Novelty Assessment Report

Paper: RealBench: A Benchmark for Complex Physical Systems with Real-World Data

PDF URL: https://openreview.net/pdf?id=y3oHMcoItR

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-27

Abstract

Predicting the evolution of complex physical systems remains a central problem in science and engineering. Despite rapid progress in scientific Machine Learning (ML) models, a critical bottleneck is the lack of expensive real-world data, resulting in most current models being trained and validated on simulated data. Beyond limiting the development and evaluation of scientific ML, this gap also hinders research into essential tasks such as sim-to-real transfer. We introduce RealPDEBench, the first benchmark for scientific ML that integrates real-world measurements with paired numerical simulations. RealPDEBench consists of five datasets, three tasks, eight metrics, and ten baselines. We first present five real-world measured datasets with paired simulated datasets across different complex physical systems. We further define three tasks, which allow comparisons between real-world and simulated data, and facilitate the development of methods to bridge the two. Moreover, we design eight evaluation metrics, spanning data-oriented and physics-oriented metrics, and finally benchmark ten representative baselines, including state-of-the-art models, pretrained PDE foundation models, and a traditional method. Experiments reveal significant discrepancies between simulated and real-world data, while showing that pretraining with simulated data consistently improves both accuracy and convergence. In this work, we hope to provide insights from real-world data, advancing scientific ML toward bridging the sim-to-real gap and real-world deployment.

Core Task Landscape

This paper addresses: **Bridging the Sim-to-Real Gap in Complex Physical System Prediction**

A total of **50 papers** were analyzed and organized into a taxonomy with **13 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Sim-to-Real Transfer Methods for Robotic Control**
- **Surveys and Frameworks for Sim-to-Real Transfer**
- **Autonomous Systems and Embodied AI**
- **Digital Twins and Virtual Monitoring Systems**
- **Physics-Informed and Hybrid Modeling Approaches**

Complete Taxonomy Tree

- Bridging the Sim-to-Real Gap in Complex Physical System Prediction Survey Taxonomy
- Sim-to-Real Transfer Methods for Robotic Control
  - Domain Randomization and Dynamics Adaptation (5 papers)
  - [5] Sim-to-real transfer of robotic control with dynamics randomization (Peng, 2018) View paper
  - [7] Closing the sim-to-real loop: Adapting simulation randomization with real world experience (Chebotar, 2019) View paper
  - [10] Neural Fidelity Calibration for Informative Sim-to-Real Adaptation (Yu Youwei, 2025) View paper
  - [13] Sampling-Based System Identification with Active Exploration for Legged Robot Sim2Real Learning (He, 2025) View paper
  - [50] Bridging the Sim-to-Real Gap with Bayesian Inference (Rothfuss, 2024) View paper
  - Perception-Based Sim-to-Real Transfer (5 papers)
  - [19] Tactile sim-to-real policy transfer via real-to-sim image translation (Church, 2022) View paper
  - [23] Sim-to-real via latent prediction: Transferring visual non-prehensile manipulation policies (Carlo Rizzardo, 2023) View paper
  - [32] Sim-to-real for robotic tactile sensing via physics-based simulation and learned latent projections (Narang, 2021) View paper
  - [37] Grasp Stability Prediction with Sim-to-Real Transfer from Tactile Sensing (Zilin Si, 2022) View paper
  - [49] Machine visual perception from sim-to-real transfer learning for autonomous docking maneuvers (Derek Worth, 2024) View paper
  - Task-Specific Robotic Sim-to-Real Applications (5 papers)
  - [18] RL from Physical Feedback: Aligning Large Motion Models with Humanoid Control (Yue, 2025) View paper
  - [28] DexNDM: Closing the Reality Gap for Dexterous In-Hand Rotation via Joint-Wise Neural Dynamics Model (Liu Xueyi, 2025) View paper
  - [30] Learning Humanoid Standing-up Control across Diverse Postures (Tao Huang, 2025) View paper
  - [36] GraphGarment: Learning Garment Dynamics for Bimanual Cloth Manipulation Tasks (Chen Wei, 2025) View paper
  - [47] Learning Agile, Vision-based Drone Flight: from Simulation to Reality (Scaramuzza, 2023) View paper
- Surveys and Frameworks for Sim-to-Real Transfer
  - General Sim-to-Real Survey Literature (4 papers)
  - [3] Sim-to-real transfer in deep reinforcement learning for robotics: a survey (Zhao Wenshuai, 2020) View paper
  - [12] Crossing the reality gap: A survey on sim-to-real transferability of robot controllers in reinforcement learning (Erica Salvato, 2021) View paper
  - [17] A survey of sim-to-real methods in rl: Progress, prospects and challenges with foundation models (Da, 2025) View paper

