International Journal of Digital Technology Driven Engineering, Vol. 1 No. 1 (2026), Article 002

ESN Publishing

Contents lists available at ESN Publishing Journals

International Journal of Digital Technology Driven Engineering

journal homepage: https://journals.esn-press.com/ijdtde

REVIEW ARTICLE

Model Updating and Data Assimilation in Engineering Digital Twins: A Cross-Domain Critical Review of Methods, Evidence, and Validation

Stelios K. Georgantzinos [1,*] and Nikos D. Lagaros [2]

[1] *Department of Aerospace Science and Technology, National and Kapodistrian University of Athens, Psachna, Greece*
[2] *Institute of Structural Analysis and Antiseismic Research, Department of Structural Engineering, School of Civil Engineering, National Technical University of Athens, 9 Heroon Polytechniou Street, Zografou Campus, GR-15780 Athens, Greece*

ARTICLE INFO

Article type:
Review Article

Keywords:
Digital Twin; model updating; data assimilation; Bayesian calibration; state and parameter estimation; uncertainty quantification; validation and credibility; engineering systems

ABSTRACT

Engineering Digital Twins rely on the continuous alignment of virtual models with evolving physical systems, yet the computational mechanisms that enable this alignment are often described inconsistently and assessed with uneven standards of evidence. This review critically examines model updating and data assimilation as the principal physical-to-virtual synchronization mechanisms of engineering Digital Twins across structural, manufacturing, aerospace, thermal, energy, robotic, and built-environment applications. Particular attention is given to the quantities being updated, including dynamic states, physical parameters, unknown inputs, boundary conditions, model structure, and model-form discrepancy, as well as to the temporal regime in which updating is performed. The synthesis shows that no inference method is universally preferable. Deterministic and finite-element updating provide transparent solutions for interpretable batch calibration; Bayesian approaches support uncertainty-aware parameter and discrepancy estimation; Kalman and particle-filter methods enable sequential state and parameter assimilation; and reduced-order, surrogate, and hybrid models address the computational demands of operational deployment. Method suitability depends on identifiability, model fidelity, nonlinearity, observation cadence, computational cost, uncertainty structure, interpretability requirements, and the consequence of prediction error. Across application domains, the most credible studies combine a complete physical-observation loop with independent or changed-condition validation, explicit uncertainty treatment, defensible latency assessment, and clearly defined applicability limits. Recurring weaknesses include conflation of calibration with validation, unsupported real-time claims, incomplete treatment of model-form error, and limited reproducibility. The review develops a unified taxonomy, an evidence-based method-selection framework, a validation-maturity hierarchy, and reporting guidance for the development of credible, continuously evaluated, and decision-relevant engineering Digital Twins.

1 Introduction

1.1 Digital Twins as Maintained Engineering Models

Digital Twins have evolved from a lifecycle-information concept into a broad engineering paradigm for monitoring, prediction, design support, maintenance, and control. Their defining promise is not merely that a virtual representation exists, but that the representation remains meaningfully connected to a physical

∗ Corresponding author. E-mail address: sgeor@uoa.gr.

counterpart as the asset, process, environment, and decision context change. This distinction separates a Digital Twin from an isolated simulation, a static digital model, or a one-way digital shadow. Broad reviews have established the importance of data connectivity, modelling, uncertainty, and lifecycle integration, while also documenting persistent inconsistency in definitions and implementation maturity [1–13]. Early aerospace and manufacturing formulations similarly emphasized asset-specific lifecycle coupling and the need to distinguish a maintained twin from a standalone model [14–17].

For engineering use, the word twin therefore implies a maintained inferential relationship. The virtual model must receive observations, determine whether prediction errors arise from changing states, parameters, loads, boundary conditions, sensor behaviour, or model inadequacy, and revise the representation without destroying its physical interpretability. This process is central to the physical-to-virtual direction of twinning. The reverse direction—using the updated model to recommend or execute actions—depends on the credibility of that inference. A Digital Twin that visualizes sensor data but cannot explain how observations revise an engineering model may be operationally useful, but it does not address the model-synchronization problem examined in this review.

The National Academies identifies data assimilation, state and parameter estimation, uncertainty-aware prediction, model calibration, and scalable computation as coupled foundational requirements for Digital Twins [18]. These requirements connect the field to mature disciplines: system identification [19–21], inverse problems, finite-element model updating [22–26], Bayesian calibration [27–34], filtering [35–43], reduced-order modelling [44–46], uncertainty quantification, and verification and validation [47, 48]. The novelty of Digital Twin research is consequently not the invention of all these methods, but their integration into recurrent, decision-oriented, and lifecycle-aware engineering workflows.

1.2 The Physical-to-Virtual Synchronization Problem

Physical-to-virtual synchronization is an inverse problem. The physical system generates incomplete, noisy, delayed, and often asynchronous observations, whereas the virtual representation contains states and quantities that are only partly observable. The updating algorithm must infer which latent quantities changed and how strongly the new observations should modify the previous model. This inference is frequently ill posed: several parameter combinations may produce similar responses, sensor placement may be insufficient, and flexible discrepancy terms may compensate for incorrect physics. A small calibration residual can therefore coexist with weak predictive validity. The inverse-problem and uncertainty-quantification literature provides the statistical regularization, prior modelling, and computational machinery required to make this inference well posed [49–52].

Synchronization is also a multi-timescale computational problem. States may require correction at the sensor rate, parameters may evolve over hours or load cycles, geometry and material properties may change through damage or wear, and model structure may require episodic revision after a regime change. High-fidelity finite-element, computational-fluid-dynamics, or multiphysics models are rarely affordable at all these timescales. Engineering Digital Twins consequently combine full-order models with reduced-order representations, surrogates, filters, probabilistic models, sparse identification, neural operators, or residual corrections. Claims of real-time capability are meaningful only when the complete measurement-to-decision latency is reported relative to the operational deadline.

Finally, synchronization is a credibility problem. An updated model may fit the data used for calibration while failing under a new load case, operating regime, sensor configuration, or damage mechanism. Credibility therefore requires more than updating accuracy: it requires verification of the implementation, separation of calibration and validation information, decomposition of uncertainty, monitoring of applicability, and evidence proportional to the risk of the supported decision. These considerations motivate a review centred on methods, evidence, and validation rather than on Digital Twin architectures alone. Established predictive-validation frameworks further show why calibration agreement alone cannot justify extrapolative or decision-relevant claims [53, 54].

1.3 Previous Reviews and the Remaining Gap

Existing Digital Twin reviews address several adjacent questions. Jones et al. characterized definitions, connections, and twinning rates across 92 publications [5], while Semeraro et al. synthesized the broader paradigm, applications, and technical challenges [6]. Rasheed et al. and the two-part review by Thelen et al. examined modelling foundations, enabling technologies, uncertainty quantification, and optimization [1, 3, 4].

Es-haghi et al. focused on methods that enable real-time analysis [55]. Recent reviews have extended the discussion to model-based Digital Twin engineering [56], engineering-development workflows and industrial potential [57], and structural-dynamics condition monitoring [58]. In parallel, established finite-element model-updating reviews provide detailed coverage of structural and material calibration [59, 60].

These reviews are individually valuable but leave a cross-domain integration gap. Broad Digital Twin surveys typically treat updating as one enabling capability among architecture, communication, visualization, analytics, optimization, and control. Real-time reviews emphasize acceleration but do not consistently distinguish state estimation, parameter calibration, model-structure revision, and discrepancy learning. Domain reviews provide greater technical depth but cannot determine whether conclusions transfer between structures, manufacturing systems, vehicles, thermal processes, buildings, batteries, and robotic systems. Conventional model-updating reviews, conversely, do not require an operational physical-to-virtual loop or assess whether a calibration study functions as a Digital Twin.

Among the closest reviews identified, none jointly asks across engineering domains: what part of the model is updated from physical observations; which inference methods are used and under which assumptions; how measurement, parameter, process, and model-form uncertainties are represented; whether online and real-time claims are computationally substantiated; and whether the updated model is validated independently for an engineering output. This combined methods-and-evidence perspective defines the contribution of the present review.

The originality boundary is established through the preceding comparison of the closest reviews in terms of updating-target differentiation, physical-evidence requirements, uncertainty treatment, model-form discrepancy, latency reporting, independent validation, cross-domain coverage, and reporting guidance.

1.4 Objectives, Scope, and Contributions

The objective is to develop an evidence-based account of model updating and data assimilation as the physical-to-virtual synchronization mechanism of engineering Digital Twins. The review is deliberately narrower than a general survey of Digital Twin architectures and broader than a review of one inference method or application domain. It includes physics-based, reduced-order, probabilistic, data-driven, and hybrid models when physical observations explicitly modify the virtual representation and the resulting engineering output is evaluated.

The review makes six contributions. First, it defines a strict evidence boundary separating physical observation-driven Digital Twins from digital models, digital shadows, conceptual architectures, and synthetic proofs of method. Second, it develops a unified taxonomy based on updating target, model representation, temporal regime, uncertainty treatment, and validation scale. Third, it compares deterministic, Bayesian, sequential, reduced-order, surrogate, and hybrid methods using engineering decision variables rather than nominal algorithm rankings. Fourth, it maps cross-domain physical evidence and proposes a validation-maturity hierarchy. Fifth, it introduces EDTURC-24, a reporting checklist for observation-updated engineering twins. Sixth, it identifies research priorities for managed model co-evolution, active sensing, continuous credibility, interoperability, and reproducible benchmarking.

The remainder of the manuscript is organized as follows. Section 2 describes the review protocol and evidence synthesis. Section 3 establishes the foundations and taxonomy of physical-observation-driven updating. Sections 4 and 5 examine offline updating and sequential data assimilation, respectively. Section 6 covers reduced-order, data-driven, and hybrid methods. Section 7 presents the method-selection framework, Section 8 synthesizes cross-domain evidence, and Section 9 addresses validation and credibility. Section 10 develops the research roadmap, and Section 11 concludes the review.

2 Review Methodology

2.1 Review Design and Research Questions

The study is designed as a PRISMA 2020-informed structured critical review supported by systematic mapping, documented open-bibliographic searching and evidence appraisal [61]. The purpose is explanatory and comparative: to determine how physical observations modify engineering Digital Twins, what evidence supports the resulting models and how method choice depends on the updating target, data cadence, model cost, uncertainty structure and context of use. PRISMA principles are used to structure transparency and auditability; the review is not presented as a subscription-index census of the literature.

Five research questions guide the synthesis. RQ1 asks which parts of engineering Digital Twins are updated from physical observations. RQ2 asks which deterministic, probabilistic, sequential, reduced-order, data-driven, and hybrid methods are used and under what assumptions. RQ3 asks how method choice varies across engineering domains, model fidelities, observation structures, and temporal requirements. RQ4 asks how uncertainty, model-form discrepancy, identifiability, computational latency, and independent validation are addressed. RQ5 asks which reporting deficiencies and research gaps prevent observation-updated twins from supporting credible engineering decisions.

2.2 Two-Corpus Strategy and Search Scope

Two linked corpora were defined because explicit Digital Twin terminology does not capture all methodological foundations, while searches for model updating and data assimilation retrieve many studies that are not Digital Twins. Corpus A contains explicit or functionally equivalent engineering Digital Twin studies that satisfy the physical-evidence criteria defined below. Corpus B contains foundational methods and boundary cases, including finite-element model updating, Bayesian calibration, filtering, inverse mapping, reduced-order modelling, model-discrepancy treatment, and synthetic Digital Twin demonstrations that inform method comparison but do not meet the full physical-twinning threshold. Reviews, standards, and credibility guidance form a separate contextual layer.

The search covered publications from 2010 through 28 July 2026 for Corpus A, with selected pre-2010 methodological foundations retained in Corpus B. Five core query families were used: Q1, "digital twin*" AND ("model updating" OR "data assimilation"); Q2, "digital twin*" AND ("calibration" OR "parameter estimation" OR "state estimation"); Q3, ("self-updating" OR "adaptive" OR "time-evolving") AND "digital twin*"; Q4, "digital twin*" AND ("Kalman filter" OR "particle filter" OR "moving-horizon estimation" OR "3D-Var" OR "4D-Var"); and Q5, ("model updating" OR "recursive identification" OR "system identification") AND ("physical measurement*" OR sensor*) AND engineering. The core concepts were preserved while punctuation and field syntax were adapted to the individual bibliographic and publisher interfaces. Detailed search records, including platform-specific query formulations, dates, fields, and discovery pathways, are available from the corresponding author upon reasonable request.

The open-search workflow was executed and documented on 28 July 2026. Because ranked public interfaces did not expose stable raw-result totals, transient hit counts were not treated as a reproducible denominator. Instead, every retained candidate was logged individually by query family, discovery route, title, DOI or canonical record, duplicate decision, eligibility result, and exclusion reason. Forty-two unique records reached full-text or authoritative-record assessment. Figure 1 reports the documented flow from the assessed records to corpus assignment. Detailed record-level screening information is available from the corresponding author upon reasonable request.

2.3 Eligibility Criteria and Corpus Assignment

A study enters Corpus A only when six elements are present: a physical engineering counterpart; observations originating from that counterpart; an engineering model or model-based virtual representation; an explicit computational update driven by the observations; an engineering prediction, diagnosis, prognosis, control, or decision output; and validation or assessment of the updated result. All six conditions are required. The criterion prevents papers from entering the evidence corpus solely because they use Digital Twin terminology.

One-time calibration is treated carefully. It may constitute a methodological foundation or commissioning step, but it is not automatically classified as an operational Digital Twin. Admission to Corpus A requires evidence that the calibration is embedded in, or explicitly supports, a repeatable physical-to-virtual process. Studies using physical data but identifying themselves as one-way digital shadows remain in Corpus B when no recurrent or decision-oriented updating loop is demonstrated. Synthetic observations, virtual sensors generated by the same simulator, and analytical benchmarks may support method development but do not satisfy the physical-observation requirement.

Studies are excluded from primary evidence when they are conceptual architectures without implemented updating, monitoring or classification studies without revision of an engineering model, visualization platforms without inference, non-engineering applications outside the review scope, reviews or standards, duplicate or superseded records, or reports with insufficient methodological information. Conference papers and preprints are not automatically excluded, but their evidence status and peer-review status are recorded explicitly.

2.4 Screening, Deduplication, and Data Extraction

The workflow comprises query logging, candidate capture, DOI and title metadata verification, deduplication, title-and-abstract screening, full-text or authoritative-record assessment, structured extraction, quality appraisal and publication-status checking. Exact DOI duplicates are merged first, followed by exact title-year matches and manual review of near-title variants, conference precursors, preprints and versions of record. Withdrawn, corrected or retracted records are explicitly checked and retained only in the audit trail when they cannot support the evidence synthesis.

The extraction matrix contains bibliographic, physical, computational, and evidentiary fields. These include engineering domain; physical counterpart and twin scale; lifecycle phase; measured quantities and sensors; model type, fidelity, and representation; updating target and specific updated quantities; method family and algorithm; temporal regime; treatment of measurement, process, parameter, and model-form uncertainties; identifiability and observability considerations; calibration and validation datasets; validation scale; baselines and metrics; computational runtime and update latency; engineering output; and data or code availability. Separating these fields enables cross-domain comparison without collapsing distinct inference problems into a single model-updating category.

The review screened 42 records at full-text or authoritative-record level. Twenty-three were assigned to Corpus A, 17 to Corpus B or contextual literature, one remained uncertain, and one withdrawn preprint was excluded from evidence. For each included or methodological study, extraction covered the physical counterpart, observations, engineering model, updating target, algorithm, uncertainty treatment, validation, computational performance, reproducibility and decision relevance.

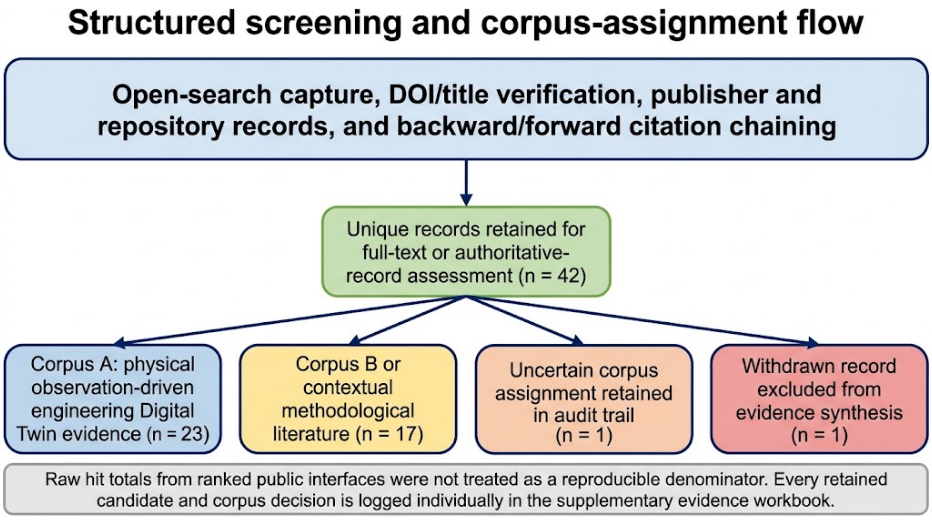

Figure 1: Structured screening and corpus-assignment flow. Forty-two unique records reached full-text or authoritative-record assessment: 23 entered Corpus A, 17 entered Corpus B or the contextual layer, one remained uncertain, and one withdrawn record was excluded. Raw hit totals from ranked public interfaces were not treated as a reproducible denominator because they were unstable; retained candidates and screening decisions were documented individually.

Corpus assignments and criterion-level appraisal records were checked against the predefined eligibility and scoring rules. The criterion-level appraisal records and record-level screening rationales are available from the corresponding author upon reasonable request.

2.5 Evidence-Quality Appraisal

Each primary study is appraised across eight domains scored from 0 to 2: physical evidence, clarity of the updating target, algorithm transparency, uncertainty treatment, independent validation, baseline comparison, computational-performance reporting, and reproducibility. Total scores range from 0 to 16. Scores of 13–16 are classified as Strong, 9–12 as Moderate, 5–8 as Limited, and 0–4 as Preliminary. The score is used as a transparent synthesis aid rather than as a universal measure of scientific quality. For every domain, 0 denotes absent or unsupported reporting, 1 denotes partial reporting or limited evidence, and 2 denotes comprehensive, decision-relevant evidence. The domain-specific appraisal anchors and study-level criterion scores were retained as part of the review records and are available from the corresponding author upon reasonable request.

The physical-evidence item distinguishes synthetic demonstrations, mixed numerical/physical studies, laboratory demonstrators, pilot-scale systems, replayed experimental datasets, and live operational deployments. Validation scoring rewards separation between calibration and validation information and evidence under changed conditions. Computational scoring considers end-to-end latency, hardware, sample or ensemble size, convergence criteria, and comparison with the operational deadline. Reproducibility scoring considers the availability of data, code, complete model definitions, and sufficient implementation detail. A threshold-sensitivity check classified 18 of 23 studies as Strong when the boundary was reduced to 12/16, 14 of 23 at the prespecified 13/16 boundary, and 9 of 23 when it was increased to 14/16. The principal cross-domain deficiencies—uncertainty treatment, independent validation, reproducibility, and end-to-end latency reporting—were unchanged across these alternatives.

2.6 Synthesis Strategy, Bias Control, and Limitations

The evidence is synthesized through structured narrative comparison, evidence mapping, and cross-tabulation rather than statistical pooling. Methods are compared according to updating target, observation structure, model fidelity, uncertainty assumptions, latency, interpretability, validation maturity, and decision context. Strong and Moderate studies contribute to the cross-domain evidence synthesis, while Corpus B studies support methodological explanation and clarify the difference between algorithmic capability and physical-twinning maturity.

Several bias controls are built into the protocol. The two-corpus strategy reduces terminology bias; the strict physical-evidence threshold reduces maturity inflation; the review-of-reviews establishes novelty and overlap boundaries; citation chaining improves recall for studies that do not use standard Digital Twin terminology; and the explicit quality rubric prevents high-profile terminology or algorithmic novelty from substituting for validation evidence. Search decisions, exclusion reasons, extracted data fields, and quality-appraisal records were documented systematically and are available from the corresponding author upon reasonable request.

