event Framework for Understanding Extreme Impacts. Wiley Interdisciplinary Reviews: Climate Change. 5(1), 113–128.

[2] Hlal, M., Baraka Munyaka, J.-C., Chenal, J., et al., 2025. Digital Twin Technology for Urban Flood Risk Management: A Systematic Review of Remote Sensing Applications and Early Warning Systems. Remote Sensing. 17(17), 3104.

[3] Zou, S., Ju, H., Zhang, J., 2025. Water Quality Management in the Age of AI: Applications, Challenges, and Prospects. Water. 17(11), 1641.

[4] Schultz, G.A., Engman, E.T., 2012. Remote Sensing in Hydrology and Water Management. Springer: Berlin/Heidelberg, Germany.

[5] Fekete, B.M., Andreu, A., Argent, R., et al., 2021. Observations, Monitoring and Data Management. In: Bogardi, J.J., Gupta, J., Nandalal, K.D.W., et al. (Eds.). Handbook of Water Resources Management: Discourses, Concepts and Examples. Springer: Berlin, Germany. pp. 385–442.

[6] Gacu, J.G., Monjardin, C.E.F., Mangulabnan, R.G.T., et al., 2025. Artificial Intelligence (AI) in Surface Water Management: A Comprehensive Review of Methods, Applications, and Challenges. Water. 17(11), 1707.

[7] Datta, S., 2022. Digital Twins: Digital Twin Meets Digital Cousin. Available from: https://hdl.handle.net/1721.1/140791 (cited 1 January 2026).

[8] Rai, N., Choi, D., Boyd, N.S., et al., 2025. Advancing Site-Specific Disease and Pest Management in Precision Agriculture: From Reasoning-Driven Foundation Models to Adaptive, Feedback-Based Learning. arXiv preprint. arXiv:2510.24650. DOI: https://doi.org/10.48550/arXiv.2510.24650

[9] Qian, C., Liu, X., Ripley, C., et al., 2022. Digital Twin—Cyber Replica of Physical Things: Architecture, Applications and Future Research Directions. Future Internet. 14(2), 64.

[10] Zhang, K., Qu, T., Zhou, D., et al., 2020. Digital Twin-Based Opti-State Control Method for a Synchronized Production Operation System. Robotics and Computer-Integrated Manufacturing. 63, 101892.

[11] Daneshgar, S., Polesel, F., Borzooei, S., et al., 2024. A Full-Scale Operational Digital Twin for a Water Resource Recovery Facility—A Case Study of Eindhoven Water Resource Recovery Facility. Water Environment Research. 96(3), e11016.

[12] Guo, M., Wu, X., Qi, H., et al., 2025. A Methodology for Sky-Space-Ground Integrated Remote Sensing Monitoring: A Digital Twin Framework for Multi-Source Data-BIM Integration in Residential Quality Monitoring. Journal of Building Engineering. 102, 111976.

[13] Ali, I., Greifeneder, F., Stamenkovic, J., et al., 2015. Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. Remote Sensing. 7(12), 16398–16421.

[14] Schaffer, B.E., Nordbotten, J.M., Rodriguez-Iturbe, I., 2015. Plant Biomass and Soil Moisture Dynamics: Analytical Results. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 471(2183), 20150179.

[15] Gupta, S., Modgil, S., Bhattacharyya, S., et al., 2022. Artificial Intelligence for Decision Support Systems in the Field of Operations Research: Review and Future Scope of Research. Annals of Operations Research. 308(1), 215–274.

[16] Kalusivalingam, A.K., Sharma, A., Patel, N., et al., 2020. Optimizing Decision-Making with AI-Enhanced Support Systems: Leveraging Reinforcement Learning and Bayesian Networks. International Journal of AI and ML. 1(2). Available from: https://www.cognitivecomputingjournal.com/index.php/IJAIML-V1/article/view/59

[17] Bhatt, G.D., Zaveri, J., 2002. The Enabling Role of Decision Support Systems in Organizational Learning. Decision Support Systems. 32(3), 297–309.

[18] Zhao, T., Song, C., Yu, J., et al., 2025. Leveraging Immersive Digital Twins and AI-Driven Decision Support Systems for Sustainable Water Reserves Management:

A Conceptual Framework. Sustainability. 17(8), 3754.