    - [38] Sim-to-real transfer in robotics: Addressing the gap between simulation and real-world performance (Naomi Chukwurah, 2024) View paper
    - Framework Development and Methodological Foundations (5 papers)
    - [22] Towards mitigating Sim2Real gaps: A formal quantitative approach (P Sangeerth, 2025) View paper
    - [39] Robot model identification and learning: A modern perspective (Tae-Yoon Lee, 2023) View paper
    - [40] A Simulation Pipeline to Facilitate Real-World Robotic Reinforcement Learning Applications (Jefferson Silveira, 2025) View paper
    - [41] Engineering AI systems and AI for engineering: compositionality and physics in learning (Neary, 2024) View paper
    - [48] From abstraction to reality: DARPA's vision for robust simâ□□toâ□□real autonomy (Noorani, 2025) View paper
- Autonomous Systems and Embodied AI
    - Autonomous Driving and Navigation (5 papers)
    - [8] A Multi-Modality Evaluation of the Reality Gap in Autonomous Driving Systems (Lambertenghi, 2025) View paper
    - [16] LEGO-Motion: Learning-Enhanced Grids with Occupancy Instance Modeling for Class-Agnostic Motion Prediction (Miao Jinyu, 2025) View paper
    - [24] Sim-to-real transfer and reality gap modeling in model predictive control for autonomous driving (Ivan Garcia-Daza, 2023) View paper
    - [25] Sim2real predictivity: Does evaluation in simulation predict real-world performance? (Kadian, 2020) View paper
    - [34] Reproducible and Low-cost Sim-to-Real Environment for Traffic Signal Control (Yiran Zhang, 2025) View paper
    - Embodied AI Frameworks and Architectures (3 papers)
    - [2] Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI (Yang Liu, 2024) View paper
    - [6] Digital twins to embodied artificial intelligence: review and perspective (Junfei Li, 2025) View paper
    - [35] Sim-to-Real Transfer for a Robotics Task: Challenges and Lessons Learned (Jakob Jonas Rothert, 2024) View paper
- Digital Twins and Virtual Monitoring Systems
    - Infrastructure and Environmental Monitoring (3 papers)
    - [1] An operational IoT-based slope stability forecast using a digital twin (Luca Piciullo, 2025) View paper
    - [27] Bridging the Reality Gap in Digital Twins with Context-Aware, Physics-Guided Deep Learning (Berges, 2025) View paper
    - [33] Enhancing urban resilience: Smart city data analyses, forecasts, and digital twin techniques at the neighborhood level (Sotiris Kotsiantis, 2024) View paper
    - Energy System Forecasting and Control (5 papers)
    - [11] Machine learning-based digital twin for predictive modeling in wind turbines (Muhammad Fahim, 2022) View paper
    - [14] Research on multi-digital twin and its application in wind power forecasting (Shuwei Liu, 2024) View paper
    - [20] A physics-based domain adaptation framework for modeling and forecasting building energy systems (Zack Xuereb Conti, 2023) View paper
    - [29] From Simulation to Reality: A Study of Reinforcement Learning Control in Operational Building Environments (Xinlin Wang, 2025) View paper
    - [46] Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison (Neha Sehrawat, 2023) View paper
    - Industrial Equipment and Process Monitoring (4 papers)
    - [9] A calibration-based hybrid transfer learning framework for RUL prediction of rolling bearing across different machines (Yafei Deng, 2023) View paper
    - [21] Transitioning from simulation to reality: applying chatter detection models to real-world machining data (Matthew Alberts, 2024) View paper
    - [43] Real-is-Sim: Bridging the Sim-to-Real Gap with a Dynamic Digital Twin for Real-World Robot Policy Evaluation (Jad Abou-Chakra, 2025) View paper
    - [44] From Simulation to Field Validation: A Digital Twin-Driven Sim2real Transfer Approach for Strawberry Fruit Detection and Sizing (Omeed Mirbod, 2025) View paper
- Physics-Informed and Hybrid Modeling Approaches
    - Physics-Informed Neural Networks and Uncertainty Quantification ★ (3 papers)
    - [0] RealBench: A Benchmark for Complex Physical Systems with Real-World Data (Anon et al., 2026) View paper
    - [15] Calibrated Physics-Informed Uncertainty Quantification (Gopakumar, 2025) View paper
    - [45] Graph networks as learnable physics engines for inference and control (Ã□lvaro SÃ¡nchezâ□□GonzÃ¡lez, 2018) View paper
    - Hybrid Transfer Learning with Physics Priors (1 papers)
    - [4] Application of transfer learning on physics-based models to enhance vessel shaft power predictions (Stamatis Mavroudis, 2025) View paper
    - Model-Based Reinforcement Learning and Control (3 papers)
    - [26] Applying machine learning to electricity price forecasting in simulated energy market scenarios (Felix Nitsch, 2024) View paper
    - [31] Guided Model-Based Policy Search Method for Fast Motor Learning of Robots With Learned Dynamics (Xiao Huang, 2025) View paper
    - [42] Digital twin: architectures, networks, and applications (Zhang, 2024) View paper