The review has four important limitations. First, ranked public scholarly interfaces do not expose stable raw totals or a single machine-readable export; the reported counts are therefore corpus-descriptive and are not estimates of literature prevalence. Second, some recent or industrial studies provide only authoritative records, accepted manuscripts, or restricted evidence packages, limiting appraisal of runtime, provenance, and calibration–validation separation. Third, the evidence base is heterogeneous across domains and contexts of use, precluding statistical pooling. Fourth, screening, extraction, and appraisal were conducted by the first author, which may introduce selection and judgement bias. The predefined criteria, evidence records, and criterion-level scores were documented systematically and are available from the corresponding author upon reasonable request. Conclusions are consequently framed as findings from the documented corpus rather than universal prevalence claims.

3 Physical-Observation-Driven Updating: Foundations

The defining computational problem of an engineering Digital Twin is not the construction of a virtual model, but the maintenance of a defensible relationship between that model and a changing physical counterpart. Conventional simulation begins with a model, prescribed inputs and assumed parameters. A Digital Twin must additionally process observations, infer unmeasured quantities, revise the virtual representation and communicate uncertainty in the resulting prediction. This physical-to-virtual pathway links Digital Twin research to system identification, inverse problems, finite-element model updating, Bayesian calibration and data assimilation [1–4, 59, 60]. The relevant methods are mature individually, but their use within Digital Twins is frequently described with inconsistent terminology and uneven validation requirements.

For the present review, model updating is defined as any explicit computational operation through which observations from a physical engineering system modify the states, parameters, inputs, boundary conditions, geometry, discrepancy representation or structure of its virtual model. This definition is deliberately stricter than common usage in the Digital Twin literature. A dashboard, data pipeline or fault classifier is not sufficient unless it alters an engineering model or a model-based belief state. Similarly, a one-time calibration experiment can provide essential methodological foundations, but it is treated as a complete Digital Twin only when it belongs to a repeatable or persistent physical-to-virtual loop. The distinction is essential because several technically advanced studies in the screened corpus use simulated observations, analytical benchmarks or one-way Digital Shadows rather than operationally updated twins [62–66].

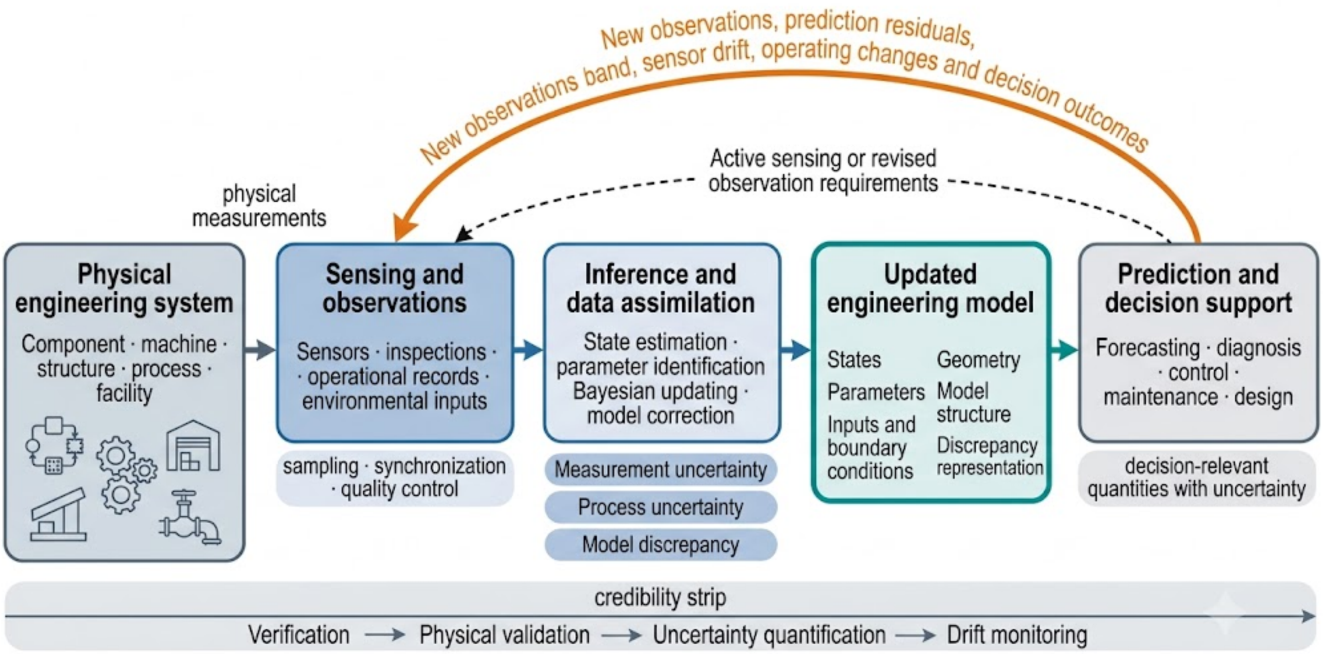

Figure 2: Physical-to-virtual updating loop adopted in this review. Measurements from the physical engineering system are processed through state estimation, parameter identification, data assimilation or model-discrepancy correction to update the virtual representation. Predictions and decisions subsequently generate new observation requirements, residual information and operating-domain changes, making synchronization a recurrent inference and credibility-management process rather than a one-time calibration exercise.

3.1 Physical Counterparts, Observations, and the Evidence Threshold

The physical counterpart may be a component, machine, structure, production process, vehicle, facility or distributed field. The screened evidence includes laboratory structures, a two-tank hydraulic system, a bridge demonstrator, a hydrogen-fired furnace, a gearbox, a machine-tool thermal test bench, aircraft flight measurements, a mobile robot, battery cycling experiments, production data centres and operational industrial-process records [67–87]. These cases differ substantially in sensor density, sampling frequency, physical scale and accessibility of ground truth. Consequently, the phrase "physical data" should not be treated as a binary quality marker. Evidence from a controlled laboratory demonstrator can provide a clearer reference state than field data, whereas field deployment better tests robustness to missing data, unknown loads, environmental variability and sensor drift.

A practical evidence hierarchy is therefore required. At the lowest level, synthetic observations are generated from the same or a closely related model used by the updating algorithm. Such studies are valuable for algorithm verification but cannot establish physical validity. The next level comprises analytical or numerical benchmarks with deliberately introduced noise or model mismatch. Laboratory and pilot-scale experiments then test the interaction between real sensors and imperfect models under controlled conditions. The highest deployment maturity is represented by operational or field data, such as continuous tower vibration monitoring, production-process measurements, real vehicle trips, flight parameters, shield-tunnel observations or production data-centre temperatures [78, 81, 84–88]. Laboratory systems can nevertheless provide stronger inferential evidence than field studies when raw observations, independent validation, model bias and computational cadence are reported comprehensively, as illustrated by the thermoacoustic twin [89]. The maturity level should therefore be reported independently of numerical accuracy.

3.2 Engineering Model Representations

The virtual representation determines both what can be updated and which inference methods are computationally feasible. Full-order physics models retain explicit governing equations and spatial resolution, but their cost often prevents repeated likelihood evaluations or sequential filtering [90–92]. Finite-element models remain dominant in structural updating, where stiffness, material properties, boundary conditions and damage parameters are inferred from modal, displacement, strain or force measurements [59, 60, 73, 76–78]. Full-order models provide interpretability and a direct connection to engineering design variables, but they can be non-identifiable when many parameters produce similar response changes.

Reduced-order models lower the state dimension while preserving selected physical relationships. Proper orthogonal decomposition, Krylov reduction and mesh-reduction methods appear in the furnace, machine-tool and bridge cases [75, 77, 80]. Their main advantage is not merely faster prediction. A reduced model can make filtering, sampling and uncertainty propagation possible within the time available between observations. However, reduction introduces an additional approximation layer whose validity may change as the physical asset leaves the training or reduction domain. The reduced basis or surrogate must therefore be treated as part of the uncertainty and validation problem, not as a neutral computational substitution.

Data-driven and hybrid representations include neural inverse mappings, sparse identified dynamics, recurrent networks, neural operators and thermophysics-informed surrogates [63–66, 69–71, 83–86, 93–98]. These approaches move computational cost from online inference to offline data generation and training. They can achieve millisecond-level parameter estimates or rapid field reconstruction, but their reliability depends on coverage of the training domain and on the consistency between simulated training data and physical measurements. The probabilistic inverse-mapping study of Kessels et al. illustrates both the promise and the limitation: physical ventilation measurements can be processed rapidly, yet observations outside the predicted intervals reveal residual model-class mismatch that the inverse network alone does not correct [71].

3.3 Updating Targets

A unified taxonomy begins with the quantity that is modified. State updating reconstructs unmeasured time-dependent variables, such as temperature fields, structural states, battery state of charge or latent fluid states. Parameter updating identifies persistent or slowly varying quantities, including stiffness, heat-transfer coefficients, damping, boundary-condition coefficients or process parameters. Input updating estimates unknown loads, environmental conditions or control inputs. Geometry and boundary-condition updating revise the spatial definition of the model. Discrepancy updating introduces an explicit correction for systematic model bias, whereas structural updating changes the model form, active terms, surrogate weights or neural representation. Joint estimation treats two or more of these quantities simultaneously.

Table 1. Updating targets and their principal engineering interpretation.

Target	Generic notation	Engineering meaning	Representative evidence	Principal risk
Dynamic state	x_k	Time-varying internal variables that are not fully measured.	Wave, thermoacoustic, thermal, tunnel, robot and battery states [72, 80, 82, 83, 88, 89, 95].	Unobservability and filter divergence.
Physical/model parameter	theta	Material, stiffness, damping, thermal, process or equipment coefficients.	Bridge, furnace, ventilation, gearbox, vehicle, datacentre, thermoacoustic and tunnel calibration [67, 68, 71, 73, 75, 79, 85, 87–89].	Non-identifiability and compensation between parameters.
Unknown input/load	u_k	Loads, environmental inputs, excitations or operating conditions.	Wind/thrust estimation, flight-condition correction and tunnel boundary-condition inference [81, 88, 96].	Input–state confounding.
Geometry/boundary condition	g, b	Supports, contact, geometry and spatial model conditions.	Cable-truss, tower and bridge FEM updating [76–78].	Over-parameterization and non-physical corrections.

Target	Generic notation	Engineering meaning	Representative evidence	Principal risk
Model discrepancy	delta_M	Systematic error not removable through physical-parameter calibration.	Bridge-bias methods, aircraft residual feedback and real-time thermoacoustic bias estimation [74, 81, 89].	Confounding with calibration parameters.
Model structure/representation	M	Active equations, model combinations, neural weights or surrogate architecture.	Time-evolving hybrid models, EKF-SINDYc, online thermal models and event-triggered re-identification [69, 84, 86, 97].	Catastrophic drift and loss of physical interpretability.
Joint target	x, theta, u	Simultaneous state, parameter and/or input inference.	Latent-space particle filtering, augmented estimation, thermoacoustic r-EnKF and tunnel EnKF-SuS [72, 88, 89, 96].	Rapid growth of dimension and weak observability.

3.4 Temporal Synchronization Regimes

Temporal terminology should reflect how frequently the physical information is used. Offline batch calibration processes a fixed dataset and produces a revised model after the experiment or inspection. Periodic updating repeats this operation at scheduled intervals or after sufficient new data have accumulated. Event-triggered updating is initiated by a detected change, such as a residual threshold, damage indicator or operating-regime transition. Online sequential methods update the state or parameters as each observation or short data window arrives. Continuous adaptive methods additionally modify the model representation or learning system throughout operation. These regimes are not interchangeable: an algorithm with fast inference is only online if it is actually connected to sequential observations and its latency is compatible with the use case.

The screened evidence reveals frequent inflation of temporal claims. Some papers describe a framework as "self-updating" or "real-time" while validating it only with model-generated observations or replayed datasets [63–66]. Conversely, an offline method can still be highly valuable for a Digital Twin when the engineering decision is slow, such as periodic structural re-calibration or fleet-level vehicle parameter estimation. The appropriate criterion is therefore use-case latency rather than a universal requirement for continuous computation. A machine-tool compensation twin may require updates faster than thermal deformation evolves, whereas a bridge model may only need recalibration after a monitoring campaign or suspected damage event.

Table 2. Temporal regimes, evidence requirements, and typical engineering use.

Regime	Data handling	Required evidence	Typical use	Common misclassification
One-time calibration	Single fixed dataset.	Independent prediction after calibration.	Model commissioning and initial parameter identification.	Described as a Digital Twin despite no repeatable data connection.
Offline batch	Accumulated dataset processed after operation/test.	Repeatable workflow and updated engineering output.	FEMU, damage assessment, fleet calibration.	Training/test split treated as operational synchronization.
Periodic	Updates at scheduled intervals or data volumes.	Multiple update cycles or distinct conditions.	Battery ageing, bridge monitoring, facility recalibration.	A single recalibration labelled continuous.
Event-triggered	Update initiated by residual/change condition.	Clear trigger and post-event validation.	Damage, regime change, maintenance.	Threshold detection without model modification.
Online sequential	Observation-by-observation or short-window inference.	Measured latency and sequential physical data.	State estimation, control, thermal compensation.	Fast simulated inference without sensor connection.
Continuous adaptive	Persistent updating of parameters or model representation.	Longitudinal stability and drift assessment.	Self-updating process or equipment twins.	Repeated optimization on synthetic data.

3.5 Uncertainty and Model-Form Discrepancy

Uncertainty is not a single quantity. Measurement uncertainty concerns sensor noise, calibration error, sampling and synchronization. Process uncertainty describes unresolved disturbances and stochastic evolution. Parameter uncertainty reflects incomplete knowledge of coefficients. Numerical and surrogate uncertainty arise from discretization, reduction and learned approximations. Model-form discrepancy represents systematic error caused by missing or incorrect physics. These sources affect updating differently and should not be merged into an undifferentiated error term [4, 74, 99].

Kalman-family methods represent uncertainty through state and parameter covariances, but these covariances are meaningful only when the model, noise assumptions and linearization are defensible. Particle filters accommodate non-Gaussian and multimodal distributions at greater computational cost. Bayesian calibration provides posterior parameter distributions and can incorporate a discrepancy process, but parameter–discrepancy confounding can make physically incorrect parameters appear statistically plausible. Imprecise-probability updating offers an alternative when prior and likelihood specifications are poorly justified [68]. Hybrid neural methods increasingly report aleatoric and epistemic uncertainty, although interval coverage under physical model mismatch remains a critical test [71].

The screened corpus shows a marked evidence gap. Strong probabilistic treatment is concentrated in a minority of Bayesian, particle-filter and ensemble studies, whereas several physically validated structural and machine-learning applications provide point estimates without formal prediction intervals. This imbalance creates a risk that the most operationally mature twins communicate the least about uncertainty. The appraisal therefore distinguishes “physical evidence strength” from “uncertainty completeness” and does not allow high accuracy alone to compensate for absent uncertainty analysis.

3.6 Identifiability, Observability, and Sensor Informativeness

A parameter can be updated only if changes in that parameter produce distinguishable changes in the observed response. Identifiability is a property of the model, excitation and data combination; observability is the corresponding state-reconstruction condition for dynamic systems. In practical Digital Twins, both are local and operating-condition dependent. Two stiffness parameters may be distinguishable under one load case and inseparable under another. Similarly, a thermal field may be observable from sparse sensors only when the reduced model and boundary-condition ensemble span the actual system behaviour.

Sensitivity analysis, Fisher-information measures, posterior correlation, rank tests and experimental-design methods should therefore precede large-scale updating. The cable-truss and gearbox studies use sensitivity to restrict the updated parameter set [76, 79], while the machine-tool ensemble study uses independent validation sensors and multiple physical models to address boundary-condition uncertainty [80]. A separate physical machine-tool study demonstrates that singular-value analysis and Group LASSO can reduce a manually selected 22-sensor layout to seven sensors while retaining substantial compensation accuracy [100]. This evidence sharpens an important distinction: observability-oriented sensor design enables updating, but it is not itself an observation-driven model update. Sensor placement should therefore be evaluated against the intended updating target, not merely against signal magnitude or reconstruction error [101–105].

The evidence supports three minimum reporting requirements. First, authors should justify why the selected observations are informative for the quantities being updated. Second, posterior or estimated parameter correlation should be reported when multiple parameters are inferred. Third, a physically plausible parameter estimate should be distinguished from a model-compensating estimate that only improves output fit. These requirements are especially important when model discrepancy is omitted, because calibration parameters may otherwise absorb structural error.

Table 3. Corpus A evidence map for the 23 physically grounded studies screened through 28 July 2026.

Study	Domain/scale	Updating target	Method/regime	Evidence
Ward et al. [67]	Facility/field	Parameters	Particle filter; continuous	Operational data; Strong (14/16)
de Angelis et al. [68]	Laboratory structure	Parameters	Imprecise probability; online	Physical tests; Strong (13/16)
Edington et al. [69]	Two-tank lab system	Model representation	Hybrid weighting; periodic	Physical experiment; Moderate (12/16)
Kessels et al. [71]	Ventilation system	Parameters	Bayesian inverse map; real-time	Real measurements; Strong (14/16)

Study	Domain/scale	Updating target	Method/regime	Evidence
Mücke et al. [72]	Wave-tank/benchmarks	States + parameters	Latent particle filter; online	Physical + numerical; Strong (16/16)
Titscher et al. [73]	Bridge demonstrator	Parameters + damage	Bayesian variational; batch	Physical scenarios; Strong (16/16)
Donato et al. [75]	Pilot furnace	States + parameters	Kalman/POD-Kriging; recursive	Experimental furnace; Strong (13/16)
Shi et al. [76]	Cable-truss lab	Geometry/parameters/BCs	Optimization FEMU; batch	Multiple tests; Moderate (10/16)
Tang et al. [77]	Laboratory bridge	Parameters + representation	GA-FEMU + surrogate; periodic	Physical experiments; Strong (14/16)
Ali et al. [78]	100 m field tower	Material + BCs	OMA-FEMU; periodic	Operational sensors; Moderate (10/16)
Xia et al. [79]	Gearbox test rig	Parameters	Surrogate/metaheuristic; periodic	Physical vibration; Moderate (12/16)
Lang et al. [80]	Machine-tool bench	Thermal states	Ensemble Kalman; online	Independent sensors; Strong (15/16)
Zhuo and Yan [81]	Aircraft flight	Inputs/parameters/discrepancy	Residual feedback; adaptive	Flight evidence; Moderate (11/16)
Kosmaga et al. [82]	Mobile robot	States + parameters	Adaptive observer; real-time	Physical robot; Strong (13/16)
Zhao et al. [83]	Battery experiments	States + parameters	TCN-LSTM/transfer; periodic	Experimental datasets; Strong (13/16)
Wu et al. [84]	Power transformer	Model representation	Online kernel learning	Operational data; Moderate (10/16)
Wang et al. [85]	Production data halls	CFD parameters	Neural-surrogate calibration	Field sensors; Strong (13/16)
Wang et al. [86]	Industrial process	States/parameters/structure	EKF-SINDYc; online	Operational data; Strong (14/16)
Park et al. [87]	Vehicle fleet/trips	Parameters	Hierarchical calibration	Real trips; Moderate (11/16)
Lang et al. [95]	Machine-tool bench/machine	Thermal states	Adaptive Kalman observer; online	Independent physical sensors; Strong (14/16)
Zimmermann et al. [97]	Five-axis machine tool	Parameters + model structure	Triggered re-identification; event-based	Physical compensation loop; Moderate (12/16)
Nóvoa et al. [89]	Annular combustor lab	States + parameters + bias/shift	Bias-regularized EnKF + ESN; real-time	Raw physical microphones + public code; Strong (16/16)
Li et al. [88]	Field-informed shield tunnel	States + parameters + BCs/degradation	EnKF-SuS; sequential	Field + controlled scenarios; Moderate (12/16)

4 Offline and Batch Model Updating

Offline and batch methods estimate model corrections from a fixed set of observations. They remain central to Digital Twin development because many engineering assets evolve slowly relative to the time required for calibration, and because expensive models can be updated during commissioning, inspections or scheduled maintenance. Batch methods also provide the foundation for sequential approaches: the objective function, parameterization, priors and discrepancy assumptions established offline frequently determine whether later online inference is stable.