[19] Ghafari, R., Samaei, S.R., 2025. An AI-Driven and Digital Twin-Based Framework for Holistic Risk and Green Project Management in Sustainable Urban Megaprojects: The Case of Tehran under Climate and Population Pressures. In Proceedings of the 18th International Conference on Mechanical, Construction, Industrial, and Civil Engineering, Saint Petersburg, Russia, 21 May 2025.

[20] Duan, H., Cao, Z., Luo, J., et al., 2025. AI-Driven Opportunities and Challenges in Lake Remote Sensing. Information Geography. 1(1), 100014.

[21] Wainwright, D.W., Oates, B.J., Edwards, H.M., et al., 2018. Evidence-Based Information Systems: A New Perspective and a Road Map for Research-Informed Practice. Journal of the Association for Information Systems. 19(11), 1035–1063.

[22] Syed, T.A., Khan, M.Y., Jan, S., et al., 2024. Integrating Digital Twins and Artificial Intelligence Multi-Modal Transformers into Water Resource Management: Overview and Advanced Predictive Framework. AI. 5(4), 1977–2017.

[23] Hipel, K.W., Ben-Haim, Y., 2002. Decision Making in an Uncertain World: Information-Gap Modeling in Water Resources Management. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 29(4), 506–517.

[24] Reed, P.M., Kasprzyk, J., 2009. Water Resources Management: The Myth, the Wicked, and the Future. American Society of Civil Engineers: Reston, VA, USA. pp. 411–413.

[25] VanDerHorn, E., Mahadevan, S., 2021. Digital Twin: Generalization, Characterization and Implementation. Decision Support Systems. 145, 113524.

[26] Keller, M.C., Coventry, W.L., 2005. Quantifying and Addressing Parameter Indeterminacy in the Classical Twin Design. Twin Research and Human Genetics. 8(3), 201–213.

[27] Tao, F., Sui, F., Liu, A., et al., 2019. Digital Twin-Driven Product Design Framework. International Journal of Production Research. 57(12), 3935–3953.

[28] Carlin, H.M., Goodall, P.A., Young, R.I.M., et al., 2024. An Interactive Framework to Support Decision-Making for Digital Twin Design. Journal of Industrial Information Integration. 41, 100639.

[29] Li, X., Li, H., He, Y., et al., 2024. Towards Deterministic-Delay Data Delivery Using Multi-Criteria Routing over Satellite Networks. Electronics. 13(23), 4822. DOI: https://doi.org/10.3390/electronics13234822

[30] Dixit, M.G., Ramesh, S., Dasgupta, P., 2014. Time-Budgeting: A Component Based Development Methodology for Real-Time Embedded Systems. Formal Aspects of Computing. 26(3), 591–621. DOI: https://doi.org/10.1007/s00165-012-0273-0

[31] Najafi, M., Zhang, K., Sadoghi, M., et al., 2017. Hardware Acceleration Landscape for Distributed Real-Time Analytics: Virtues and Limitations. In Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), Atlanta, GA, USA, 5–8 June 2017; pp. 1938–1948. DOI: https://doi.org/10.1109/ICDCS.2017.194

[32] Manure, A., Bengani, S., S, S., 2023. Robustness and Reliability. In Introduction to Responsible AI: Implement Ethical AI Using Python. Apress: Berkeley, CA, USA. pp. 133–158. DOI: https://doi.org/10.1007/978-1-4842-9982-1_5

[33] Johannesson, P., Bergman, B., Svensson, T., et al., 2013. A Robustness Approach to Reliability. Quality and Reliability Engineering International. 29(1), 17–32. DOI: https://doi.org/10.1002/qre.1294

[34] Templ, M., 2023. Enhancing Precision in Large-Scale Data Analysis: An Innovative Robust Imputation Algorithm for Managing Outliers and Missing Values. Mathematics. 11(12), 2729. DOI: https://doi.org/10.3390/math11122729

[35] Grieves, M., Vickers, J., 2016. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In: Kahlen, J., Flumerfelt, S., Alves, A. (Eds.). Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. Springer: Berlin, Germany. pp. 85–113.

[36] Alshammari, K., 2021. Access Management for Digital Twins in the Built Environment [PhD Thesis]. Cardiff University: Cardiff, UK.

[37] Grigg, N.S., 2025. Digital Transformation in Water Utilities: Status, Challenges, and Prospects. Smart Cities. 8(3), 99.