Narrative

Core task: bridging the sim-to-real gap in complex physical system prediction. The field addresses the challenge of ensuring that models trained or validated in simulation can reliably predict real-world behavior across diverse physical systems. The taxonomy reveals five main branches: Sim-to-Real Transfer Methods for Robotic Control focuses on domain randomization, policy adaptation, and reinforcement learning techniques that enable robots to generalize from synthetic to physical environments (e.g., Dynamics Randomization[5], Closing Sim-to-Real Loop[7]). Surveys and Frameworks provide conceptual overviews and methodological guidance (Sim-to-Real Survey[3]). Autonomous Systems and Embodied AI emphasize end-to-end learning and perception-action loops in agents operating in real environments (Embodied AI Survey[2]). Digital Twins and Virtual Monitoring Systems create persistent virtual replicas for industrial assets, infrastructure, and energy systems (Digital Twin Turbines[11], Multi Digital Twin[14]). Physics-Informed and Hybrid Modeling Approaches integrate domain knowledge, neural networks, and uncertainty quantification to improve predictive fidelity and calibration (Calibrated Physics Informed[15], Graph Physics Engines[45]).

A particularly active line of work explores how to blend data-driven flexibility with physical constraints, trading off model expressiveness against interpretability and sample efficiency. Another contrasting theme is whether to adapt simulators to match reality through system identification and calibration, or to learn robust policies that tolerate discrepancies via randomization and domain adaptation. RealBench[0] sits within the Physics-Informed and Hybrid Modeling branch, specifically addressing physics-informed neural networks and uncertainty quantification. It shares methodological kinship with Calibrated Physics Informed[15], which also emphasizes calibration and uncertainty-aware prediction, and with Graph Physics Engines[45], which leverages structured representations of physical interactions. Where some works prioritize pure learning or pure physics, RealBench[0] occupies a middle ground by systematically benchmarking how well hybrid approaches can close the sim-to-real gap when physical priors and neural flexibility are combined with rigorous uncertainty estimates.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Calibrated Physics-Informed Uncertainty Quantification

Authors: Gopakumar, Vignesh, Gray, Ander, Zanisi, et al. (13 authors total) | **Year/Venue**: 2025 • International Conference on Machine Learning | **URL**: View paper

Abstract

Simulating complex physical systems is crucial for understanding and predicting phenomena across diverse fields, such as fluid dynamics and heat transfer, as well as plasma physics and structural mechanics. Traditional approaches rely on solving partial differential equations (PDEs) using numerical methods, which are computationally expensive and often prohibitively slow for real-time applications or large-scale simulations. Neural PDEs have emerged as efficient alternatives to these costly nume...