4.1 Deterministic Inverse Updating

A common deterministic formulation minimizes the mismatch between measured and simulated responses subject to regularization or physical constraints. For observations y and model predictions h(theta), a weighted objective can be written as

$$J(\boldsymbol{\theta}) = [\mathbf{y} - \mathbf{h}(\boldsymbol{\theta})]^{\mathsf{T}}\mathbf{W}[\mathbf{y} - \mathbf{h}(\boldsymbol{\theta})] + \lambda R(\boldsymbol{\theta}) \tag{1}$$

where W weights the observations, R(theta) penalizes implausible or unstable solutions and lambda controls the regularization strength. The formulation is attractive because it is transparent and compatible with gradient-based optimization, sensitivity analysis and established finite-element software. Its apparent simplicity is deceptive: the solution depends on parameter scaling, objective selection, sensor noise, model discrepancy and the shape of the feasible region.

Finite-element model updating is the most mature engineering implementation of deterministic calibration [59, 60]. The cable-truss, laboratory bridge and field-tower studies update stiffness-related, geometric or boundary-condition quantities using static or modal measurements [76–78]. Their physical evidence is valuable because it demonstrates that updating can correct as-built deviations and changed conditions. However, deterministic agreement should not be interpreted as unique physical identification. Several parameter combinations may reduce modal or strain errors, particularly when only a small number of modes or response locations are measured.

4.2 Sensitivity, Regularization, and Parameter Selection

Parameter selection is often more important than the optimizer. Updating every uncertain quantity creates an underdetermined problem and encourages compensation between parameters. Local sensitivity matrices, global sensitivity indices and correlation analysis can identify parameters that are both influential and distinguishable. The gearbox study combines sensitivity analysis with a response-surface approximation before metaheuristic optimization, while the cable-truss study restricts the update to parameters that materially influence the selected structural responses [76, 79]. These steps reduce cost and improve numerical conditioning, although they do not by themselves prove global identifiability.

Regularization incorporates prior engineering knowledge into a deterministic problem. Typical forms penalize deviation from nominal parameters, impose spatial smoothness, restrict damage to sparse regions or enforce physical bounds. Regularization should be selected according to the anticipated correction mechanism. A smoothness penalty is defensible for distributed material fields but may suppress a localized crack. A nominal-value penalty is useful when manufacturing tolerances are known, but it can bias the solution if the physical asset has genuinely degraded. Reporting only the final objective value is therefore insufficient; the influence of the regularization term should be demonstrated through sensitivity or ablation analysis.

4.3 Bayesian Calibration and Model Updating

Bayesian calibration represents the unknown parameters probabilistically. Given observations y, the posterior distribution is

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta}) \tag{2}$$

where $p(\boldsymbol{\theta})$ is the prior and $p(\mathbf{y} \mid \boldsymbol{\theta})$ is the likelihood. The posterior communicates both the parameter estimate and its uncertainty, and it permits propagation of uncertainty into predictions. The bridge-demonstrator study uses variational Bayesian inference to calibrate a high-dimensional finite-element model and detect damage while retaining computational feasibility [73]. The continuous farm study compares particle filtering with static and sequential Kennedy–O'Hagan calibration, illustrating how posterior updating can be embedded in a recurrent Digital Twin workflow [67]. The Ford application introduces a hierarchical structure that separates parameters shared across vehicles or trips from trip-specific quantities [87].

A central challenge is model-form discrepancy. If the observation model is written as y = h(theta) + delta_M + epsilon, both theta and the discrepancy delta_M can explain the same residual. Flexible discrepancy models may protect predictions from structural bias but weaken physical parameter identification. Orthogonal or modular discrepancy strategies attempt to reduce this confounding [74]. The evidence currently remains stronger methodologically than physically: the bridge-bias study uses simulated observations and identifies real-measurement validation as future work. This gap is important because discrepancy functions that perform well on synthetic data may absorb unmodelled operational effects differently in a real asset.

4.4 Metaheuristics and Surrogate-Assisted Updating

Population-based and metaheuristic optimization is frequently selected when the objective is non-convex, discontinuous or generated by commercial simulation software without reliable gradients. Genetic algorithms, firefly methods and related approaches can explore broad parameter domains, but they require many model evaluations and provide limited information about uncertainty or identifiability. Surrogates mitigate the cost by approximating the mapping from parameters to responses. The gearbox study uses a polynomial response surface before firefly optimization, while the timely bridge system combines genetic-algorithm FEM updating with mesh reduction and a radial-basis surrogate [77, 79].

Surrogate-assisted calibration introduces a two-stage error structure: parameter estimates depend on both the physical model and the surrogate. Validation should therefore include surrogate error at the updated parameter region, not only global training accuracy. The Kalibre framework demonstrates a stronger approach for expensive CFD calibration: a thermophysics-informed neural surrogate is iteratively calibrated against production sensor data, and the inferred parameters are transferred back to the CFD model for validation [85]. This return to the governing simulator is crucial because a surrogate can match sensors while violating unobserved physical fields.

4.5 Critical Assessment of Offline Methods

Offline updating is appropriate when decisions are slow, datasets are accumulated over campaigns, or the cost of uncertainty quantification exceeds the available online budget. Its strengths are methodological transparency, compatibility with high-fidelity models and the ability to perform detailed diagnostic analyses. Its limitations are delayed response to change, possible reuse of the same data for calibration and validation, and weak evidence that the calibrated model will remain valid under future operating conditions.

The screened corpus indicates that offline methods achieve some of the strongest physical validation in structural and facility applications, but frequently omit probabilistic uncertainty and runtime reporting. Bayesian methods provide richer uncertainty descriptions, yet their assumptions and discrepancy choices are more difficult to validate. A defensible method-selection rule is therefore not "Bayesian is superior to deterministic" but: use deterministic updating when a low-dimensional, identifiable correction with clear physical bounds is required; use Bayesian updating when uncertainty propagation materially affects decisions and the prior/likelihood can be justified; and use surrogate or metaheuristic methods only when their additional approximation and computational cost are independently evaluated.

Table 4. Comparison of principal offline and batch updating approaches.

Approach	**Best suited to**	**Principal strengths**	**Critical limitations**	**Minimum validation evidence**
Gradient/sensitivity-based optimization	Differentiable models with moderate parameter dimension.	Efficient, interpretable, compatible with adjoints.	Local minima; derivative and scaling sensitivity.	Independent response quantities and parameter-sensitivity analysis.
Regularized inverse updating	Ill-posed spatial or multi-parameter problems.	Encodes engineering constraints and stabilizes solutions.	Regularization can dominate the physical evidence.	Regularization sensitivity and recovery tests.
Bayesian calibration	Decisions requiring posterior uncertainty.	Probability distributions and uncertainty propagation.	Prior/likelihood dependence; calibration–discrepancy confounding.	Posterior predictive checks and held-out physical validation.
Variational Bayesian inference	High-dimensional posterior approximation.	Lower cost than sampling and fast repeated inference.	Approximation bias and under-dispersion.	Comparison with sampling or benchmark posterior where feasible.
Metaheuristic optimization	Black-box, non-smooth or multimodal objectives.	Broad search and implementation simplicity.	High evaluation cost; no natural UQ; stochastic variability.	Repeated runs, convergence analysis and parameter plausibility.
Surrogate-assisted calibration	Expensive CFD/FEM or many-query calibration.	Major cost reduction and easy optimization.	Surrogate extrapolation and compounded model error.	Local surrogate validation and return-to-simulator confirmation.

5 Online and Sequential Data Assimilation

Sequential data assimilation updates a model as observations arrive. The generic dynamic system is represented by a state-evolution equation and an observation equation:

$$\mathbf{x}_{k+1} = F(\mathbf{x}_k, \boldsymbol{\theta}_k, \mathbf{u}_k, M_k) + \boldsymbol{\eta}_k \tag{3}$$

$$\mathbf{y}_k = H(\mathbf{x}_k, \boldsymbol{\theta}_k, \mathbf{u}_k, M_k) + \boldsymbol{\epsilon}_k \tag{4}$$

Here $\mathbf{x}_k$ denotes states, $\boldsymbol{\theta}_k$ parameters, $\mathbf{u}_k$ unknown or prescribed inputs, M_k the model representation, $\boldsymbol{\eta}_k$ process uncertainty and $\boldsymbol{\epsilon}_k$ measurement uncertainty. Sequential inference estimates the filtering distribution $p(\mathbf{x}_k, \boldsymbol{\theta}_k \mid \mathbf{y}_{1:k})$, or a broader joint distribution including inputs and discrepancy. The method must balance three competing requirements: fidelity to the engineering model, consistency with noisy observations and computation before the next decision deadline. Unified Bayesian treatments of filtering and smoothing provide the broader probabilistic basis for these sequential formulations [106, 107].

5.1 Kalman-Filter Family

The classical Kalman filter is optimal for linear systems with Gaussian process and observation noise. Engineering Digital Twins are rarely exactly linear, but the Kalman structure remains attractive because it produces recursive state estimates and covariances with low computational overhead. The extended Kalman filter linearizes nonlinear dynamics locally, while the unscented Kalman filter propagates selected sigma points to represent nonlinear transformations. Ensemble Kalman methods use a sample of model realizations and are better suited to high-dimensional systems when explicit covariance propagation is infeasible. Iterative ensemble inversion, multiple-data-assimilation smoothers, and operational EnKF reviews clarify the roles of subspace restriction, covariance inflation, localization, and ensemble design [108–110].

The screened physical evidence demonstrates the versatility of this family. The furnace study assimilates temperature and NO measurements into a POD–Kriging reduced combustion model using steady-state and recursive Kalman variants [75]. An adaptive Kalman state observer reconstructs unmeasured thermal fields from sparse sensors on a physical test bench and five-axis machine tool, reducing thermal-error RMSE from 67.4 to 33.3 μm [111]. The later machine-tool study extends this concept to an enhanced ensemble Kalman filter over multiple reduced physical models, thereby representing boundary-condition and model uncertainty while compensating geometric error in real time [80]. The thermoacoustic Digital Twin uses a bias-regularized EnKF to infer acoustic states, time-varying physical parameters, model bias and measurement shift simultaneously from raw microphone data [89]. The shield-tunnel study extends EnKF with subset simulation within a recursive Bayesian framework to update a multiphysics structure–soil model from new observations [88]. The industrial-process study combines extended-Kalman recursive coefficient estimation with sparse nonlinear identification and model-predictive control [86]. These examples show distinct uses of the same inference family: single-model state reconstruction, multi-model uncertainty representation, explicit discrepancy estimation, non-Gaussian sequential calibration, reduced-field updating, and adaptive model-structure estimation.

Kalman covariances should not be interpreted automatically as credible uncertainty. Their validity depends on the process and measurement models, covariance tuning, ensemble construction and linearization. In practice, covariance matrices are often selected heuristically to stabilize the filter. A strong Digital Twin paper should report the physical meaning or tuning method of these matrices, test sensitivity to their values and evaluate innovation residuals. Independent sensors, as used in both machine-tool state-observer studies, provide strong evidence because they test field reconstruction beyond assimilated measurements [80, 111]. The thermoacoustic study goes further by estimating structured model bias and measurement shift rather than forcing all residuals into state or parameter covariance [89]. The methodological progression from a single reduced-model observer [111], to an ensemble of physically plausible reduced models [80], to bias-aware state–parameter–discrepancy inference [89] shows how distinct uncertainty sources can be separated rather than conflated.

5.2 Particle Filters and Non-Gaussian Assimilation

Particle filters approximate the filtering distribution with weighted samples and can represent nonlinear, non-Gaussian and multimodal uncertainty. Their generality is valuable for damage states, parameter ambiguity and strongly nonlinear processes, but the number of particles required grows rapidly with state dimension. The continuous farm study shows that particle filtering can outperform or simplify repeated Bayesian calibration in a low-dimensional operational model, although an update with 10,000 particles still requires minutes [67]. High-dimensional particle-filter reviews show that proposal design, localization, hybridization, and adaptive resampling are central to avoiding degeneracy [112].

The deep latent-space particle filter addresses dimensionality by learning a compact representation of the state and its dynamics before applying particle filtering [72]. The method combines a Wasserstein autoencoder, learned latent dynamics and particle assimilation, producing non-Gaussian state and parameter distributions. Its evaluation includes a real wave-tank experiment and reports orders-of-magnitude acceleration relative to high-fidelity filtering. This architecture illustrates a broader trend: reduced or learned representations are becoming part of the assimilation algorithm rather than only fast forward models. The resulting uncertainty is credible only if the latent representation preserves the observation-sensitive dynamics and remains valid under operational changes.

5.3 Joint State, Parameter, and Input Estimation

A Digital Twin often needs more than state estimation. Parameters may drift through wear, inputs may be unmeasured and model coefficients may change between regimes. One strategy augments the state vector with parameters or inputs. Another uses dual filters, alternating between state and parameter updates. Augmentation is simple but can create weakly observable, high-dimensional systems. Dual estimation separates time scales but may propagate errors between filters.

Simulation-based wind-turbine work demonstrates augmented estimation of states and unmeasured loads, but it remains in the methodological corpus because the published applications use virtual measurements [113]. The latent particle-filter study performs joint state–parameter inference with physical wave observations [72]. The mobile-robot study uses an adaptive nonlinear observer to estimate time-varying parameters and maintain position estimates during data loss [82]. The thermoacoustic twin jointly estimates acoustic states, time-varying parameters, model bias and measurement shift [89], whereas the shield-tunnel framework updates time-invariant and time-varying quantities, boundary conditions and degradation-related states within a sequential Bayesian scheme [88]. These cases emphasize that joint estimation should be justified through observability and excitation. A constant operating condition may provide insufficient information to separate a changing parameter, an unknown input and structured model discrepancy.

5.4 Variational and Moving-Horizon Methods

Variational data assimilation estimates the trajectory that minimizes a cost over an observation window, while moving-horizon estimation repeatedly solves a constrained finite-window optimization problem. These methods are attractive when physical constraints, bounds and non-Gaussian penalties must be imposed explicitly. They can exploit adjoint gradients and handle delayed or asynchronous observations, but the repeated optimization may be expensive and sensitive to initialization. Corpus A is dominated by filtering rather than variational or moving-horizon implementations. This pattern may reflect either a genuine application gap or incomplete capture despite the method-specific search terms reported in Section 2.2. Comparative process-estimation studies also document the trade-off between EKF efficiency, constrained moving-horizon estimation, and particle-based posterior representation [114, 115].

Comparative data-assimilation research shows that no single method dominates across system nonlinearity, observation density, noise structure and computational constraints [116]. Digital Twin reviews should therefore avoid ranking algorithms independently of the problem. An ensemble filter may be appropriate for a large field with approximately Gaussian uncertainty, while a particle method may be preferable for a low-dimensional multimodal damage state. Moving-horizon estimation may be favoured when hard constraints and recent-data windows are central to control.

5.5 Real-Time Feasibility and Latency Reporting

Real-time capability is a system property, not an algorithm label. The relevant quantity is the end-to-end latency from measurement acquisition to an updated engineering output. This latency includes sensor transmission, preprocessing, inference, model propagation, uncertainty calculation and communication to the decision layer. Reporting only neural-network inference time or only the reduced model solve time is insufficient.

The current evidence ranges from millisecond inverse-map inference to multi-minute particle-filter updates and multi-hour CFD calibration [67, 70, 80, 85]. All can be operationally acceptable for different decisions. The machine-tool twin is exemplary because it relates computational speed to the measurement time scale and demonstrates approximately 47-times-faster-than-real-time execution [80]. The thermoacoustic twin assimilates raw microphone observations at approximately 1707 Hz and then predicts autonomously beyond the assimilation window [89]. Kalibre requires about five hours, but its target is periodic calibration of production data-hall CFD models rather than closed-loop millisecond control [85]. The shield-tunnel work is designed for parallel sequential updating, but the accessible record does not provide a complete measurement-to-update latency [88]. Accordingly, EDTURC-24 requires reporting of hardware, data-window length, model-solve time, inference time, total latency, update frequency and the maximum latency allowed by the use case.

Table 5. Sequential assimilation methods and conditions governing their suitability.

Method	Assumptions/ representation	Advantages	Limitations	Representative evidence
Kalman filter	Linear dynamics; Gaussian uncertainty.	Low cost and exact recursive covariance under assumptions.	Inadequate for strong nonlinearity or multimodality.	Foundational state/input estimation and physical machine-tool state reconstruction [95, 96].
Extended Kalman filter	Locally differentiable nonlinear model.	Efficient adaptive state/parameter estimation.	Linearization bias and Jacobian requirements.	Industrial EKF-SINDYc updating [86].
Unscented Kalman filter	Moderately nonlinear, approximately Gaussian system.	No explicit Jacobian; improved nonlinear moment propagation.	Cost grows with augmented dimension.	No physically grounded UKF study met the Corpus A eligibility threshold in the documented search.
Ensemble Kalman filter	High-dimensional fields with approximately Gaussian ensemble statistics.	Scalable covariance approximation and multi-model ensembles.	Sampling error, localization/tuning and Gaussian update.	Machine-tool ensemble compensation, real-time thermoacoustic bias-aware assimilation and shield-tunnel EnKF-SuS [80, 88, 89].
Particle filter	Nonlinear/non-Gaussian, usually lower-dimensional state.	Multimodal posterior and flexible likelihood.	Particle degeneracy and high computational cost.	Continuous calibration and latent-space PF [67, 72].
Variational / moving horizon	Windowed optimization with differentiable or constrained model.	Hard constraints, asynchronous data and trajectory estimation.	Repeated optimization and local minima.	No physically grounded 3D-Var, 4D-Var, or moving-horizon Digital Twin study met the Corpus A eligibility threshold in the documented search.
Adaptive observer / recursive identification	Structured parameter/state dynamics.	Very fast, interpretable and control-oriented.	Limited UQ and dependence on persistent excitation.	Mobile robot, sparse process identification and event-triggered thermal-model updating [82, 86, 97].

6 Reduced-Order, Data-Driven, and Hybrid Updating

High-fidelity engineering models are often too expensive for repeated calibration or assimilation. Reduced-order, surrogate and data-driven methods address this bottleneck, but they also alter the epistemic structure of the Digital Twin. A full-order model has explicit discretization and modelling assumptions. A learned surrogate adds training-domain, architecture and optimization assumptions. Hybrid updating is therefore not simply faster inference; it is an additional model layer that must be calibrated, validated and monitored. Broader hybrid-modelling literature emphasizes that physical constraints, data coverage, and model-form uncertainty must be treated jointly rather than as separate design choices [117, 118].

6.1 Reduced-Order Models and Multi-Fidelity Updating

Projection-based reduced-order models preserve a low-dimensional subspace of the governing solution. Proper orthogonal decomposition is used in the furnace twin, Krylov reduction in the machine-tool twin, and mesh/surrogate reduction in the timely bridge system [75, 77, 80]. These methods retain a recognizable relationship to the original equations, which facilitates engineering interpretation and boundary-condition changes. They are particularly suitable for filtering because the state covariance or ensemble is propagated in a much smaller space. Classical reduced-basis theory and newer nonlinear-manifold reduced-order models provide complementary foundations for this offline-online decomposition [119, 120].

The principal limitation is domain dependence. A basis extracted from nominal thermal or flow fields may omit the modes activated by damage, new boundary conditions or a new operating regime. Online basis enrichment, switching between local reduced models and error indicators are therefore important research directions. Multi-fidelity Digital Twins should report not only the error of each model level, but also the rule by which the system selects or corrects a level during operation.

6.2 Surrogate-Assisted Calibration and Inverse Mapping

A forward surrogate approximates the model response h(theta), enabling rapid optimization or sampling. An inverse surrogate approximates the mapping from response features to parameters. Kalibre is a forward-surrogate calibration architecture in which a thermophysics-informed neural model reduces the number of CFD evaluations while preserving final validation in the CFD simulator [85]. The gearbox and bridge cases use polynomial or radial-basis surrogates within optimization loops [77, 79]. These examples demonstrate that surrogate validation should be local to the calibration region and should be followed by confirmation in the governing model [121].