[38] Ferko, E., Bucaioni, A., Behnam, M., 2022. Architecting Digital Twins. IEEE Access. 10, 50335–50350.

[39] Islam, F.A.S., 2025. A Multi-Dimensional AI Framework for Sustainable Drinking Water Management: Integrating Federated Learning, Digital Twins, and Blockchain. Journal of Engineering Research and Reports. 27(6), 466–492.

[40] Alliance for IoT and Edge Computing Innovation (AIOTI), 2024. Edge Driven Digital Twins in Distributed Energy Systems: Role and Opportunities for Hybrid Data Driven Solutions. AIOTI: Brussels, Belgium.

[41] Movva, S.S., Devineni, S.K., Meitivyeki, M.M., et al., 2024. The Future of Digital-Physical Interactions. Journal of Technology and System. 6(3). Available from: https://carijournals.org/product/the-future-of-digital-physical-interactions/?srsltid=AfmBOooAkJxtTnWROSHuNPiB2Y0ZilfjiVFV6AiDAalLNAh0BY_lqMCR

[42] Ma, L., Xu, Z., Fan, L., et al., 2025. Multi-Source Data Integration and Model Coupling for Watershed Eco-Assessment Systems: Progress, Challenges, and Prospects. Processes. 13(9), 2998.

[43] Wu, X., Lu, G., Wu, Z., 2023. Remote Sensing Tech-

nology in the Construction of Digital Twin Basins: Applications and Prospects. Water. 15(11), 2040.

[44] Schimel, D.S., Asner, G.P., Moorcroft, P., 2013. Observing Changing Ecological Diversity in the Anthropocene. Frontiers in Ecology and the Environment. 11(3), 129–137.

[45] Singh, R., Paramanik, S., Bhattacharya, B.K., et al., 2020. Modelling of Evapotranspiration Using Land Surface Energy Balance and Thermal Infrared Remote Sensing. Tropical Ecology. 61(1), 42–50.

[46] Liou, Y.-A., Kar, S.K., 2014. Evapotranspiration Estimation with Remote Sensing and Various Surface Energy Balance Algorithms—A Review. Energies. 7(5), 2821–2849.

[47] Friedl, M.A., 2002. Forward and Inverse Modeling of Land Surface Energy Balance Using Surface Temperature Measurements. Remote Sensing of Environment. 79(2–3), 344–354.

[48] de Sousa, G.P.B., Demattê, J.A.M., Chabrillat, S., et al., 2025. A Sustainable Agricultural Development Index (SADI): Bridging Soil Health, Management, and Socioeconomic Factors. Remote Sensing. 17, 4039.

[49] Russell, D.M., Swanson, D., 2019. Transforming Information into Supply Chain Agility: An Agility Adaptation Typology. The International Journal of Logistics Management. 30(1), 329–355.

[50] Fang, W., Qin, H., Liu, G., et al., 2023. A Method for Spatiotemporally Merging Multi-Source Precipitation Based on Deep Learning. Remote Sensing. 15(17), 4160.

[51] Ma, Y., Zhang, D., Zhang, Y., et al., 2023. Advancements and Challenges in Deep Learning-Driven Marine Data Assimilation: A Comprehensive Review. Computational Research Progress in Applied Science and Engineering: Transactions of Applied Sciences. 9(4), 2876.

[52] Kalogiannidis, S., Spinthiropoulos, K., Kalfas, D., et al., 2025. Integration of Remote Sensing and GIS for Urban Sprawl Monitoring in European Cities. European Journal of Geography. 16(2), 75–90.

[53] Senanayake, I.P., Pathira Arachchilage, K.R.L., Yeo, I.-Y., et al., 2024. Spatial Downscaling of Satellite-Based Soil Moisture Products Using Machine Learning Techniques: A Review. Remote Sensing. 16(12), 2067.

[54] Wei, Z., Fengping, W., Junfei, C., 2023. Methods, Progresses, and Challenges of Passive Microwave Soil Moisture Spatial Downscaling. National Remote Sensing Bulletin. 26(9), 1699–1722.

[55] Göppert, A., Grahn, L., Rachner, J., et al., 2023. Pipeline for Ontology-Based Modeling and Automated Deployment of Digital Twins for Planning and Control of Manufacturing Systems. Journal of Intelligent Manufacturing. 34(5), 2133–2152.