Relationship Analysis

Both papers belong to the Physics-Informed Neural Networks and Uncertainty Quantification category, focusing on neural network approaches that incorporate physical laws for complex system simulation. While RealPDEBench addresses the sim-to-real gap by providing paired real-world and simulated datasets across multiple physical systems (fluid dynamics, combustion, FSI) to benchmark prediction models, the candidate paper focuses specifically on developing a physics-informed conformal prediction framework for uncertainty quantification in neural PDE solvers without requiring labeled data. The key difference is that RealPDEBench provides a comprehensive benchmark infrastructure with real-world data to evaluate model performance, whereas the candidate paper contributes a methodological approach for calibrated uncertainty estimation in physics-informed models, particularly applied to plasma physics and fusion reactor applications.

---

2. Graph networks as learnable physics engines for inference and control

Authors: Ã□lvaro SÃ¡ncheza□□GonzÃ¡lez, Nicolas Heess, Jost Tobias Springenberg, Josh Merel, Martin Riedmiller, et al. (7 authors total) | **Year/Venue**: 2018 | **URL**: View paper

Abstract

Understanding and interacting with everyday physical scenes requires rich knowledge about the structure of the world, represented either implicitly in a value or policy function, or explicitly in a transition model. Here we introduce a new class of learnable models--based on graph networks--which implement an inductive bias for object- and relation-centric representations of complex, dynamical systems. Our results show that as a forward model, our approach supports accurate predictions from real...

Relationship Analysis

Both papers belong to the Physics-Informed Neural Networks and Uncertainty Quantification category, focusing on neural network approaches for complex physical system prediction. While RealBench addresses the sim-to-real gap by providing paired real-world and simulated datasets across multiple physical systems (fluid dynamics, combustion) with comprehensive benchmarking tasks and metrics, the candidate paper focuses on graph networks as learnable physics engines for inference and control in robotic systems. The key difference is that RealBench emphasizes benchmark creation and evaluation of existing models on real-world data, whereas the candidate paper introduces a novel graph network architecture for learning forward models and performing system identification in both simulated and real robotic environments.

Contributions Analysis

Overall novelty summary. The paper introduces RealPDEBench, a benchmark integrating real-world measurements with paired numerical simulations for scientific machine learning. It resides in the 'Physics-Informed Neural Networks and Uncertainty Quantification' leaf, which contains only three papers total. This leaf sits within the broader 'Physics-Informed and Hybrid Modeling Approaches' branch, indicating a relatively sparse research direction compared to the more crowded robotic control branches (15 papers across three leaves). The focus on benchmark infrastructure for PDE prediction distinguishes it from the sibling papers, which emphasize calibration methods and graph-based physics engines.

The taxonomy reveals that neighboring leaves address hybrid transfer learning with physics priors (1 paper) and model-based reinforcement learning (3 papers), both emphasizing policy learning rather than benchmark construction. The broader field structure shows that most sim-to-real work concentrates on robotic control (15 papers) and digital twin monitoring (13 papers), with physics-informed modeling receiving less attention (5 papers total). RealPDEBench diverges from these directions by targeting scientific ML evaluation infrastructure rather than control policies or industrial monitoring, occupying a niche at the intersection of data-driven learning and physics-based simulation validation.

Among 30 candidates examined, the contribution-level analysis shows varied novelty profiles. The paired real-world and simulated dataset contribution (10 candidates examined, 0 refutable) appears most distinctive, as no prior work provides this specific benchmark infrastructure. The three task categories (10 candidates, 0 refutable) also show no direct overlap. However, the comprehensive evaluation framework (10 candidates, 1 refutable) encounters at least one candidate offering overlapping metrics or evaluation approaches. Given the limited search scope, these statistics suggest the benchmark infrastructure itself is relatively novel, while the evaluation methodology has more substantial prior work within the examined candidates.

Based on the top-30 semantic matches and taxonomy structure, the work addresses a sparse research direction with limited direct competition in its specific leaf. The benchmark contribution appears more novel than the evaluation framework, though the restricted search scope means additional relevant work may exist beyond the candidates examined. The taxonomy context suggests this represents a meaningful but incremental step in a less-explored corner of the broader sim-to-real transfer landscape.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: RealPDEBench benchmark with paired real-world and simulated data

Description: The authors present RealPDEBench, the first scientific ML benchmark that systematically pairs real-world experimental measurements with numerical simulations across five complex physical systems. This benchmark includes more than 700 trajectories covering fluid dynamics and combustion scenarios, enabling systematic evaluation of models on real-world data and investigation of the sim-to-real gap.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Computational, Data-Driven, and Physics-Informed Machine Learning Approaches for Microstructure Modeling in Metal Additive Manufacturing

URL: View paper

Brief Assessment

Microstructure Additive Manufacturing[62] focuses on microstructure modeling in metal additive manufacturing using computational and ML approaches, not on creating benchmarks that pair real-world experimental measurements with numerical simulations for scientific ML evaluation across diverse physical systems.