Inverse mappings can provide extremely fast online estimates because the expensive parameter search is learned offline. The deterministic Kessels study achieves millisecond inference but is validated only with simulated experiments [70]. Its probabilistic extension incorporates Bayesian neural-network uncertainty and is evaluated on real ventilation measurements [71]. The comparison is instructive: a method can be computationally mature before its physical evidence matures, and real observations may expose model mismatch that is invisible in synthetic test sets. Inverse updating should therefore include out-of-distribution detection and a fallback mechanism when the measured feature vector lies outside the simulation-supported domain. Modern surveys of neural uncertainty further underline the need to distinguish data, model, distributional, and approximation uncertainty [122].

6.3 Sparse Identification, Online Learning, and Continual Adaptation

Sparse identification seeks compact governing equations from data by selecting active terms from a candidate library. Recursive updating can adapt these terms or their coefficients as the process changes. The EKF-SINDYc process twin combines sparse nonlinear identification, extended-Kalman recursion and model-predictive control, producing quantitative improvements on industrial data [86]. Event-triggered online learning provides a complementary architecture: the Thermal Adaptive Learning Control system monitors physical residuals, declares a No-Good condition when action limits are exceeded, collects a high-frequency measurement campaign and autonomously re-identifies the thermal-error model before returning it to compensation service [123]. The first approach continuously tracks a structured dynamic model; the second avoids unnecessary recalibration until evidence of divergence is detected. Both are more interpretable than unconstrained continual neural adaptation, but neither eliminates the need for rollback, update validation and drift monitoring.

Online learning also appears in transformer temperature prediction, battery state estimation and mobile-robot dynamics [82–84]. These applications update neural or adaptive-observer representations using new operational or experimental data. Their central risk is stability–plasticity: the model must adapt to genuine change without forgetting previously valid behaviour or amplifying sensor anomalies. Validation should therefore be longitudinal. A single before/after accuracy value cannot reveal catastrophic forgetting, delayed drift or performance under a return to an earlier operating regime. Concept-drift literature provides explicit taxonomies for detection, characterization, and adaptation in nonstationary data streams [124].

Replay-based battery datasets occupy an intermediate evidence level. The observations are physical and independent of the model, but the Digital Twin is not necessarily connected live to the specific cell. Such studies should be labelled as experimental-data Digital Twin demonstrations rather than operational deployments. This distinction allows the review to recognize valuable physical evidence without overstating synchronization maturity.

6.4 Physics-Informed Learning and Neural Operators

Physics-informed neural networks and neural operators incorporate governing equations, conservation relationships or simulator structure into learning. They are promising for Digital Twins because they can reconstruct unmeasured fields, solve inverse problems and generalize across parameters more efficiently than pointwise surrogates. The additive-manufacturing neural-operator framework couples a multiphysics melt-pool model, Fourier neural operators, calibration and uncertainty components [63]. The optical-fibre study uses a differentiable propagation model and physics-informed loss for parameter estimation [66]. The self-updating thermal PINN introduces physics-stabilization to prevent drift across update cycles [65].

In the strict evidence assessment, all three remain in Corpus B because the reported demonstrations rely primarily on simulations or analytical reference data. This does not reduce their methodological importance; it identifies the next validation requirement. Physical sensor assimilation should test whether the physics constraints are sufficient when the governing model itself is incomplete. In addition, neural-operator uncertainty should be evaluated under parameter and geometry shifts, not only within the training distribution.

6.5 Residual and Model-Discrepancy Learning

Residual learning corrects the difference between model predictions and observations rather than replacing the engineering model. It can preserve interpretability and reduce the data required for learning. The aircraft study implements a measurement–calibration–residual-feedback–correction loop using multidimensional flight information [81]. The thermoacoustic study provides a more explicit decomposition: a reservoir computer estimates model bias and measurement shift, while a bias-regularized EnKF updates states and physical parameters from raw measurements [89]. These approaches are significant because they treat model error as an evolving operational quantity rather than a fixed calibration residual.

Residual correction can nevertheless hide physical deterioration or parameter changes if the correction term is too flexible. A discrepancy model should therefore be decomposed where possible into interpretable sources: input error, parameter drift, sensor bias or shift, numerical error and remaining structural discrepancy. Bayesian Gaussian-process discrepancy methods provide a probabilistic formulation, but current bridge evidence is largely simulation-based [74]. The bias-regularized thermoacoustic architecture demonstrates that physically constrained discrepancy estimation can operate in real time and generalize to an unseen operating condition [89]. Combining such structured residual models with long-duration field data is one of the clearest open research directions emerging from the synthesis.

6.6 Cross-Method Synthesis and Selection Principles

The choice between reduced-order, surrogate, data-driven and hybrid updating should be based on the information and latency structure of the problem. Projection-based reduction is appropriate when a trusted full-order model exists and the dominant physical subspace is stable. Forward surrogates are useful when repeated model evaluation is the principal bottleneck. Inverse surrogates are appropriate when the parameter domain can be comprehensively simulated and millisecond inference is required. Sparse identification is valuable when interpretable, control-oriented dynamic equations are sought. Physics-informed learning is promising when governing constraints are reliable but data are sparse. Residual learning is appropriate when the physics model captures the dominant behaviour but retains systematic bias.

Three safeguards apply across all hybrid methods. First, the training and calibration domain must be stated in physical terms, and an out-of-domain indicator should accompany predictions. Second, the hybrid model should be validated on observations not used to select its architecture or weights. Third, the final engineering quantity should be checked against the original high-fidelity model or physical system whenever possible. The strongest studies in the screened corpus follow at least two of these principles, while weaker studies report only surrogate test error or synthetic reconstruction accuracy.

Table 6. Reduced-order, data-driven, and hybrid updating strategies.

Strategy	Online role	Evidence advantages	Principal failure mode	Recommended safeguard
Projection ROM	Fast state propagation and filtering.	Retains explicit connection to governing equations.	Basis invalid outside sampled regimes.	Residual error indicator and basis enrichment.

Strategy	Online role	Evidence advantages	Principal failure mode	Recommended safeguard
Forward surrogate	Fast optimization, sampling and calibration.	Can be returned to the original simulator for confirmation.	Local fit used outside calibration region.	Local validation near inferred parameters.
Inverse mapping	Direct parameter/state inference.	Millisecond online prediction after offline training.	Non-unique inverse and simulation-to-real mismatch.	Probabilistic output, OOD detection and fallback solver.
Sparse identification	Adaptive interpretable dynamics and control.	Compact equations and low online cost.	Candidate-library bias and unstable term switching.	Persistence tests, coefficient regularization and physical constraints.
PINN/neural operator	Field reconstruction and parametric solution maps.	Embeds differential constraints and scales across states.	Physics model remains incomplete; weak physical validation.	Experimental assimilation and extrapolation tests.
Residual/discrepancy model	Correct systematic engineering-model bias.	Preserves a physics baseline and targets the remaining error.	Correction absorbs damage or wrong parameters.	Decompose error sources and constrain residual structure.
Transfer/continual learning	Adaptation to ageing and new regimes.	Reuses prior models with limited new data.	Catastrophic forgetting and drift amplification.	Longitudinal validation and rollback/freeze mechanisms.

Across Corpus A, high evidence scores occur across multiple method families. The common determinants are a complete physical-observation loop, a clearly defined updating target, independent or changed-condition validation, explicit uncertainty and latency reporting, and reproducible implementation detail. Algorithmic sophistication, optimized sensing, or one-time physical identification alone does not establish recurrent Digital Twin synchronization.

7 Evidence-Based Method Comparison and Selection

Method selection for an engineering Digital Twin is a constrained inference-design problem rather than a ranking of algorithms by nominal accuracy. The same method may be appropriate for one twin and indefensible for another because the updating target, data cadence, trigger policy, model cost, uncertainty structure and consequence of error differ. The screened evidence corpus reinforces this point: Strong studies are distributed across Kalman filtering, particle filtering, Bayesian calibration, finite-element updating, inverse machine learning and surrogate-assisted calibration [67, 68, 72, 73, 75, 77, 80, 82, 83, 85, 86, 111]. Event-triggered model re-identification contributes a distinct Moderate-evidence pathway that is valuable when recalibration should occur only after statistically or operationally meaningful divergence [123]. No method family dominates when evidence quality, physical validation and computational feasibility are considered jointly.

The selection process should therefore begin with the context of use: the engineering decision supported by the twin, the tolerated prediction error, the required update interval and the risk associated with an incorrect result. This is consistent with credibility frameworks that relate the required verification and validation evidence to how strongly a decision relies on model output [18, 99, 125]. A method that is adequate for trend monitoring may be inadequate for closed-loop control, remaining-life certification or safety-critical intervention.

7.1 Decision Variables Governing Method Choice

Nine decision variables repeatedly distinguish successful applications from nominal demonstrations. First, the updating target determines whether the problem is filtering, calibration, input reconstruction, model selection or discrepancy correction. Second, system nonlinearity and posterior geometry determine whether Gaussian approximations are credible. Third, the dimension of the state and parameter spaces governs whether particle-based or full Bayesian methods are computationally feasible. Fourth, observation cadence and latency determine whether inference must be recursive. Fifth, the cost of the forward model determines whether reduction or surrogate modelling is necessary. Sixth, data quantity and excitation govern identifiability. Seventh, interpretability requirements constrain the admissible use of black-box learning. Eighth, model-form error determines whether parameter adjustment alone is meaningful. Ninth, the consequence of decision error determines the required credibility evidence.

Table 7. Decision variables controlling the selection of an updating method.

Decision variable	Relatively benign condition	Demanding condition	Methodological response
Updating target	Single state or small parameter vector	Joint states, parameters, inputs, bias or model structure	Match the inference architecture to the target; avoid estimating non-identifiable quantities jointly without additional information.
Dynamics and nonlinearity	Linear or weakly nonlinear dynamics	Strong nonlinearities, discontinuities or switching behaviour	Use EKF/UKF cautiously; consider EnKF, particle filtering, moving-horizon estimation or hybrid observers.
Posterior geometry	Approximately unimodal and Gaussian	Multimodal, bounded, skewed or heavy-tailed	Use particle methods, imprecise probability or robust/nonparametric inference rather than covariance-only summaries.
Dimension	Low-dimensional state and parameter vectors	High-dimensional fields or many uncertain parameters	Apply reduced-order, latent-space, ensemble or variational representations; verify that reduction preserves observability.
Observation cadence	Infrequent experiments or inspections	Streaming sensors and rapidly drifting conditions	Use batch calibration for the former and recursive filtering, identification or continual learning for the latter.
Forward-model cost	Milliseconds to seconds per solve	High-fidelity FE/CFD or multiphysics simulations	Use ROMs, surrogates or multi-fidelity hierarchies, followed by high-fidelity spot checks.
Interpretability	Prediction only	Physical diagnosis, certification or control authority	Prefer identifiable physical parameters, sparse equations or constrained residuals over unrestricted black-box adaptation.
Model discrepancy	Small residual after calibration	Persistent structured residuals or extrapolation bias	Introduce explicit discrepancy/residual models and prevent them from absorbing physical deterioration.
Decision risk	Advisory visualization	Safety-, cost- or mission-critical action	Increase independent validation, UQ, traceability, monitoring and conservative applicability limits.

7.2 Comparative Suitability of Principal Method Families

Deterministic and finite-element model updating remain appropriate when the model is interpretable, the number of update parameters is limited and repeatable experiments provide informative responses. They are comparatively transparent and compatible with established structural-dynamics workflows, as illustrated by cable-truss, bridge, tower and gearbox studies [76–79]. Their weakness is not the optimization itself but the frequent absence of uncertainty, non-uniqueness analysis and validation outside the calibration condition.

Bayesian calibration is preferable when posterior uncertainty is a required output and the cost of repeated model evaluation is manageable or can be reduced. The bridge demonstrator and continuous-calibration studies show the advantages of posterior distributions and explicit measurement uncertainty [67, 73]. Bayesian methods are not automatically credible: an expressive discrepancy model can confound parameters, and a posterior conditioned on calibration data is not independent validation. Variational or approximate inference can make high-dimensional calibration tractable, but approximation error should be reported.

Kalman-family methods are particularly effective for sequential state estimation when dynamics are known and uncertainty can be represented by means, covariances or ensembles. The furnace, single-model machine-tool observer, ensemble machine-tool twin, industrial-process, thermoacoustic and shield-tunnel studies demonstrate how filtering can connect physical observations to operational outputs [75, 80, 86, 88, 89, 111]. EnKF and particle methods are preferable as nonlinearity, dimensionality or non-Gaussianity increases. The latent-space particle filter demonstrates that high-dimensional assimilation becomes feasible when reduction and inference are designed together [72]. The thermoacoustic study demonstrates the additional value of explicit bias regularization, while EnKF-SuS addresses non-Gaussian recursive updating in infrastructure models [88, 89]. Event-triggered re-identification is preferable when the model remains adequate for long periods and new high-rate measurements should be acquired only after residual deterioration [123].

Data-driven inverse maps and online-learning models are effective when the forward problem can generate a representative training domain or when long operational histories are available. Their advantage is low online latency; their principal risk is unobserved extrapolation. The mechanical-ventilation, robot, battery, transformer and data-centre studies demonstrate distinct versions of this pattern [71, 82–85]. The model must carry an applicability indicator, and physical measurements should be used to test—not merely train—

the inverse relationship.

Table 8. Evidence-based comparison of updating-method families.

Method family	Best suited to	Principal strengths	Principal limitations	Representative evidence
Deterministic/ FEMU	Small to moderate parameter sets; interpretable physics; batch tests	Transparent objectives; direct engineering parameters; established sensitivity tools	Point estimates; local minima; parameter correlation; weak UQ in many applications	Cable truss, laboratory bridge, JRC tower [76–78]
Bayesian calibration	Sparse experiments; posterior uncertainty; parameter/bias inference	Posterior distributions; prior information; formal noise treatment	Computational cost; prior sensitivity; parameter–discrepancy confounding	Continuous farm calibration, bridge demonstrator, Ford vehicles [67, 73, 87]
Kalman/EKF/ UKF	Recursive state estimation with moderate nonlinearity	Low latency; recursive covariance or ensemble propagation; mature implementation	Gaussian/local-linear assumptions; tuning, sampling and model-bias sensitivity	Furnace, machine tools, industrial process, thermoacoustics and shield tunnel [75, 80, 86, 88, 89, 95]
EnKF/particle filters	Nonlinear, high-dimensional or non-Gaussian sequential inference	Ensemble/non-Gaussian uncertainty; joint estimation	Sampling cost; degeneracy; localization/reduction choices	Latent particle filter, continuous particle calibration, thermoacoustic r-EnKF and tunnel EnKF-SuS [67, 72, 88, 89]
ROM/surrogate-assisted update	Expensive FE/CFD/multiphysics forward models	Large speedups; supports optimization and sequential inference	Reduction error; limited training domain; surrogate uncertainty	Bridge, furnace and data-centre twins [75, 77, 85]
Inverse ML/online learning	High-rate estimation and large simulation/operational datasets	Very low online cost; flexible nonlinear mappings	Simulation-to-real mismatch; non-unique inverse mappings; out-of-distribution failure; limited interpretability; and continual-learning drift.	Ventilation, robot, battery and transformer studies [71, 82–84]
Sparse identification	Interpretable adaptive dynamics and control	Compact equations; coefficient tracking; control integration	Library dependence; term-selection instability; limited discrepancy treatment	Industrial EKF-SINDYc application [86]
Residual/ discrepancy update	Trusted dominant physics with persistent structured error	Corrects systematic bias; supports hybrid fidelity	Can conceal deterioration; attribution and extrapolation are difficult	Aircraft residual feedback, bridge-bias foundations and thermoacoustic bias/shift estimation [74, 81, 89]

7.3 A Structured Selection Framework

Figure 3 presents a practical selection pathway. The first branch separates dynamic state/input estimation from calibration of parameters, structure or model bias. The second branch assesses whether Gaussian and low-dimensional assumptions are defensible. The third assesses whether the forward model is sufficiently inexpensive for direct inference. The final branch is not an algorithmic choice but a credibility gate: the selected method must be verified, validated on physical evidence, accompanied by uncertainty and monitored after deployment.

For transparent comparison among feasible candidates, a normalized multi-criteria score may be used during method design, provided it is not presented as universal empirical evidence. Let Fm denote physical and statistical fidelity, Um uncertainty adequacy, Lm latency suitability, Im interpretability, Rm reproducibility and Cm computational and implementation cost for method m. A project-specific selection utility can be written as

Equation (5) is an illustrative, project-specific decision aid proposed by the authors. It has not been empirically validated and must not be interpreted as a universal ranking formula; variables and weights require normalization, stakeholder justification, and sensitivity analysis for each context of use.

$$S_m = w_F F_m + w_U U_m + w_L L_m + w_I I_m + w_R R_m - w_C C_m \tag{5}$$

where the weights are defined by the context of use and decision risk. This formulation makes trade-offs

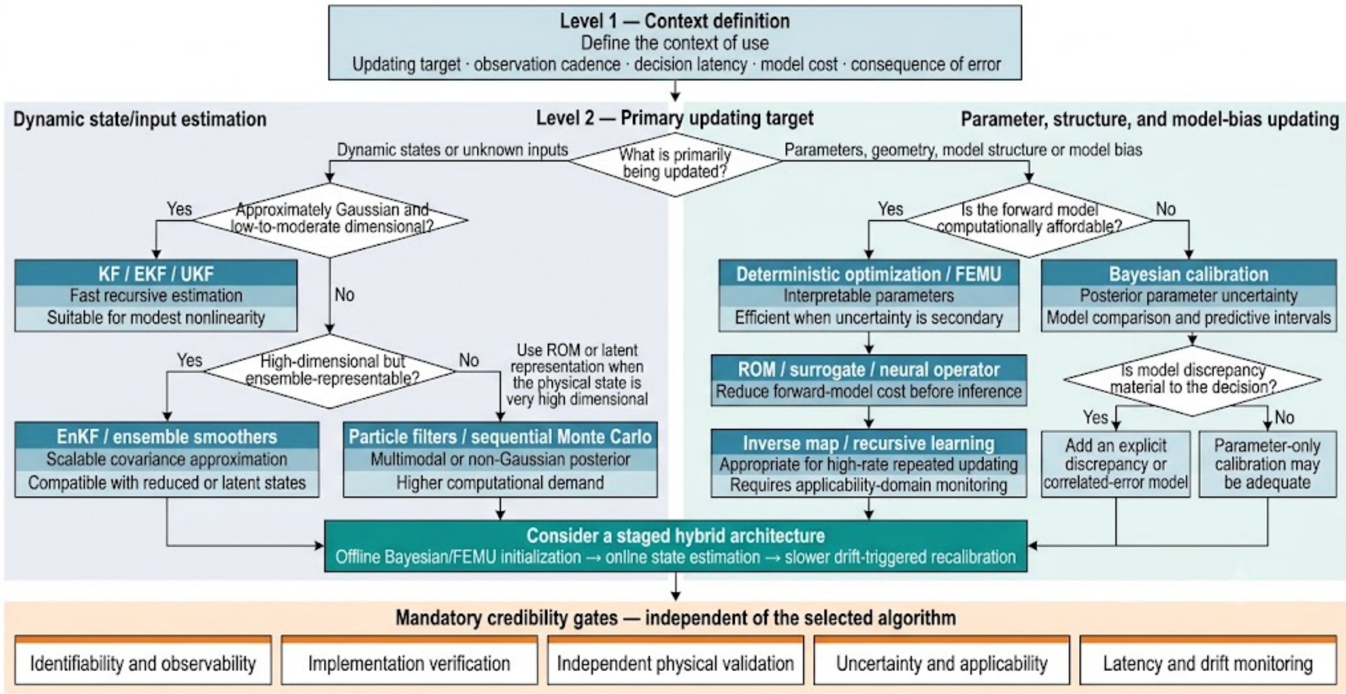

Figure 3: Evidence-based pathway for selecting model-updating and data-assimilation methods. The updating target, posterior characteristics, state dimension, forward-model cost and required decision cadence define an initial method family. The pathway does not prescribe an automatic optimum: identifiability, implementation verification, independent physical validation, uncertainty adequacy, latency and post-deployment monitoring remain mandatory for every method.

explicit: a millisecond inverse model may score highly for latency but poorly for extrapolatory fidelity; a full Bayesian calibration may score highly for uncertainty but poorly for update cadence. The score should therefore support—not replace—engineering judgement and validation [126].