[56] Coito, T., Martins, M.S., Firme, B., et al., 2022. Assessing the Impact of Automation in Pharmaceutical Quality Control Labs Using a Digital Twin. Journal of Manufacturing Systems. 62, 270–285.

[57] Eriksson, K., Markussen, C., 2023. Quality Assurance of Digital Twins. In Proceedings of the ASME 2023 42nd International Conference on Ocean, Offshore and Arctic Engineering, Melbourne, Australia, 11–16 June 2023. DOI: https://doi.org/10.1115/OMAE2023-105285

[58] Curl, J.M., Nading, T., Hegger, K., et al., 2019. Digital Twins: The Next Generation of Water Treatment Technology. Journal of the American Water Works Association. 111(12), 44–50.

[59] Song, P., Wang, C., Zhang, W., et al., 2020. Urban Multi-Source Water Supply in China: Variation Tendency, Modeling Methods and Challenges. Water. 12(4), 1199.

[60] Rahman, M.A., Shahrior, M.F., Iqbal, K., et al., 2025. Enabling Intelligent Industrial Automation: A Review of Machine Learning Applications with Digital Twin and Edge AI Integration. Automation. 6(3), 37.

[61] Sheraz, M., Chuah, T.C., Lee, Y.L., et al., 2024. A Comprehensive Survey on Revolutionizing Connectivity through Artificial Intelligence-Enabled Digital Twin Network in 6G. IEEE Access. 12, 49184–49215.

[62] Vidyalakshmi, G., Gopikrishnan, S., Boulila, W., et al., 2025. Digital Twins and Cyber-Physical Systems: A New Frontier in Computer Modeling. Computer Modeling in Engineering & Sciences. 143(1), 51.

[63] Abu-Elkheir, M., Hayajneh, M., Ali, N.A., 2013. Data Management for the Internet of Things: Design Primitives and Solution. Sensors. 13(11), 15582–15612.

[64] Shi, F., Li, Q., Zhu, T., et al., 2018. A Survey of Data Semantization in Internet of Things. Sensors. 18(1), 313.

[65] Bao, J., Guo, D., Li, J., et al., 2019. The Modelling and Operations for the Digital Twin in the Context of Manufacturing. Enterprise Information Systems. 13(4), 534–556.

[66] Fantozzi, I.C., Santolamazza, A., Loy, G., et al., 2025. Digital Twins: Strategic Guide to Utilize Digital Twins to Improve Operational Efficiency in Industry 4.0. Future Internet. 17(1), 41.

[67] Wang, Q., Wu, W., Qian, L., et al., 2023. Design and Implementation of Secure and Reliable Information Interaction Architecture for Digital Twins. China Communications. 20(2), 79–93.

[68] Song, J., Kim, J., Lee, M.J., et al., 2022. Differential Diagnosis between Parkinson's Disease and Atypical Parkinsonism Based on Gait and Postural Instability: Artificial Intelligence Using an Enhanced Weight Voting Ensemble Model. Parkinsonism & Related Disorders. 98, 32–37.

[69] Li, S., Cheng, W., Liu, Z., et al., 2025. Key Technologies and Applications of Digital Twin Hydraulic Engineering. Digital Twin. 2(1), 2486862.

[70] Brocca, L., Barbetta, S., Camici, S., et al., 2024. A Digital Twin of the Terrestrial Water Cycle: A Glimpse into the Future through High-Resolution Earth Observations. Frontiers in Science. 1, 1190191.

[71] Jiang, Y., Li, M., Li, M., et al., 2022. Digital Twin-Enabled Real-Time Synchronization for Planning, Scheduling, and Execution in Precast On-Site Assembly. Automation in Construction. 141, 104397.

[72] Pan, Y., Qu, T., Wu, N., et al., 2021. Digital Twin-Based Real-Time Production Logistics Synchronization System in a Multi-Level Computing Architecture. Journal of Manufacturing Systems. 58, 246–260.

[73] Tan, B., Matta, A., 2024. The Digital Twin Synchronization Problem: Framework, Formulations, and Analysis. IISE Transactions. 56(6), 652–665.

[74] Thelen, A., Zhang, X., Fink, O., et al., 2023. A Comprehensive Review of Digital Twin—Part 2: Roles of Uncertainty Quantification and Optimization, a Battery Digital Twin, and Perspectives. Structural and Multidisciplinary Optimization. 66(1), 23.