---

2. Bulk Low-Inertia Power Systems Adaptive Fault Type Classification Method Based on Machine Learning and Phasor Measurement Units Data

URL: View paper

Brief Assessment

Fault Classification PMU[64] focuses on power system fault classification using PMU data and machine learning, not on scientific ML benchmarks pairing real-world measurements with numerical simulations for physical systems like fluid dynamics or combustion.

---

3. Scientific Machine Learning (SciML) - How the Fusion of AI and Physics is Giving Rise to Promising Simulation Methodologies

URL: View paper

Brief Assessment

Scientific Machine Learning[69] focuses on neural network architectures for silicide formation simulations in semiconductor technology, not on creating benchmarks that pair real-world experimental measurements with numerical simulations across physical systems.

---

4. Evaluating Universal Machine Learning Force Fields Against Experimental Measurements

URL: View paper

Brief Assessment

Universal Force Fields[70] focuses on evaluating machine learning force fields for atomistic simulations against experimental mineral structures, not on pairing real-world measurements with numerical simulations for scientific ML benchmarks in fluid dynamics and combustion systems.

---

5. Morpheus: Benchmarking Physical Reasoning of Video Generative Models with Real Physical Experiments

URL: View paper

Brief Assessment

Morpheus[66] focuses on evaluating video generation models' adherence to physical conservation laws using real-world videos, not on pairing real-world measurements with numerical simulations for scientific ML benchmarking.

---

6. On the prediction of critical heat flux using a physics-informed machine learning-aided framework

URL: View paper

Brief Assessment

Critical Heat Flux[67] focuses on a physics-informed ML framework for predicting critical heat flux in boiling systems, not on creating benchmarks that pair real-world measurements with numerical simulations for scientific ML evaluation.

---

7. Physics-informed deep-learning applications to experimental fluid mechanics

URL: View paper

Brief Assessment

Physics Fluid Mechanics[63] focuses on physics-informed neural networks for super-resolution of flow-field data from experimental measurements, not on creating a systematic benchmark pairing real-world and simulated datasets across multiple physical systems for scientific ML evaluation.

---

8. Real-time Fusion of Multi-Source Monitoring Data with Geotechnical Numerical Model Results using Data-driven and Physics-informed Sparse Dictionary Learning

URL: View paper

Brief Assessment

Sparse Dictionary Geotechnical[68] focuses on geotechnical engineering applications using sparse dictionary learning for model updating with monitoring data, not on creating a systematic benchmark pairing real-world measurements with numerical simulations across diverse physical systems for scientific ML evaluation.

---

9. Filtered partial differential equations: a robust surrogate constraint in physics-informed deep learning framework

URL: View paper

Brief Assessment

Filtered PDE Surrogate[61] focuses on improving physics-informed neural network training through filtered PDE constraints to handle noisy/sparse data, not on creating benchmarks that pair real-world measurements with simulations for systematic evaluation.

---

10. Predicting fusion ignition at the National Ignition Facility with physics-informed deep learning.

URL: View paper

Brief Assessment

Fusion Ignition Prediction[65] focuses on predicting fusion ignition outcomes using physics-informed deep learning, not on creating benchmarks that pair real-world measurements with numerical simulations for scientific ML evaluation.

Contribution 2: Three task categories for comparing real-world and simulated data

Description: The authors define three training paradigms: training on simulated data, training on real-world data, and pretraining on simulated data followed by finetuning on real-world data. These tasks enable systematic comparison of the strengths and limitations of both data types and provide a foundation for developing methods that effectively combine them.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Domain-adaptive neural networks improve supervised machine learning based on simulated population genetic data

URL: View paper

Brief Assessment

[Final Audit Failure] The model insisted on a refutation claim but failed to provide verifiable evidence after multiple retries. Marked as cannot_refute for safety. Please manually verify the candidate text.