7.4 Multi-Stage and Hybrid Inference Architectures

The strongest operational systems often use more than one updating method. A high-fidelity model may generate an offline training or reduction basis; a surrogate or reduced model performs online inference; a filter assimilates streaming observations; a discrepancy estimator treats persistent bias; and periodic high-fidelity recalibration prevents accumulated surrogate error. The furnace, bridge, machine-tool, data-centre and thermoacoustic examples each implement part of this architecture [75, 77, 80, 85, 89]. This indicates that the meaningful comparison is frequently between inference pipelines rather than isolated algorithms.

A defensible pipeline separates four timescales: fast state correction, intermediate parameter adaptation, slower model-structure revision and episodic high-fidelity revalidation. Updating all quantities at the sensor rate is neither necessary nor statistically stable. State variables may be corrected each measurement cycle, boundary conditions hourly, degradation parameters after informative events and model structure only after persistent residual evidence. This hierarchy reduces confounding and computational cost.

Table 9. Proposed staged workflow for engineering Digital Twin updating.

Stage	Primary objective	Typical methods	Required evidence / gate
1. Define context of use	Specify decision, outputs, error tolerance, latency and risk	Requirements analysis; uncertainty budget	Traceable use case and acceptance criteria
2. Establish baseline model	Verify equations, code, units, interfaces and numerical convergence	Analytical tests; mesh/time-step studies; benchmark problems	Solver/code verification completed before calibration
3. Design observations	Ensure targets are identifiable and operating conditions are informative	Sensitivity, observability, information-gain and sensor-placement analyses	Documented sensor bandwidth, uncertainty and synchronization

Stage	Primary objective	Typical methods	Required evidence / gate
4. Calibrate/update	Infer states, parameters, inputs or discrepancy	FEMU, Bayesian calibration, filters, inverse ML or hybrid pipelines	Convergence diagnostics, uncertainty and parameter plausibility
5. Validate	Test predictions on conditions not used for fitting	Held-out experiments, changed loads, damage or operational periods	Predefined metrics and acceptance thresholds
6. Deploy and monitor	Track latency, residuals, drift and applicability	Residual tests, change detection, scheduled recalibration	Versioned updates, rollback rules and audit trail

8 Cross-Domain Engineering Evidence

The 23-study Corpus A spans structural and geotechnical engineering, engineering dynamics, manufacturing, aerospace, combustion, energy, built environments, automotive systems, robotics and batteries. Fourteen studies meet the Strong evidence threshold and nine are Moderate. This distribution is descriptive of the documented review corpus and should not be interpreted as a prevalence estimate for the entire literature. It nevertheless supports cross-domain comparison of updating targets, validation maturity and engineering value.

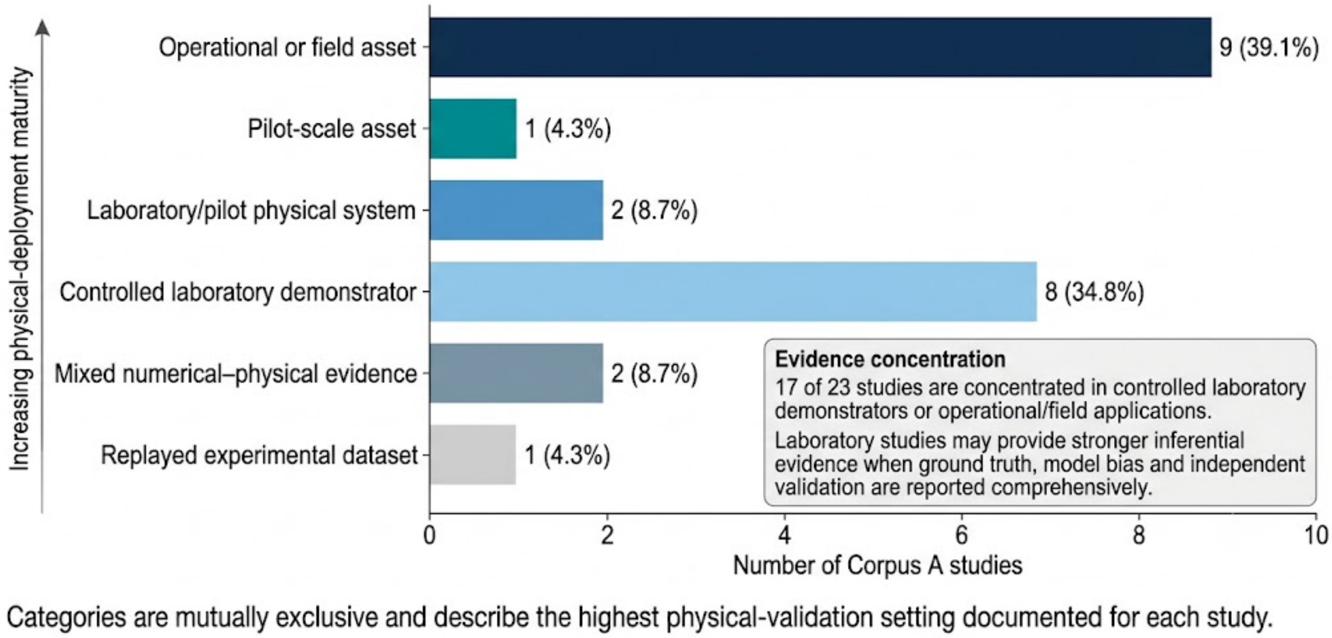

Figure 4: Physical-validation setting of the 23 studies included in Corpus A. Operational or field assets form the largest category, followed closely by controlled laboratory demonstrators. The categories describe the highest physical-validation setting documented for each study and should not be interpreted as a direct ranking of evidence quality, because controlled laboratory studies may provide superior ground truth, independent validation and model-discrepancy characterization.

8.1 Civil, Structural, and Aerospace Systems

Structural applications provide the clearest lineage from conventional model updating to Digital Twins. Physical responses such as modal frequencies, accelerations, displacements, cable forces and strains constrain finite-element or low-order mechanical models. The bridge demonstrator provides the strongest probabilistic case, combining physical damage scenarios, high-dimensional Bayesian variational inference and openly available evidence [73]. Cable-truss, bridge and tower studies show that deterministic FEMU remains effective when parameters are carefully selected and multiple test conditions are available [76–78].

The principal structural limitation is the tendency to equate reduced calibration residual with validation. An updated modal model may reproduce measured frequencies while remaining inaccurate for untested loads, damage modes or nonlinear response. The bridge work in [77] improves this situation by coupling updating with reduced/surrogate models and physical dynamic-load validation. The JRC tower extends evidence to a field-scale asset, although detailed UQ and computational reporting are limited [78]. The shield-tunnel study broadens sequential infrastructure evidence beyond bridges by applying recursive EnKF-SuS updating to a thermo-hydro-mechanical structure–soil model informed by real energy-tunnel observations [88]. Its main reporting limitation is that the accessible evidence does not fully separate field-data validation from synthetic degradation scenarios.

The aircraft study differs by emphasizing multi-dimensional flight parameters and residual-feedback model correction [81]. Its importance lies in the operating-domain challenge: loads and boundary conditions change continuously, and laboratory calibration cannot span the flight envelope. Aerospace twins therefore require explicit input uncertainty, residual monitoring and applicability limits [127]. The remaining challenge is reproducibility because flight data, model details and runtime information are rarely public.

8.2 Manufacturing, Machine Tools, and Industrial Processes

Manufacturing applications demonstrate the strongest connection between updating and immediate operational action. The gearbox study uses physical–virtual signal matching to update dynamic parameters and support fault diagnosis [79]. Two successive machine-tool studies reconstruct thermal fields from sparse sensors: the first uses a Kalman state observer over a reduced finite-element model [111], while the later ensemble architecture combines multiple reduced physical models to represent uncertain boundary conditions and compensate geometric error in real time [80]. A separate event-triggered system monitors residual degradation, acquires new physical training data and autonomously re-identifies the compensation model [123]. The industrial-process study combines EKF-based recursive sparse identification with model-predictive control, linking updating directly to control performance [86]. Together, these studies provide the clearest progression from state estimation, through uncertainty-aware multi-model inference, to autonomous model maintenance.

Computational architecture is central in this domain. Full FE, CFD or process models are often too slow for the operational cycle, so model reduction, ensemble representations and surrogate calibration are not optional enhancements. The machine-tool Digital Twin operates approximately 47 times faster than the measurement timescale, whereas Kalibre requires several hours but updates an expensive CFD representation of production data halls [80, 85]. Both are credible within their contexts because the relevant latency differs: seconds for compensation versus hours for facility recalibration.

Additive manufacturing provides sophisticated examples of physics-informed surrogates and stochastic calibration [128–130], but the strict screening exposes a maturity gap. The LPBF digital-shadow study uses physical evidence and rigorous calibration yet remains a one-way, offline representation [62]. The neural-operator study demonstrates online-capable calibration computationally but not through a completed physical sensor loop [63]. These studies belong in the methodological synthesis, while the application evidence should remain distinct until persistent physical updating is demonstrated.

8.3 Energy, Thermal Systems, and Built Environments

Energy, thermal and combustion applications span markedly different updating timescales. The hydrogen furnace uses physical temperature and emission measurements to assimilate a reduced combustion representation [75]. The azimuthal-thermoacoustic twin assimilates raw sparse microphone data at high rate and jointly estimates physical states, time-varying parameters, model bias and measurement shift [89]. The transformer model applies online learning to changing temperature behaviour [84]. The battery study updates state estimators periodically through transfer learning using experimental cycling histories [83]. The underground-farm model uses particle filtering for continuous environmental calibration [67]. These cases illustrate how the required inference architecture changes from kilohertz instability tracking to minute-, cycle- and facility-scale adaptation.

Built-environment applications also reveal the distinction between operational evidence and synthetic commissioning. Kalibre calibrates CFD models using sensors from two production data halls [85], while the smart-building generative-calibration method uses model-generated observations and therefore remains in Corpus B [64]. The methodological performance of the latter is strong, including probabilistic calibration and sub-second inference, but the evidence does not demonstrate physical synchronization. This distinction should be maintained explicitly in cross-domain comparisons.

8.4 Robotics, Biomedical Dynamics, and General Engineering Systems

Robotic and engineering-dynamics twins emphasize low latency and incomplete observations. The wheeled-robot study estimates static and time-varying parameters in real time and maintains useful position prediction during data loss [82]. The imprecise-probability study updates a low-order structure from repeated laboratory experiments without imposing a precise prior or likelihood [68]. The time-evolving hybrid twin adjusts the balance between physical and data-driven models as new tank-system data arrive [69]. These approaches are especially valuable when a fixed model class cannot remain adequate throughout operation.

The ventilation study demonstrates a different hybrid pattern: physics-generated simulations train a probabilistic inverse model, which is subsequently tested on physical measurements [71]. This separation between simulation training and real-data testing is stronger than purely synthetic validation, but measurements outside the predicted intervals reveal model-class discrepancy. Such cases support the use of explicit out-of-domain and discrepancy indicators rather than ever-wider predictive intervals.

Table 10. Cross-domain synthesis of Corpus A evidence.

Domain cluster	Representative updating targets and methods	Physical evidence and engineering value	Dominant limitations
Civil/structural/ geotechnical/ aerospace	FEM parameters, damage states, loads and residual correction; deterministic and Bayesian updating	Laboratory bridges/cable structures, field tower, shield tunnel and flight measurements; diagnosis, degradation tracking, fatigue and structural prediction [73, 76–78, 81, 88]	Calibration–validation conflation; weak UQ in deterministic studies; restricted data access
Manufacturing/ industrial	Machine states, thermal fields, dynamic coefficients, model structure and CFD parameters; filters, triggered online learning, sparse identification and surrogate calibration	Gearbox test rig, multiple machine-tool systems, industrial process records and production data halls; diagnosis, compensation, autonomous maintenance and control [79, 80, 85, 86, 95, 97]	Proprietary data; incomplete model-form uncertainty; inconsistent latency reporting; physical validation sometimes conflated with updating
Energy/thermal/ combustion/built environment	Environmental, combustion, acoustic, battery and transformer states/parameters; PF, Kalman/EnKF, transfer and online learning	Operational farm, furnace experiments, real-time annular-combustor measurements, battery cycling and transformer records [67, 75, 83, 84, 89]	Replay rather than live deployment in some cases; fast multiphysics bias estimation remains rare
Engineering dynamics/biomedical	Mechanical parameters, model weights and latent states; imprecise probability, hybrid weighting and inverse models	Laboratory structures/tanks, real ventilation measurements and wave-tank observations [68, 69, 71, 72]	Small-scale systems; applicability to complex assets requires demonstration
Robotics/ mechatronics	Time-varying dynamic parameters and hidden states; adaptive observer/inverse dynamics	Physical mobile-robot experiments; resilient estimation under missing data [82]	Limited probabilistic uncertainty and reproducibility

8.5 Cross-Domain Patterns and Maturity

Cross-domain maturity is governed primarily by physical-loop completeness, independent validation, uncertainty treatment, and operational reporting—not by algorithmic novelty. Methodologically advanced studies based on simulated observations, fixed-model accuracy compensation, or one-time physical identification can be scientifically strong without satisfying the recurrent physical-observation loop [131–136]. The evidence supports four non-equivalent levels: virtual proof of method, physical laboratory updating, pilot or replayed operational evidence, and live field deployment. Each level supports different claims and should be reported separately.

9 Validation, Credibility, and Evidence Quality

Credibility is the degree to which the evidence supporting a Digital Twin is adequate for its intended use. It is not a permanent property of a model and cannot be inferred from a low calibration error alone. NIST identifies verification, validation and uncertainty quantification as central to Digital Twin credibility, while the National Academies recommends embedding VVUQ throughout design, deployment and monitoring [18, 99]. ASME V&V 40 provides a transferable risk-based principle: the rigor of credibility evidence should increase with the consequence of the decision and the influence of the computational model on that decision [125]. ISO 23247 provides a manufacturing Digital Twin framework but does not by itself guarantee the credibility of a particular predictive model [137].

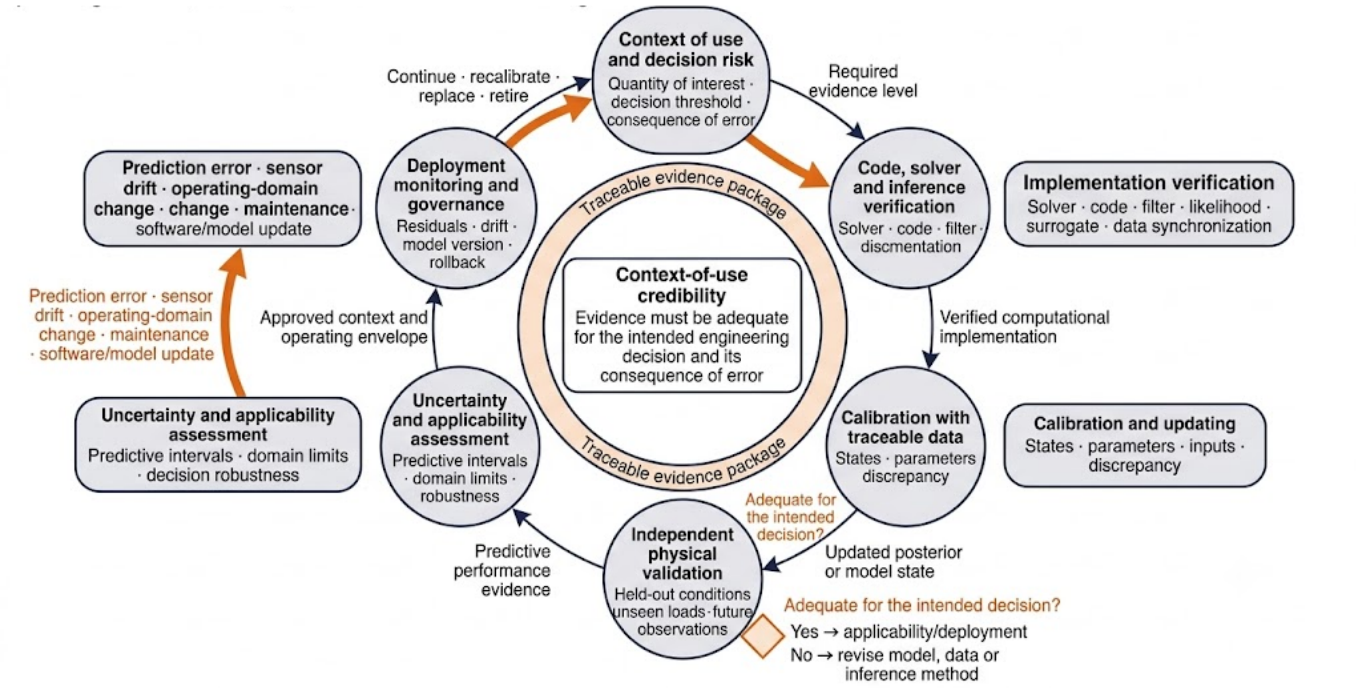

Figure 5: Risk- and context-of-use-driven credibility lifecycle for observation-updated engineering Digital Twins. Verification, calibration, independent physical validation, uncertainty assessment and deployment monitoring form a recurrent evidence process. Prediction errors, sensor drift, operating-domain changes, maintenance actions and model or software updates trigger renewed assessment; therefore, validation is not a terminal event.

9.1 Verification, Calibration, and Validation Are Distinct

Code verification asks whether the implementation correctly solves the chosen mathematical problem. Calculation verification estimates numerical error for the specific simulation. Calibration estimates uncertain parameters or model components from data. Validation evaluates predictive agreement for quantities and conditions relevant to the context of use. These tasks are complementary but not interchangeable. An optimization algorithm can converge to a parameter set for an incorrectly implemented model, and a calibrated model can fit observations while producing biased predictions under changed conditions.

For observation-updated twins, the distinction extends to the inference implementation. Filter equations, Jacobians, likelihoods, surrogate transformations and data synchronization must be verified. Unit tests should include known-state or known-parameter synthetic problems, but synthetic tests cannot replace physical validation. The screened corpus contains many papers with strong numerical verification but no physical observation loop; they are retained as Corpus B precisely because methodological correctness and Digital Twin evidence are different claims.

9.2 Calibration–Validation Separation

The minimum defensible design separates information used to estimate the model from information used to judge its prediction. Separation can be temporal, operational or experimental. Temporal separation uses later operating periods; operational separation uses changed loads, speeds, temperatures or degradation states; experimental separation uses a different specimen or test campaign. The bridge demonstrator, latent particle-filter and industrial-process studies use distinct scenarios or test segments more clearly than studies that report only residual reduction on calibration data [72, 73, 86]. The thermoacoustic study additionally forecasts beyond the assimilation window and tests a unified bias estimator at an unseen equivalence ratio [89]. The shield-tunnel study combines field-informed observations with controlled synthetic degradation cases, but the partition between calibration and validation should be reported more explicitly [88].

Continuous updating complicates the separation because all new observations may enter the posterior. A practical approach is rolling validation: reserve selected sensors or time windows, update with the remaining observations and evaluate one-step and multi-step predictions on the reserved information. Alternatively, maintain a shadow validation channel that never influences the current update. After deployment, prequential scoring—predict first and update after the observation arrives—provides an honest measure of forecasting performance.

9.3 Uncertainty Decomposition and Model-Form Discrepancy

The predictive error of an updated twin can be decomposed conceptually into numerical, input, parameter, measurement and model-form contributions. For a quantity of interest q, a useful bookkeeping relation is

$$e_q = e_{\text{num}} + e_{\text{input}} + e_{\text{param}} + \delta_M + \epsilon \tag{6}$$

where the terms denote numerical error, input uncertainty, parameter uncertainty, model-form discrepancy and measurement noise. The terms are not generally independent, so Equation (6) is an uncertainty-accounting structure rather than a variance-summation formula. Its purpose is to prevent all residual error from being attributed to adjustable parameters.