[75] Paredis, R., Vangheluwe, H., 2024. Modelling and Simulation-Based Evaluation of Twinning Architectures and Their Deployment. In Proceedings of the 14th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH), Dijon, France, 2024; pp. 170–182.

[76] Mihai, S., Yaqoob, M., Hung, D.V., et al., 2022. Digital Twins: A Survey on Enabling Technologies, Challenges, Trends and Future Prospects. IEEE Communications Surveys & Tutorials. 24(4), 2255–2291.

[77] Rathore, M.M., Shah, S.A., Shukla, D., et al., 2021. The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities. IEEE Access. 9, 32030–32052.

[78] Bolbotinović, Ž., Milić, S.D., Janda, Ž., et al., 2025. AI-Powered Digital Twin in the Industrial IoT. International Journal of Electrical Power & Energy Systems. 167, 110656.

[79] Ciliberti, F.G., Berardi, L., Laucelli, D.B., et al., 2023. From Digital Twin Paradigm to Digital Water Services. Journal of Hydroinformatics. 25(6), 2444–2459.

[80] Wu, Z.Y., Chew, A., Meng, X., et al., 2023. High Fidelity Digital Twin-Based Anomaly Detection and Localization for Smart Water Grid Operation Management. Sustainable Cities and Society. 91, 104446.

[81] van der Valk, H., Haße, H., Möller, F., et al., 2022. Archetypes of Digital Twins. Business & Information Systems Engineering. 64(3), 375–391.

[82] Al Zami, M.B., Shaon, S., Quy, V.K., et al., 2025. Digital Twin in Industries: A Comprehensive Survey. arXiv preprint. arXiv:2412.00209. DOI: https://doi.org/10.48550/arXiv.2412.00209

[83] Mayr, S., Kuenzer, C., Gessner, U., et al., 2019. Validation of Earth Observation Time-Series: A Review for Large-Area and Temporally Dense Land Surface Products. Remote Sensing. 11(22), 2616.

[84] Emerson, R.A., Hendry, D.F., 1996. An Evaluation of Forecasting Using Leading Indicators. Journal of Forecasting. 15(4), 271–291.

[85] Yusupov, R.M., Musaev, A.A., Grigoriev, D.A., 2021. Evaluation of Statistical Forecast Method Efficiency in the Conditions of Dynamic Chaos. In Proceedings of the 2021 IV International Conference on Control in Technical Systems (CTS), Saint Petersburg, Russia, 21–23 September 2021; pp. 178–180.

[86] Liu, Y., Ong, S., Nee, A., 2022. State-of-the-Art Survey on Digital Twin Implementations. Advances in Manufacturing. 10(1), 1–23.

[87] Minerva, R., Lee, G.M., Crespi, N., 2020. Digital Twin in the IoT Context: A Survey on Technical Features, Scenarios, and Architectural Models. Proceedings of the IEEE. 108(10), 1785–1824.

[88] Thelen, A., Zhang, X., Fink, O., et al., 2022. A Comprehensive Review of Digital Twin—Part 1: Modeling and Twinning Enabling Technologies. Structural and Multidisciplinary Optimization. 65(12), 354.

[89] Alsubaih, A.A.S., Sepehrnoori, K., Delshad, M., et al., 2025. A Comprehensive Review of Well Integrity Challenges and Digital Twin Applications across Conventional, Unconventional, and Storage Wells. Energies. 18(17), 4757.

[90] Uhlenkamp, J.-F., Hauge, J.B., Broda, E., et al., 2022. Digital Twins: A Maturity Model for Their Classification and Evaluation. IEEE Access. 10, 69605–69635.

[91] Klar, R., Arvidsson, N., Angelakis, V., 2023. Digital Twins' Maturity: The Need for Interoperability. IEEE Systems Journal. 18(1), 713–724.

[92] Dietrich, C.F., 2017. Uncertainty, Calibration and Probability: The Statistics of Scientific and Industrial Measurement. Routledge: London, UK.

[93] Byaruhanga, N., Kibirige, D., Gokool, S., et al., 2024. Evolution of Flood Prediction and Forecasting Models for Flood Early Warning Systems: A Scoping Review. Water. 16(13), 1763.

[94] Forouheshfar, Y., Ayadi, R., Moghadas, O., 2025. Enhancing System Resilience to Climate Change through Artificial Intelligence: A Systematic Literature Review. Frontiers in Climate. 7, 1585331.