Model discrepancy is the least consistently treated component in the Corpus A. Bayesian bias identification is methodologically well developed but still largely validated on virtual bridge observations [74]. The aircraft study uses residual feedback with physical flight measurements [81], and Kalibre corrects sensor-observed CFD residuals through a surrogate-calibration loop [85]. The thermoacoustic twin provides the clearest physical online example: model bias and measurement shift are estimated by a reservoir computer and incorporated into a bias-regularized EnKF rather than being absorbed into physical parameters [89]. The real-bridge Bayesian identification study by Koune et al. shows that treating spatially and temporally correlated residuals as independent can produce overconfident predictive intervals, even when parameter posteriors appear precise [136]. The risk is that an unconstrained discrepancy term can conceal damage, sensor bias or changing inputs. Discrepancy models should therefore be constrained, monitored, validated outside the assimilation window and decomposed into interpretable sources where possible.

9.4 Validation Metrics and Acceptance Criteria

Validation metrics should be selected before examining the final results and should correspond to the engineering decision. State-estimation applications may use normalized RMSE, innovation statistics and coverage probability. Structural updating may use modal-frequency error, mode-shape correlation, strain or displacement error and damage-detection performance. Calibration studies should report posterior predictive error and interval coverage, not only parameter estimates. Control and compensation applications should report closed-loop improvement, stability and latency. Remaining-life models require calibration of probability statements and decision-relevant false-positive/false-negative consequences.

A dimensionless validation metric facilitates comparisons across operating points. For observations yi, predictions yhat_i and a domain-specific normalization scale s, the normalized prediction error is

$$E_N = \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{s} \right)^2 \right]^{1/2} \tag{7}$$

The scale may be an observation range, reference magnitude, measurement standard deviation or acceptance tolerance, but it must be stated. Probabilistic twins should additionally report interval coverage and sharpness. A wide interval may achieve nominal coverage while providing little decision value; coverage and width must be interpreted together.

Table 11. Validation and credibility levels proposed for engineering Digital Twin studies.

Level	Evidence description	Claims supported	Claims not yet supported
V0: Algorithm check	Analytical or synthetic known-truth problem	Implementation behaviour and numerical convergence	Physical realism or operational performance
V1: Virtual-system benchmark	High-fidelity simulator provides observations/ground truth	Method comparison under controlled model assumptions	Robustness to sensor bias and real model discrepancy
V2: Physical laboratory validation	Instrumented specimen, rig or process with repeatable conditions	Physical feasibility and local predictive accuracy	Field robustness, long-term drift and fleet variability

Level	Evidence description	Claims supported	Claims not yet supported
V3: Pilot / replayed operational evidence	Pilot asset or archived experimental/operational dataset	Performance under realistic data quality and varied conditions	Live latency, maintenance integration and autonomous decisions
V4: Live field deployment	Connected operational asset with monitored updates and decisions	Operational effectiveness within a stated domain	Generalization beyond the validated context of use
V5: Decision-grade credibility	Risk-based VVUQ, audit trail, monitoring, independent evidence and governance	Use in specified high-consequence decisions	Use outside the documented applicability domain

9.5 Computational Credibility and Real-Time Claims

A real-time claim is meaningful only relative to the observation and decision intervals. Define the latency ratio as

$$\rho_{\mathrm{lat}} = \frac{t_{\mathrm{update}}}{\Delta t_{\mathrm{decision}}} \tag{8}$$

Values below unity indicate that an update completes before the next decision deadline, but communication, preprocessing, storage and actuator latency must be included. Reporting only neural-network inference time is insufficient when feature extraction, high-fidelity model calls or recalibration dominate. The reduced-model Kalman observer and its subsequent ensemble extension provide exemplary physical evidence because computation is evaluated against the machine-tool measurement and compensation cycle [80, 111]. The thermoacoustic Digital Twin assimilates raw observations at approximately 1707 Hz, directly linking inference cadence to the physical oscillation timescale [89]. The event-triggered compensation system additionally shows that latency must include the diagnostic trigger, new data acquisition and re-identification—not merely the final prediction call [123]. The continuous farm calibration is adequate for its environmental cadence despite minute-scale updates [67], while Kalibre is operationally useful with an hours-scale calibration because facility configurations change much more slowly [85]. The shield-tunnel study is designed for efficient parallel recursive updating but illustrates a common reporting gap: an operationally framed method without a complete end-to-end latency value [88].

Computational credibility also includes numerical stability, failure handling and resource dependence. Studies should report hardware, software, precision, ensemble/sample size, convergence criteria and the cost of both offline preparation and online inference. For stochastic methods, effective sample size, filter degeneracy or posterior convergence should be reported. For surrogates, the cost of generating the training basis should not be omitted.

9.6 Evidence-Quality Synthesis

The evidence appraisal uses eight criteria scored from 0 to 2: physical evidence, clarity of the updating target, algorithm transparency, uncertainty treatment, validation, baseline comparison, computational reporting and reproducibility. Fourteen of the 23 Corpus A studies are Strong and nine Moderate. Physical evidence and target clarity are generally high because they are inclusion requirements. The largest deficits are explicit uncertainty treatment, independent validation, open evidence packages and complete latency reporting. The thermoacoustic study reaches 16/16 through unusually complete evidence [89], while the shield-tunnel study remains Moderate because detailed runtime, reproducibility and field-versus-synthetic validation information are limited [88]. The evidence also shows why the total score must be interpreted alongside corpus assignment: a physically validated sensor-placement or one-time Bayesian identification study can score strongly while remaining outside Corpus A because it lacks recurrent synchronization [100, 135, 136].

Strong evidence is not restricted to open-source studies, since industrial data may legitimately remain proprietary. Nevertheless, reproducibility can be improved through model equations, parameter ranges, preprocessing steps, synthetic surrogates of confidential data, pseudocode and precise hardware/software reporting. Proprietary evidence should not become unverifiable evidence. Independent replication is especially important when an updated twin directly controls or certifies a physical asset.

9.7 Proposed Reporting Checklist

Table 12 converts the synthesis into a 24-item reporting checklist. The checklist is intended for authors, reviewers and Digital Twin developers. It does not prescribe a single model architecture; it ensures that the physical loop, inference assumptions and credibility evidence are visible.

Table 12. Engineering Digital Twin Updating Reporting Checklist (EDTURC-24), proposed from the review synthesis.

Group	ID	Reporting item	Minimum information
System and use	R1	Physical counterpart	Identify the asset/process, scale and lifecycle phase.
System and use	R2	Context of use	State the decision supported and consequence of error.
System and use	R3	Quantity of interest	Define prediction/diagnostic outputs and acceptance tolerances.
System and use	R4	Temporal regime	State one-time, periodic, event-triggered, sequential or continuous updating.
Data	R5	Observation provenance	Distinguish physical, replayed, synthetic and high-fidelity-simulation observations.
Data	R6	Sensors and synchronization	Report sensor type, location, rate, time alignment and missing-data handling.
Data	R7	Measurement uncertainty	Quantify calibration, noise, bias, drift and resolution.
Data	R8	Data partition	Identify calibration, validation and test periods/specimens/conditions.
Model	R9	Model representation	Describe physics, reduced, surrogate, data-driven or hybrid components.
Model	R10	Fidelity and cost	Report dimensions, mesh/order, solver and representative runtime.
Model	R11	Updating target	List states, parameters, inputs, geometry, discrepancy or structure updated.
Model	R12	Identifiability	Report sensitivity, observability, excitation or parameter-correlation analysis.
Inference	R13	Algorithm and implementation	Provide equations, settings, initialization and convergence/stopping rules.
Inference	R14	Prior/regularization choices	Justify priors, bounds, penalties and hyperparameters.
Inference	R15	Uncertainty propagation	Describe process, parameter, input and predictive uncertainty.
Inference	R16	Model discrepancy	State whether discrepancy is ignored, modelled or learned and how confounding is controlled.
Validation	R17	Verification	Report code/solver tests, numerical convergence and known-truth checks.
Validation	R18	Independent physical validation	Use observations or conditions not used in fitting.
Validation	R19	Metrics and thresholds	Predefine engineering and probabilistic performance measures.
Validation	R20	Baseline comparison	Compare unupdated model and credible alternative methods.
Deployment	R21	Latency and hardware	Report total pipeline latency relative to the decision interval.
Deployment	R22	Applicability domain	Define operating range and out-of-domain/drift detection.
Deployment	R23	Update governance	Record model versions, approval, rollback and recalibration triggers.
Reproducibility	R24	Evidence availability	Provide code/data where possible or sufficient substitutes for replication.

The evidence indicates that credible model updating is a systems-engineering workflow rather than an isolated inverse algorithm. The method must be compatible with the target, uncertainty and latency; the observation design must make the target identifiable; the updated prediction must be validated independently; and the system must continue to monitor applicability after deployment. These requirements form the basis for the method-selection framework, cross-domain comparison and reporting checklist proposed in Sections 7–9.

10 Open Challenges and Research Roadmap

The remaining scientific challenge is sustained, credible co-evolution as assets, sensors, operating domains, and decision contexts change. The roadmap therefore prioritizes continuous validation, bias-aware inference, active sensing, model portfolios, interoperability, reproducibility, and governance rather than isolated gains in synchronization frequency.

The roadmap proposed here is therefore organized around a transition from episodic calibration to governed model co-evolution. Near-term work should make evidence comparable and validation continuous. Medium-term work should integrate nonstationarity, active sensing and interoperable model portfolios. Longer-term work should enable decision-grade autonomy while preserving explicit applicability limits, human oversight and rollback mechanisms.

10.1 From Episodic Updating to Managed Co-Evolution

Most reported updating studies assume that the model class remains adequate and that new observations modify only a fixed set of states or parameters. Physical assets violate this assumption through ageing, repairs, component replacement, software changes, environmental shifts and changing operating policies. A future Digital Twin must distinguish at least three events: ordinary state variation, gradual parameter drift and a structural regime change that invalidates the current model class. Treating all three as parameter noise can produce apparently stable but physically misleading estimates.

Continuous validation provides the missing supervisory layer. Validation metrics should be evaluated at runtime, not only during commissioning, and should trigger diagnosis before automatic recalibration. Mertens and Denil demonstrate how runtime validation can detect divergence and initiate parameter correction in a cyber-physical system [138]. The event-triggered machine-tool study supplies complementary physical evidence: control-limit violations initiate new data collection and model re-identification rather than unconditional continual learning [123]. The thermoacoustic twin shows that structured model bias and measurement shift can be inferred continuously without forcing residuals into physical parameters [89], while the shield-tunnel study illustrates sequential updating under long-term degradation [88]. The next research step is a hierarchical response: update states when innovations are statistically consistent; update parameters when persistent but explainable drift is detected; activate discrepancy correction or model replacement when the residual structure changes; and suspend decision authority when none of the available models is credible [139].

A practical co-evolution trigger may combine a validation statistic, an out-of-distribution score and a decision-risk multiplier:

Equation (9) is a conceptual trigger structure rather than a validated operational rule. Its terms, normalization, weights, and action thresholds must be calibrated prospectively and evaluated for false alarms, missed changes, and decision consequences.

$$G_k = w_v V_k + w_d D_k + w_r R_k \tag{9}$$

where Vk measures predictive inconsistency, Dk measures departure from the calibrated operating domain and Rk represents the consequence of using the twin for the current decision. An update, escalation or shutdown threshold should be defined from the context of use rather than selected retrospectively.

10.2 Nonstationarity, Abrupt Changes, and Model-Discrepancy Confounding

Sequential filtering handles changing states efficiently, but it does not automatically separate physical parameter evolution from systematic model bias. Offline Bayesian calibration, conversely, can represent discrepancy but commonly assumes stationarity. Real-bridge evidence shows that unmodelled spatial and temporal residual correlation can itself be mistaken for parameter certainty [136]. Online Bayesian Recursive Projected Calibration has recently been proposed to track gradual evolution while maintaining separation between calibration parameters and discrepancy, with restart mechanisms for abrupt regime shifts [140]. Although current demonstrations remain largely synthetic or simulation based, this direction addresses one of the most important gaps identified by the present review.

Future studies should report whether a detected change is interpreted as state evolution, parameter drift, sensor bias, input error or model inadequacy. This distinction is necessary for engineering action: a sensor recalibration, a maintenance intervention and a model-class replacement are not interchangeable responses. Research is needed on identifiable joint estimation under sparse excitation, especially when parameters and discrepancy generate similar output effects. Orthogonal discrepancy models, physics-informed priors, designed excitation and multi-sensor data can reduce confounding, but their effectiveness must be tested on physical assets with known interventions.

10.3 Active Observation, Sensor Steering, and Information-Seeking Twins

The studies in the main corpus treat observations primarily as exogenous inputs. This passive paradigm is inefficient when sensing is costly, intermittent or weakly informative. An active twin should select measurements, excitation or operating actions that reduce uncertainty while respecting safety and production constraints. The recently published active-inference framework formulates a twin as a partially observable Markov decision process that balances goal-directed utility with epistemic information gain [141]. Its railway-bridge demonstration remains virtual, but the conceptual shift is important: actions can be chosen partly because they improve the model. The physical machine-tool sensor-placement study demonstrates the enabling layer for this transition by selecting a sparse sensor set through model-ensemble observability rather than signal amplitude alone [100]. Future work must close the loop by adapting sensing decisions during operation and demonstrating improved downstream updating or control. Bayesian optimal experimental design supplies a rigorous information-theoretic basis for choosing such sensors and excitations in PDE-constrained inverse problems [142].

A general active-observation objective can be written as

Equation (10) is likewise illustrative. It formalizes the trade-off between engineering utility, information gain, and sensing cost, but requires domain-specific utility definitions, safety constraints, and physical validation before deployment.

$$a^* = \arg\max_a \{\mathbb{E}[U_{\text{eng}}] + \beta I(z; y \mid a) - \lambda C(a)\} \tag{10}$$

where Ueng is engineering utility, I(z;y|a) is the expected information gained about hidden quantities z from an observation or action a, and C(a) represents sensing, energy, disruption and safety costs. Research priorities include observability-aware sensor scheduling, excitation design for nonlinear systems, adaptive sampling rates, and coordination between estimation and control. Physical validation should demonstrate that the information-seeking action improves subsequent prediction or decision quality, not merely that it increases a formal information metric.

10.4 Multi-Fidelity, Hybrid, and Model-Portfolio Updating

No single model representation is optimal across all lifecycle decisions. High-fidelity models provide physical detail but are rarely fast enough for continual inference; reduced and learned models are efficient but may fail outside their training domain. Future twins should manage portfolios of models with explicit fidelity, validity and cost metadata. A supervisory layer should select, combine or replace models according to the current decision, observation quality and available computation. The machine-tool ensemble and the time-evolving hybrid model provide early examples of this principle [69, 80].

Important open problems include uncertainty transfer between fidelities, conservation and stability after learned correction, online detection of surrogate extrapolation, and propagation of calibration evidence when a model is replaced. Neural operators and physics-informed learning should be evaluated not only on interpolation accuracy but also on update stability, posterior calibration and failure recovery. High-fidelity models should remain available as periodic anchors for reduced twins, while discrepancies identified during operation should inform both surrogate retraining and revision of the governing model assumptions.

10.5 Continuous Credibility, Out-of-Domain Detection, and Governance

A validated twin can become invalid without an obvious software failure. Sensor ageing, maintenance, changes in material properties and new operating regimes can silently move the system outside the validated domain. Continuous credibility therefore requires monitored applicability, versioned evidence and an explicit response when confidence deteriorates. The update algorithm itself must be treated as a safety-relevant component: poorly chosen automatic updates can make a previously adequate model worse.

A governance framework should record observation provenance, model and parameter versions, validation status, update triggers, approvals, rejected updates and rollback points. High-consequence decisions should use staged authority: advisory predictions first, supervised actions second and autonomous intervention only after evidence has accumulated under representative conditions. NIST is advancing manufacturing Digital Twin test methods, standards and a planned VVUQ extension of ISO 23247, reflecting the need to combine implementation architecture with measurement science and assurance [137, 143].

10.6 Interoperability, Federated Twins, and Semantic Traceability

Most current implementations are vertically integrated demonstrators. Scaling to fleets, supply chains and systems of systems requires exchange not only of sensor values but also of units, uncertainty, model assumptions, calibration history and validity ranges. ISO 23247 provides a general manufacturing framework, while current standards work emphasizes reusable architectures and data/model exchange [137, 143]. However, syntactic interoperability does not guarantee inferential interoperability: two twins may exchange a parameter while assigning it different physical meanings, priors or timestamps.

Research should develop machine-readable contracts for quantities of interest, model inputs, uncertainty semantics, causal dependencies and evidence lineage. Federated updating must also address conflicting observations, different model fidelities, privacy constraints and local ownership. A federated twin should not merge posteriors or model corrections unless their conditioning data, assumptions and correlations are understood. Standardized provenance and calibration manifests could make model updates auditable across organizations without requiring disclosure of proprietary raw data.

10.7 Data Sufficiency, Benchmarking, and Reproducibility

The literature frequently specifies the algorithm in more detail than the data required to make it work. Data sufficiency depends on the context of use, target dimension, excitation, sensor placement and tolerated uncertainty. Recent NIST work explicitly frames Digital Twins as descriptive, diagnostic, predictive, prescriptive or autonomous systems with different data requirements [144]. The machine-tool sensor-placement study provides direct physical evidence that substantially fewer sensors can be sufficient when selection is tied to model observability and validated under changing boundary conditions [100]. Similar distinctions and ablation studies are needed across engineering domains.

The field requires benchmark packages that combine a physical asset or traceable experiment, synchronized raw observations, sensor-calibration information, model versions, predefined calibration and validation partitions, decision-relevant quantities of interest, and latency constraints. Synthetic benchmarks remain valuable for known-truth verification, but they should be paired with physical tests that expose model-form error. Reporting should distinguish open data from reproducible evidence: proprietary studies can still provide equations, pseudocode, parameter bounds, preprocessing steps, hardware and blinded validation metrics.

10.8 Real-Time Scalability and Sustainable Computation

Real-time feasibility is increasingly achieved through reduction, surrogates and accelerated inference, but offline costs are often omitted. The environmental and operational cost of a twin includes sensor acquisition, communication, high-fidelity training simulations, repeated retraining and uncertainty propagation. A method that reduces online inference to milliseconds may still be unsuitable if it requires frequent large-scale retraining or cannot run near the asset.

Future work should report lifecycle computational cost, energy use and hardware dependence alongside latency. Edge–cloud partitioning should be designed from the decision cadence: fast state estimation and safety logic near the asset; heavier recalibration, model selection and uncertainty audits at slower supervisory layers. Event-triggered updating can reduce unnecessary computation, but triggers must be validated to avoid missing gradual degradation. Sustainable twins will require reusable reduced models, incremental learning, uncertainty-aware caching and model retirement policies.

Table 13. Priority research challenges and evidence required for progress.

Challenge	Current limitation	Research objective	Minimum convincing evidence
Nonstationary co-evolution	Fixed model class and stationary-noise assumptions	Separate state variation, parameter drift, sensor bias and regime change	Physical asset with documented interventions and prospective change detection
Bias-aware online calibration	Filtering and discrepancy modelling are commonly separated	Jointly track parameters and structured discrepancy without confounding	Known-bias benchmark followed by independent physical validation
Active sensing	Observations are treated as externally fixed	Select sensors/actions by engineering utility and information gain	Physical demonstration showing improved downstream prediction or decision

Challenge	Current limitation	Research objective	Minimum convincing evidence
Model portfolios	One model is used across incompatible decisions	Select and combine models using fidelity, cost and validity metadata	Safe switching/combination with uncertainty continuity and rollback
Continuous credibility	Validation is mainly pre-deployment	Monitor applicability and trigger update, escalation or suspension	Runtime validation with controlled drift and false-alarm analysis
Interoperability	Data exchange lacks uncertainty and evidence semantics	Machine-readable provenance, units, priors, validity and calibration manifests	Cross-platform twin exchange with preserved predictive meaning
Data sufficiency	Sensor quantity is chosen heuristically	Link sensing requirements to identifiability and context of use	Sensor-ablation and observability study on a physical system
Scalable UQ	High-dimensional uncertainty remains expensive	Real-time non-Gaussian inference with calibrated coverage	Latency, coverage, sharpness and failure-mode comparison
Reproducibility	Proprietary or incomplete evidence limits replication	Share executable methods or adequate substitutes	Independent reproduction or blinded benchmark evaluation
Autonomous decisions	Updating and control are validated separately	Assure the coupled inference–decision loop	Risk-based closed-loop testing with safe fallback and human override

10.9 Staged Research Roadmap

Figure 6 and Table 14 organize these priorities into three horizons. The horizons are capability-based rather than predictions of calendar adoption. Progress to the next horizon should depend on evidence maturity: reliable autonomy cannot be obtained by adding a decision agent to a twin that lacks continuous validation, identifiable updating and traceable data.

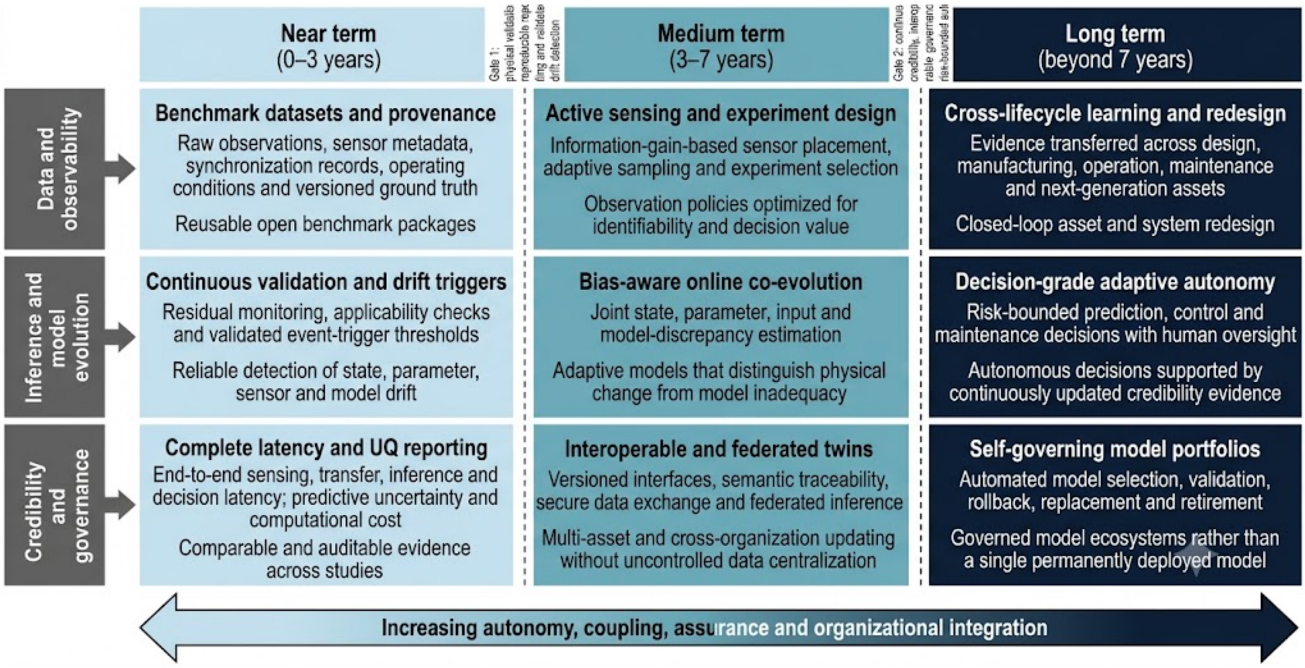

Figure 6: Capability-based research roadmap for observation-updated engineering Digital Twins. Near-term priorities establish reproducible data, continuous validation and complete uncertainty and latency reporting. Medium-term capabilities enable active sensing, bias-aware online model co-evolution and interoperable federated architectures. Long-term progress targets cross-lifecycle learning, decision-grade adaptive autonomy and governed model portfolios. Advancement between horizons depends on demonstrated evidence maturity rather than elapsed calendar time.

Table 14. Capability-based roadmap and measurable milestones.

Horizon	Primary objective	Representative milestones	Exit evidence
Near term (0–3 years)	Make updating studies comparable and continuously assessable	Common terminology; physical/synthetic provenance labels; EDTURC-24 adoption; runtime validation; complete latency and UQ reporting	Multi-domain benchmark studies with independent physical validation and prospective metrics

Horizon	Primary objective	Representative milestones	Exit evidence
Medium term (3–7 years)	Enable bias-aware, active and interoperable co-evolution	Regime-shift handling; active sensor scheduling; model portfolios; semantic calibration manifests; federated uncertainty exchange	Cross-platform pilot deployments with safe model switching and quantified decision benefit
Long term (>7 years)	Achieve decision-grade adaptive autonomy	Self-governing update policies; coordinated sensing/control; fleet and lifecycle learning; automated evidence accumulation and rollback	Long-duration field operation under changing conditions with audited safety, resilience and human oversight

10.10 Limitations of the Current Evidence Base and Review

Current evidence is concentrated in selected structural, manufacturing, energy, combustion, and controlled-system applications, and several advanced methods rely on simulation or replayed datasets. Search-interface instability, heterogeneous outputs, and incomplete reporting preclude prevalence estimates and quantitative meta-analysis. The conclusions are therefore corpus-descriptive. Detailed search, screening, appraisal, and publication-status records are available from the corresponding author upon reasonable request.

11 Conclusions

This review examined how physical observations are converted into updated engineering models that can function as Digital Twins. The synthesis shows that model updating is not a single algorithmic task. It is a coupled process comprising observation design, state and parameter inference, model-form assessment, uncertainty propagation, computational acceleration, independent validation and lifecycle governance. A model becomes a credible observation-updated twin only when this process is repeatable and linked to a defined engineering decision.

Four conclusions follow. First, updating targets and temporal regimes must be stated explicitly because states, parameters, inputs, discrepancy, and model structure require different inference assumptions. Second, no method family is universally superior: selection depends on identifiability, uncertainty, model cost, cadence, interpretability, and decision risk. Third, physical-loop completeness and independent predictive validation matter more than algorithm novelty. Fourth, computational performance must be assessed against the complete sensing-to-decision interval and monitored after deployment.

The review contributes a six-axis taxonomy, an evidence-based selection framework, a validation-maturity hierarchy, and EDTURC-24. These elements support a disciplined workflow: Observe → diagnose → infer → update → quantify uncertainty → validate → decide → monitor. Their scope is the documented 42-record corpus and the 23-study physical-evidence subset, not a prevalence estimate for the entire field.

Future progress requires actively observed, continuously validated, and governed model ecosystems that can distinguish ordinary state correction from parameter drift, structural model failure, and loss of decision authority. Active inference, bias-aware calibration, model portfolios, federated evidence exchange, and continuous credibility monitoring should therefore be judged by independent physical evidence, calibrated uncertainty, transparent computational cost, and safe performance under change.

The central engineering criterion is therefore not whether a Digital Twin can be updated, but whether every update produces a model that is more decision-relevant, more credible and no less physically defensible than the model it replaces. Progress should be measured by independent predictive evidence, calibrated uncertainty, transparent computational cost and safe performance under change—not by synchronization frequency or model complexity alone.

Author contributions

S.K.G.: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing—original draft, and Writing—review and editing. N.D.L.: Methodology, Validation, Supervision, and Writing—review and editing.

Data availability

Data are available from the corresponding author upon reasonable request.

Declaration of generative AI assistance

Generative AI tools were used to support language editing, document structuring, visual preparation, and bibliographic workflow management. The authors remain responsible for source verification, scientific interpretation, corpus decisions, appraisal scores, and the final manuscript.

Ethics approval and consent

Not applicable. This review synthesized published literature and did not involve human participants, animals, clinical interventions, or identifiable personal data.

Funding

This research received no external funding.

Declaration of competing interest

S.K.G. and N.D.L. are editors of IJDTDE.

References

[1] Rasheed A, San O, Kvamsdal T. Digital Twin: Values, Challenges and Enablers From a Modeling Perspective. IEEE Access. 2020. doi:10.1109/ACCESS.2020.2970143.

[2] Wagg DJ, Worden K, Barthorpe RJ, Gardner P. Digital Twins: State-of-the-Art and Future Directions for Modeling and Simulation in Engineering Dynamics Applications. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B. 2020. doi:10.1115/1.4046739.

[3] Thelen A, et al. A Comprehensive Review of Digital Twin-Part 1: Modeling and Twinning Enabling Technologies. Structural and Multidisciplinary Optimization. 2022. doi:10.1007/s00158-022-03425-4.

[4] Thelen A, et al. A Comprehensive Review of Digital Twin-Part 2: Roles of Uncertainty Quantification and Optimization, a Battery Digital Twin, and Perspectives. Structural and Multidisciplinary Optimization. 2023. doi:10.1007/s00158-022-03410-x.

[5] Jones D, Snider C, Nassehi A, Yon J, Hicks B. Characterising the Digital Twin: A Systematic Literature Review. CIRP Journal of Manufacturing Science and Technology. 2020;29:36–52. doi:10.1016/j.cirpj.2020.02.002.

[6] Semeraro C, Lezoche M, Panetto H, Dassisti M. Digital Twin Paradigm: A Systematic Literature Review. Computers in Industry. 2021;130:103469. doi:10.1016/j.compind.2021.103469.

[7] Kritzinger W, Karner M, Traar G, Henjes J, Sihn W. Digital Twin in Manufacturing: A Categorical Literature Review and Classification. IFAC-PapersOnLine. 2018;51(11):1016–1022. doi:10.1016/j.ifacol.2018.08.474.

[8] Negri E, Fumagalli L, Macchi M. A Review of the Roles of Digital Twin in CPS-Based Production Systems. Procedia Manufacturing. 2017;11:939–948. doi:10.1016/j.promfg.2017.07.198.

[9] Tao F, Zhang H, Liu A, Nee AYC. Digital Twin in Industry: State-of-the-Art. IEEE Transactions on Industrial Informatics. 2019;15(4):2405–2415. doi:10.1109/TII.2018.2873186.

[10] Fuller A, Fan Z, Day C, Barlow C. Digital Twin: Enabling Technologies, Challenges and Open Research. IEEE Access. 2020;8:108952–108971. doi:10.1109/ACCESS.2020.2998358.

[11] Cimino C, Negri E, Fumagalli L. Review of Digital Twin Applications in Manufacturing. Computers in Industry. 2019;113:103130. doi:10.1016/j.compind.2019.103130.

[12] Madni AM, Madni CC, Lucero SD. Leveraging Digital Twin Technology in Model-Based Systems Engineering. Systems. 2019;7(1):7. doi:10.3390/systems7010007.

[13] Tao F, Qi Q, Wang L, Nee AYC. Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison. Engineering. 2019;5(4):653–661. doi:10.1016/j.eng.2019.01.014.

[14] Glaessgen EH, Stargel DS. The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles. In: 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. Honolulu, HI; 2012. doi:10.2514/6.2012-1818.

[15] Tuegel EJ, Ingraffea AR, Eason TG, Spottswood SM. Reengineering Aircraft Structural Life Prediction Using a Digital Twin. International Journal of Aerospace Engineering. 2011;2011:154798. doi:10.1155/2011/154798.

[16] Wright L, Davidson S. How to Tell the Difference Between a Model and a Digital Twin. Advanced Modeling and Simulation in Engineering Sciences. 2020;7:13. doi:10.1186/s40323-020-00147-4.

[17] Schleich B, Anwer N, Mathieu L, Wartzack S. Shaping the Digital Twin for Design and Production Engineering. CIRP Annals. 2017;66(1):141–144. doi:10.1016/j.cirp.2017.04.040.

[18] National Academies of Sciences, Engineering, and Medicine. Foundational Research Gaps and Future Directions for Digital Twins. Washington, DC: The National Academies Press; 2024. doi:10.17226/26894.

[19] Ljung L. Perspectives on System Identification. Annual Reviews in Control. 2010;34(1):1–12. doi:10.1016/j.arcontrol.2009.12.001.

[20] Farrar CR, Worden K. An Introduction to Structural Health Monitoring. Philosophical Transactions of the Royal Society A. 2007;365(1851):303–315. doi:10.1098/rsta.2006.1928.

[21] Worden K, Manson G. The Application of Machine Learning to Structural Health Monitoring. Philosophical Transactions of the Royal Society A. 2007;365(1851):515–537. doi:10.1098/rsta.2006.1938.

[22] Mottershead JE, Friswell MI. Model Updating in Structural Dynamics: A Survey. Journal of Sound and Vibration. 1993;167(2):347–375. doi:10.1006/jsvi.1993.1340.

[23] Beck JL, Katafygiotis LS. Updating Models and Their Uncertainties. I: Bayesian Statistical Framework. Journal of Engineering Mechanics. 1998;124(4):455–461. doi:10.1061/(ASCE)0733-9399(1998)124:4(455).

[24] Katafygiotis LS, Beck JL. Updating Models and Their Uncertainties. II: Model Identifiability. Journal of Engineering Mechanics. 1998;124(4):463–467. doi:10.1061/(ASCE)0733-9399(1998)124:4(463).

[25] Vanik MW, Beck JL, Au SK. Bayesian Probabilistic Approach to Structural Health Monitoring. Journal of Engineering Mechanics. 2000;126(7):738–745. doi:10.1061/(ASCE)0733-9399(2000)126:7(738).

[26] Kerschen G, Worden K, Vakakis AF, Golinval JC. Past, Present and Future of Nonlinear System Identification in Structural Dynamics. Mechanical Systems and Signal Processing. 2006;20(3):505–592. doi:10.1016/j.ymssp.2005.04.008.

[27] Kennedy MC, O'Hagan A. Bayesian Calibration of Computer Models. Journal of the Royal Statistical Society: Series B. 2001;63(3):425–464. doi:10.1111/1467-9868.00294.

[28] Bayarri MJ, Paulo R, Berger JO, Sacks J, Cafeo JA, Cavendish J, Lin CH, Tu J. A Framework for Validation of Computer Models. Technometrics. 2007;49(2):138–154. doi:10.1198/004017007000000092.

[29] Higdon D, Gattiker J, Williams B, Rightley M. Computer Model Calibration Using High-Dimensional Output. Journal of the American Statistical Association. 2008;103(482):570–583. doi:10.1198/016214507000000888.

[30] Brynjarsdóttir J, O'Hagan A. Learning about Physical Parameters: The Importance of Model Discrepancy. Inverse Problems. 2014;30(11):114007. doi:10.1088/0266-5611/30/11/114007.

[31] Plumlee M. Bayesian Calibration of Inexact Computer Models. Journal of the American Statistical Association. 2017;112(519):1274–1285. doi:10.1080/01621459.2016.1211016.

[32] Arendt PD, Apley DW, Chen W, Lamb D, Gorsich D. Improving Identifiability in Model Calibration Using Multiple Responses. Journal of Mechanical Design. 2012;134(10):100909. doi:10.1115/1.4007573.

[33] Gu M, Wang L. Scaled Gaussian Stochastic Process for Computer Model Calibration and Prediction. SIAM/ASA Journal on Uncertainty Quantification. 2018;6(4):1555–1583. doi:10.1137/17M1159890.

[34] Morrison RE, Oliver TA, Moser RD. Representing Model Inadequacy: A Stochastic Operator Approach. SIAM/ASA Journal on Uncertainty Quantification. 2018;6(2):457–496. doi:10.1137/16M1106419.

[35] Kalman RE. A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering. 1960;82(1):35–45. doi:10.1115/1.3662552.

[36] Julier SJ, Uhlmann JK. Unscented Filtering and Nonlinear Estimation. Proceedings of the IEEE. 2004;92(3):401–422. doi:10.1109/JPROC.2003.823141.

[37] Evensen G. Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. Journal of Geophysical Research: Oceans. 1994;99(C5):10143–10162. doi:10.1029/94JC00572.

[38] Evensen G. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. Ocean Dynamics. 2003;53:343–367. doi:10.1007/s10236-003-0036-9.

[39] Arulampalam MS, Maskell S, Gordon N, Clapp T. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. IEEE Transactions on Signal Processing. 2002;50(2):174–188. doi:10.1109/78.

978374.

[40] Doucet A, Godsill S, Andrieu C. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. Statistics and Computing. 2000;10(3):197–208. doi:10.1023/A:1008935410038.

[41] Rao CV, Rawlings JB, Mayne DQ. Constrained State Estimation for Nonlinear Discrete-Time Systems: Stability and Moving Horizon Approximations. IEEE Transactions on Automatic Control. 2003;48(2):246–258. doi:10.1109/TAC.2002.808470.

[42] Carrassi A, Bocquet M, Bertino L, Evensen G. Data Assimilation in the Geosciences: An Overview of Methods, Issues, and Perspectives. WIREs Climate Change. 2018;9(5):e535. doi:10.1002/wcc.535.

[43] Bannister RN. A Review of Operational Methods of Variational and Ensemble-Variational Data Assimilation. Quarterly Journal of the Royal Meteorological Society. 2017;143:607–633. doi:10.1002/qj.2982.

[44] Benner P, Gugercin S, Willcox K. A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems. SIAM Review. 2015;57(4):483–531. doi:10.1137/130932715.

[45] Peherstorfer B, Willcox K, Gunzburger M. Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization. SIAM Review. 2018;60(3):550–591. doi:10.1137/16M1082469.

[46] Rowley CW. Model Reduction for Fluids, Using Balanced Proper Orthogonal Decomposition. International Journal of Bifurcation and Chaos. 2005;15(3):997–1013. doi:10.1142/S0218127405012429.

[47] Oberkampf WL, Trucano TG. Verification and Validation in Computational Fluid Dynamics. Progress in Aerospace Sciences. 2002;38(3):209–272. doi:10.1016/S0376-0421(02)00005-2.

[48] Roy CJ, Oberkampf WL. A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing. Computer Methods in Applied Mechanics and Engineering. 2011;200(25–28):2131–2144. doi:10.1016/j.cma.2011.03.016.

[49] Stuart AM. Inverse Problems: A Bayesian Perspective. Acta Numerica. 2010;19:451–559. doi:10.1017/S0962492910000061.

[50] Kaipio J, Somersalo E. Statistical and Computational Inverse Problems. New York: Springer; 2005. doi:10.1007/b138659.

[51] Marzouk YM, Najm HN, Rahn LA. Stochastic Spectral Methods for Efficient Bayesian Solution of Inverse Problems. Journal of Computational Physics. 2007;224(2):560–586. doi:10.1016/j.jcp.2006.10.010.

[52] Smith RC. Uncertainty Quantification: Theory, Implementation, and Applications. Philadelphia: SIAM; 2014. doi:10.1137/1.9781611973228.

[53] Oliver TA, Terejanu G, Simmons CS, Moser RD. Validating Predictions of Unobserved Quantities. Computer Methods in Applied Mechanics and Engineering. 2015;283:1310–1335. doi:10.1016/j.cma.2014.08.023.

[54] Trucano TG, Swiler LP, Igusa T, Oberkampf WL, Pilch M. Calibration, Validation, and Sensitivity Analysis: What's What. Reliability Engineering & System Safety. 2006;91(10–11):1331–1357. doi:10.1016/j.ress.2005.11.031.

[55] Es-haghi MS, Anitescu C, Rabczuk T. Methods for Enabling Real-Time Analysis in Digital Twins: A Literature Review. Computers & Structures. 2024;297:107342. doi:10.1016/j.compstruc.2024.107342.

[56] Zech P, Barat S, Nast B, Oakes B, Michael J, Zschaler S, Barn B, Breu R. Model-Based Digital Twin Engineering: Insights, Challenges, and Future Directions. Software and Systems Modeling. 2026. doi:10.1007/s10270-026-01368-8.

[57] Kopp L, Harfmann P, Mangaluru Ramananda A, Kley M. Digital Twins in Engineering: A Literature Review and Expert Survey for Potential Analysis. Engineering Research. 2026;90:8. doi:10.1007/s10010-026-00940-4.

[58] Adimass SA, Żak A. Review of Digital Twin in Structural Dynamics to Improve Condition Monitoring. Structural Control and Health Monitoring. 2026;2026:2777297. doi:10.1155/stc/2777297.

[59] Ereiz S, Duvnjak I, Jimenez-Alonso JF. Review of Finite Element Model Updating Methods for Structural Applications. Structures. 2022. doi:10.1016/j.istruc.2022.05.041.

[60] Chen Y, et al. Finite Element Model Updating for Material Model Calibration: A Review and Guide to Practice. Archives of Computational Methods in Engineering. 2025. doi:10.1007/s11831-024-10200-9.

[61] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

[62] Li Y, et al. Statistical Parameterized Physics-Based Machine Learning Digital Shadow Models for Laser Powder Bed Fusion Process. Additive Manufacturing. 2024. doi:10.1016/j.addma.2024.104214.

[63] Liu J, et al. Deep Neural Operator Enabled Digital Twin Modeling for Additive Manufacturing. Advances in Computational Science and Engineering. 2024. doi:10.3934/acse.2024010.

[64] Eneyew DD, Capretz MAM, Bitsuamlak GT. Continuous Model Calibration Framework for Smart-Building Digital Twin: A Generative Model-Based Approach. Applied Energy. 2024. doi:10.1016/j.apenergy.2024.124080.

[65] Karkadakattil A. A Physics-Stabilized Self-Updating Digital Twin Framework Using Physics-Informed Neural Networks for Thermal Field Prediction. Applied Computer Systems. 2026. doi:10.2478/acss-2026-0003.

[66] Jiang Z, Karlsson M, Agrell E, Hager C. PIDT: Physics-Informed Digital Twin for Optical Fiber Parameter Estimation. Optical Fiber Communication Conference. 2026. doi:10.1364/OFC.2026.M4K.3.

[67] Ward R, Choudhary R, Gregory A, Jans-Singh M, Girolami M. Continuous Calibration of a Digital Twin: Comparison of Particle Filter and Bayesian Calibration Approaches. Data-Centric Engineering. 2021;2:e15. doi:10.1017/dce.2021.12.

[68] de Angelis M, et al. Robust Online Updating of a Digital Twin with Imprecise Probability. Mechanical Systems and Signal Processing. 2023;185:109877. doi:10.1016/j.ymssp.2022.109877.

[69] Edington T, et al. A Time-Evolving Digital Twin Tool for Engineering Dynamics Applications. Mechanical Systems and Signal Processing. 2023;185:109971. doi:10.1016/j.ymssp.2022.109971.

[70] Kessels BM, Fey RHB, van de Wouw N. Real-Time Parameter Updating for Nonlinear Digital Twins Using Inverse Mapping Models and Transient-Based Features. Nonlinear Dynamics. 2023. doi:10.1007/s11071-023-08354-5.

[71] Kessels BM, et al. Uncertainty Quantification in Real-Time Parameter Updating for Digital Twins Using Bayesian Inverse Mapping Models. Nonlinear Dynamics. 2025. doi:10.1007/s11071-024-10608-9.

[72] Mucke NT, Bohte SM, Oosterlee CW. The Deep Latent Space Particle Filter for Real-Time Data Assimilation with Uncertainty Quantification. Scientific Reports. 2024. doi:10.1038/s41598-024-69901-7.

[73] Titscher T, van Dijk T, Kadoke D, Robens-Radermacher A, Herrmann R, Unger JF. Bayesian Model Calibration and Damage Detection for a Digital Twin of a Bridge Demonstrator. Engineering Reports. 2023;5:e12669. doi:10.1002/eng2.12669.

[74] Arcones A, et al. Model Bias Identification for Bayesian Calibration of Stochastic Digital Twins of Bridges. Applied Stochastic Models in Business and Industry. 2024. doi:10.1002/asmb.2897.

[75] Donato L, Galletti C, Parente A. Self-Updating Digital Twin of a Hydrogen-Powered Furnace Using Data Assimilation. Applied Thermal Engineering. 2024;236:121431. doi:10.1016/j.applthermaleng.2023.121431.

[76] Shi G, et al. Digital Twin-Based Model Updating Method for Mechanical Behaviors Analysis of Cable Truss Structure. Journal of Constructional Steel Research. 2024;220:108917. doi:10.1016/j.jcsr.2024.108917.

[77] Tang A, Shi X, Qin G, Shi H, Mei X. A Digital Twin-Based System for Timely Simulation of Bridge Structural Performance with Finite Element Model Updating. Structure and Infrastructure Engineering. 2025. doi:10.1080/15732479.2025.2587043.

[78] Ali H, Bocchino G, Bono F, Perotti F, Martinelli L. Finite Element Model Updating for Digital Twin Development Using Operational Modal Analysis: The JRC Atmospheric Observatory Tower Case Study. Lecture Notes in Civil Engineering. 2025. doi:10.1007/978-3-031-96110-6_32.

[79] Xia J, et al. Digital Twin-Assisted Gearbox Dynamic Model Updating Toward Fault Diagnosis. Frontiers of Mechanical Engineering. 2023. doi:10.1007/s11465-023-0748-0.

[80] Lang S, et al. Overcoming Uncertainty With an Ensemble of Physical Models and Real-Time Measurements: Thermal Error Compensation Using Kalman Filters in a Digital Twin. In: ICTIMT 2025 Proceedings. 2026. doi:10.1007/978-3-032-01194-7_25.

[81] Zhuo Y, Yan B. A Key Component in Digital Twin of Aircraft Structures: Multi-Dimensional Flight-Parameter Measurements. Chinese Journal of Aeronautics. 2025. doi:10.7527/S1000-6893.2025.32375.

[82] Kosmaga SR, et al. Digital Twin-Enabled Real-Time Parameter Estimation of a Wheeled Mobile Robot. Systems Science & Control Engineering. 2026. doi:10.1080/21642583.2026.2634459.

[83] Zhao K, et al. Digital Twin-Supported Battery State Estimation Based on TCN-LSTM Neural Networks and Transfer Learning. CSEE Journal of Power and Energy Systems. 2025. doi:10.17775/CSEEJPES.2024.00900.

[84] Wu T, et al. An Online Learning Method for Constructing a Self-Update Digital Twin Model for Transformer Temperature Prediction. Applied Thermal Engineering. 2024. doi:10.1016/j.applthermaleng.2023.121728.

[85] Wang R, Zhou X, Dong L, Wen Y, Tan R, Chen L, Wang G, Zeng F. Kalibre: Knowledge-Based Neural Surrogate Model Calibration for Data Center Digital Twins. Proceedings of ACM BuildSys. 2020. doi:10.1145/3408308.3427982.

[86] Wang J, et al. Adaptive Digital Twin Modeling with Control: Integration of Extended Kalman Filter-Based Recursive Sparse Nonlinear Identification with Model Predictive Control. Sensors. 2026;26:1734. doi:10.3390/s26051734.

[87] Park J, Li C, Byon E, Bailey W, Hussain Q, Lonari Y. Hierarchical Parameter Calibration of Digital Twins at Ford Motor Company. INFORMS Journal on Applied Analytics. 2026. doi:10.1287/inte.2025.0259.

[88] Li X, Ye L, Bian X. Sequential Data Assimilation for Digital Twin Modeling of Shield Tunnel Structure-Soil

Interaction Systems. Tunnelling and Underground Space Technology. 2026;168(2):107168. doi:10.1016/j.tust.2025.107168.

[89] Nóvoa A, Noiray N, Dawson JR, Magri L. A Real-Time Digital Twin of Azimuthal Thermoacoustic Instabilities. Journal of Fluid Mechanics. 2024;1001:A49. doi:10.1017/jfm.2024.1052.

[90] Georgantzinos SK, Antoniou PA, Spitas C. A Multi-Scale Computational Framework for the Hygro-Thermo-Mechanical Analysis of Laminated Composite Structures with Carbon Nanotube Inclusions. Results in Engineering. 2023;17:100904. doi:10.1016/j.rineng.2023.100904.

[91] Georgantzinos SK, Giannopoulos GI, Markolefas SI. Vibration Analysis of Carbon Fiber-Graphene-Reinforced Hybrid Polymer Composites Using Finite Element Techniques. Materials. 2020;13(19):4225. doi:10.3390/ma13194225.

[92] Antoniou PA, Stamoulis KP, Georgantzinos SK. Progressive Damage and Failure Analysis of CNT-Reinforced Laminated Nanocomposite Structures: A Multiscale Modeling Framework. Engineering Failure Analysis. 2025;173:109452. doi:10.1016/j.engfailanal.2025.109452.

[93] Brunton SL, Proctor JL, Kutz JN. Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems. Proceedings of the National Academy of Sciences. 2016;113(15):3932–3937. doi:10.1073/pnas.1517384113.

[94] Raissi M, Perdikaris P, Karniadakis GE. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. Journal of Computational Physics. 2019;378:686–707. doi:10.1016/j.jcp.2018.10.045.

[95] Lu L, Jin P, Pang G, Zhang Z, Karniadakis GE. Learning Nonlinear Operators via DeepONet Based on the Universal Approximation Theorem of Operators. Nature Machine Intelligence. 2021;3:218–229. doi:10.1038/s42256-021-00302-5.

[96] Kovachki N, Li Z, Liu B, Azizzadenesheli K, Bhattacharya K, Stuart AM, Anandkumar A. Neural Operator: Learning Maps Between Function Spaces with Applications to PDEs. Journal of Machine Learning Research. 2023;24(89):1–97.

[97] Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-Informed Machine Learning. Nature Reviews Physics. 2021;3:422–440. doi:10.1038/s42254-021-00314-5.

[98] Voulgaris S, Chandrinos S, Chamatidis I, Kazakis G, Tsakalis P, Mitropoulou CC, Georgantzinos SK, Istrati D, Lagaros ND. Machine Learning and Data-Driven Methods for Steel Constitutive Modeling: A State-of-the-Art Review and Validation. Discover Civil Engineering. 2025;2:228. doi:10.1007/s44290-025-00389-4.

[99] Shao G, Hightower J, Schindel W. Credibility Consideration for Digital Twins in Manufacturing. Manufacturing Letters. 2023;35:24–28. doi:10.1016/j.mfglet.2022.11.009.

[100] Lang S, Zorzini M, Scholze S, Mayr J, Bambach M. Sensor Placement Utilizing a Digital Twin for Thermal Error Compensation of Machine Tools. Journal of Manufacturing Systems. 2025. doi:10.1016/j.jmsy.2025.03.003.

[101] Kammer DC. Sensor Placement for On-Orbit Modal Identification and Correlation of Large Space Structures. Journal of Guidance, Control, and Dynamics. 1991;14(2):251–259. doi:10.2514/3.20635.

[102] Papadimitriou C. Optimal Sensor Placement Methodology for Parametric Identification of Structural Systems. Journal of Sound and Vibration. 2004;278(4–5):923–947. doi:10.1016/j.jsv.2003.10.063.

[103] Manohar K, Brunton BW, Kutz JN, Brunton SL. Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns. IEEE Control Systems Magazine. 2018;38(3):63–86. doi:10.1109/MCS.2018.2810460.

[104] Chaloner K, Verdinelli I. Bayesian Experimental Design: A Review. Statistical Science. 1995;10(3):273–304. doi:10.1214/ss/1177009939.

[105] Huan X, Marzouk YM. Simulation-Based Optimal Bayesian Experimental Design for Nonlinear Systems. Journal of Computational Physics. 2013;232(1):288–317. doi:10.1016/j.jcp.2012.08.013.

[106] Evensen G, Vossepoel FC, van Leeuwen PJ. Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem. Cham: Springer; 2022. doi:10.1007/978-3-030-96709-3.

[107] Reich S, Cotter C. Probabilistic Forecasting and Bayesian Data Assimilation. Cambridge: Cambridge University Press; 2015. doi:10.1017/CBO9781107706804.

[108] Iglesias MA, Law KJH, Stuart AM. Ensemble Kalman Methods for Inverse Problems. Inverse Problems. 2013;29(4):045001. doi:10.1088/0266-5611/29/4/045001.

[109] Emerick AA, Reynolds AC. Ensemble Smoother with Multiple Data Assimilation. Computers & Geosciences. 2013;55:3–15. doi:10.1016/j.cageo.2012.03.011.

[110] Houtekamer PL, Zhang F. Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. Monthly Weather Review. 2016;144(12):4489–4532. doi:10.1175/MWR-D-15-0440.1.

[111] Lang S, Talleri S, Mayr J, Wegener K, Bambach M. Kalman Filter-Driven State Observer for Thermal Error Compensation in Machine Tool Digital Twins. Manufacturing Letters. 2024. doi:10.1016/j.mfglet.2024.09.025.

[112] van Leeuwen PJ, Künsch HR, Nerger L, Potthast R, Reich S. Particle Filters for High-Dimensional Geoscience Applications: A Review. Quarterly Journal of the Royal Meteorological Society. 2019;145(723):2335–2365. doi:10.1002/qj.3551.

[113] Branlard E, et al. Augmented Kalman Filter with a Reduced Mechanical Model to Estimate Tower Loads on a Land-Based Wind Turbine: A Step Towards Digital-Twin Simulations. Wind Energy Science. 2020;5:1155-1167. doi:10.5194/wes-5-1155-2020.

[114] Haseltine EL, Rawlings JB. Critical Evaluation of Extended Kalman Filtering and Moving-Horizon Estimation. Industrial & Engineering Chemistry Research. 2005;44(8):2451–2460. doi:10.1021/ie034308l.

[115] Rawlings JB, Bakshi BR. Particle Filtering and Moving Horizon Estimation. Computers & Chemical Engineering. 2006;30(10–12):1529–1541. doi:10.1016/j.compchemeng.2006.05.031.

[116] Alvarez-Cuesta M, et al. Which Data Assimilation Method to Use and When. Environmental Research Letters. 2024. doi:10.1088/1748-9326/ad3143.

[117] Willard J, Jia X, Xu S, Steinbach M, Kumar V. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. ACM Computing Surveys. 2023;55(4):66. doi:10.1145/3514228.

[118] Duraisamy K, Iaccarino G, Xiao H. Turbulence Modeling in the Age of Data. Annual Review of Fluid Mechanics. 2019;51:357–377. doi:10.1146/annurev-fluid-010518-040547.

[119] Quarteroni A, Manzoni A, Negri F. Reduced Basis Methods for Partial Differential Equations: An Introduction. Cham: Springer; 2016. doi:10.1007/978-3-319-15431-2.

[120] Lee K, Carlberg KT. Model Reduction of Dynamical Systems on Nonlinear Manifolds Using Deep Convolutional Autoencoders. Journal of Computational Physics. 2020;404:108973. doi:10.1016/j.jcp.2019.108973.

[121] Chandrinos S, Voulgaris S, Mitropoulou CC, Gkara M, Georgantzinos SK, Lagaros ND. Statistically Grounded Selection and Validation of Surrogate Models for Civil-Engineering Decision Support: ANN, Kriging and SVM Across Three Case Studies. Machine Learning for Computational Science and Engineering. 2026;2:35. doi:10.1007/s44379-026-00075-x.

[122] Gawlikowski J, Tassi CRN, Ali M, Lee J, Humt M, Feng J, et al. A Survey of Uncertainty in Deep Neural Networks. Artificial Intelligence Review. 2023;56(Suppl 1):1513–1589. doi:10.1007/s10462-023-10562-9.

[123] Zimmermann N, Breu T, Mayr J, Wegener K. Autonomously Triggered Model Updates for Self-Learning Thermal Error Compensation. CIRP Annals. 2021. doi:10.1016/j.cirp.2021.04.029.

[124] Lu J, Liu A, Dong F, Gu F, Gama J, Zhang G. Learning under Concept Drift: A Review. IEEE Transactions on Knowledge and Data Engineering. 2019;31(12):2346–2363. doi:10.1109/TKDE.2018.2876857.

[125] ASME. ASME V&V 40-2018: Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices. New York: American Society of Mechanical Engineers; 2018.

[126] Georgantzinos SK, Kastanos G, Tseni AD, Kostopoulos V. Efficient Optimization of the Multi-Response Problem in the Taguchi Method through Advanced Data Envelopment Analysis Formulations Integration. Computers & Industrial Engineering. 2024;197:110618. doi:10.1016/j.cie.2024.110618.

[127] Georgantzinos SK, Kostopoulos G, Papadopoulou E, Voulgaris S, Tseni A, Bakalis P, Gkara MN, Mitropoulou CC, Kazakis G, Munteanu A, Gancet J, Lagaros ND. Design and Prototyping of Topology-Optimized, Additively Manufactured Lightweight Components for Space Robotic Systems: A Case Study. Advanced Engineering Materials. 2026;28(4):e202502311. doi:10.1002/adem.202502311.

[128] Georgantzinos SK, Papadopoulou E, Bakalis P. Prediction of Anisotropic Mechanical Behavior in Fused Deposition Modeling-Printed Polylactic Acid Structures with Straight-Line Microstructure Using Taguchi Design. Journal of Materials Engineering and Performance. 2025;34(10):10218-10234. doi:10.1007/s11665-025-12381-1.

[129] Georgantzinos SK, Papadopoulou E. Digital Twin for Fused Filament Fabrication (FFF). Zenodo; 2026. doi:10.5281/zenodo.18270596.

[130] Georgantzinos SK, Papadopoulou E, Kostopoulos G, Voulgaris S, Tseni A, Bakalis P, Gkara M, Mitropoulou CC, Lagaros ND. A Comprehensive Review of Additive Manufacturing for Space Applications: Materials, Advances, Challenges, and Future Directions. Advanced Engineering Materials. 2025;27(22):e202501082. doi:10.1002/adem.202501082.

[131] Torzoni M, et al. A Digital Twin Framework for Civil Engineering Structures. Computer Methods in Applied Mechanics and Engineering. 2024;418:116584. doi:10.1016/j.cma.2023.116584.

[132] Chen H, Li W, Bao J, Shen Y. Multi-Scale Error-Triggered Latent Assimilation for Reduced-Order Digital Twins of Gas-Solid Flow Reactors. Chemical Engineering Journal. 2026. doi:10.1016/j.cej.2026.179496.

[133] Mucke NT, Pandey P, Jain S, Bohte SM, Oosterlee CW. A Probabilistic Digital Twin for Leak Localization in Water Distribution Networks Using Generative Deep Learning. Sensors. 2023;23:6179. doi:10.3390/s23136179.

[134] Irino N, Kobayashi A, Shinba Y, Kawai K, Spescha D, Wegener K. Digital Twin Based Accuracy Compensation. CIRP Annals. 2023. doi:10.1016/j.cirp.2023.04.088.

[135] Rozsas A, Slobbe A, Martini G, Jansen R. Structural and Load Parameter Estimation of a Real-World Reinforced Concrete Slab Bridge Using Measurements and Bayesian Statistics. Structural Concrete. 2022. doi:10.1002/suco.202100913.

[136] Koune I, Rozsas A, Slobbe A, Cicirello A. Bayesian System Identification for Structures Considering Spatial and Temporal Correlation. Data-Centric Engineering. 2023. doi:10.1017/dce.2023.18.

[137] ISO. ISO 23247-1:2021, Automation Systems and Integration—Digital Twin Framework for Manufacturing—Part 1: Overview and General Principles. Geneva: International Organization for Standardization; 2021.

[138] Mertens J, Denil J. Reusing Model Validation Methods for the Continuous Validation of Digital Twins of Cyber-Physical Systems. Software and Systems Modeling. 2025;24:1427–1449. doi:10.1007/s10270-024-01225-6.

[139] Gama J, Žliobaitė I, Bifet A, Pechenizkiy M, Bouchachia A. A Survey on Concept Drift Adaptation. ACM Computing Surveys. 2014;46(4):44. doi:10.1145/2523813.

[140] Xu Y, Park C. Online Bayesian Calibration under Gradual and Abrupt System Changes. arXiv preprint. 2026. doi:10.48550/arXiv.2605.06612.

[141] Torzoni M, Maisto D, Manzoni A, Donnarumma F, Pezzulo G, Corigliano A. Active Digital Twins via Active Inference. Engineering Applications of Artificial Intelligence. 2026;174:114519. doi:10.1016/j.engappai.2026.114519.

[142] Alexanderian A. Optimal Experimental Design for Infinite-Dimensional Bayesian Inverse Problems Governed by PDEs: A Review. Inverse Problems. 2021;37(4):043001. doi:10.1088/1361-6420/abe10c.

[143] Shao G. Manufacturing Digital Twin Standards. In: Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems Companion. 2024:370–377. doi:10.1145/3652620.3688250.

[144] Kibira D, Shao G. Data Requirements for a Digital Twin of a CNC Machine Tool. In: Proceedings of the 2025 Winter Simulation Conference; 7–10 December 2025; Seattle, WA, USA. Published 2026.