---

2. Interpretable machine learning for science with PySR and SymbolicRegression. jl

URL: View paper

Brief Assessment

PySR[71] focuses on symbolic regression for discovering interpretable mathematical expressions from data, not on training paradigms comparing real-world versus simulated data in physical systems.

---

3. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis

URL: View paper

Brief Assessment

Real World RL[77] focuses on defining challenges for deploying RL in real-world systems (robotics, healthcare, recommender systems), not on training paradigms comparing real-world versus simulated data in scientific machine learning for physical systems like PDEs.

---

4. Physics informed synthetic image generation for deep learning-based detection of wrinkles and folds

URL: View paper

Brief Assessment

Synthetic Wrinkle Detection[79] focuses on a two-stage training methodology (synthetic pretraining followed by real data finetuning) for defect detection in manufacturing, not on systematic comparison frameworks across multiple training paradigms for scientific ML benchmarking.

---

5. Next-generation deep learning based on simulators and synthetic data

URL: View paper

Brief Assessment

Simulators Synthetic Data[74] focuses on synthetic data generation methods (graphics pipelines, GANs, fusion models) and domain adaptation techniques for deep learning, not on defining systematic task categories for comparing real-world versus simulated training paradigms in scientific ML benchmarks.

---

6. MLReal: Bridging the gap between training on synthetic data and real data applications in machine learning

URL: View paper

Brief Assessment

MLReal[76] focuses on domain adaptation techniques for seismic waveform data using cross-correlation and convolution operations, not on defining systematic task categories for comparing training paradigms across real-world and simulated data in scientific ML.

---

7. Transfer-learning: Bridging the gap between real and simulation data for machine learning in injection molding

URL: View paper

Brief Assessment

Injection Molding Transfer[80] focuses on transfer learning in injection molding manufacturing, not on defining systematic training paradigms for comparing real-world and simulated data across scientific ML domains.

---

8. From realâ□□world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges

URL: View paper

Brief Assessment

Individualized Treatment Effects[75] focuses on medical treatment effect estimation from observational patient data, not on comparing real-world and simulated data in physical systems or scientific machine learning contexts.

---

9. Combining machine learning and simulation to a hybrid modelling approach: Current and future directions

URL: View paper

Brief Assessment

Hybrid Modelling Approach[72] focuses on combining machine learning with simulation models in engineering contexts, not on defining training paradigms for comparing real-world versus simulated data in scientific ML benchmarks.

---

10. Merging physics-based synthetic data and machine learning for thermal monitoring of lithium-ion batteries: the role of data fidelity

URL: View paper

Brief Assessment

Thermal Monitoring Batteries[78] focuses on thermal monitoring of lithium-ion batteries using physics-based synthetic data and machine learning, not on general scientific ML benchmarks with systematic task categorization for comparing real-world and simulated data across multiple physical systems.

Contribution 3: Comprehensive evaluation framework with data-oriented and physics-oriented metrics

Description: The authors introduce a comprehensive evaluation framework consisting of eight metrics that assess model performance from both data-oriented perspectives (such as RMSE and MAE) and physics-oriented perspectives (such as Fourier Space Error and Kinetic Energy Error). They benchmark ten representative baselines including state-of-the-art models and pretrained foundation models using this framework.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Dataâ□□driven physicsâ□□based digital twins via a library of componentâ□□based reducedâ□□order models

URL: View paper

Brief Assessment

Component Reduced Models[53] focuses on digital twins for structural systems using Bayesian state estimation and component-based reduced-order models, not on evaluation frameworks for scientific ML models with combined data-oriented and physics-oriented metrics.

---

2. Physics-based vs. data-driven 24-hour probabilistic forecasts of precipitation for northern tropical Africa

URL: View paper

Brief Assessment

Physics Data Precipitation[56] focuses on precipitation forecasting evaluation metrics (CRPS, Brier score, reliability diagrams) rather than a general framework for scientific ML combining data-oriented and physics-oriented assessments across diverse physical systems.

---

3. Physics-informed machine learning for advancing computational medical imaging: integrating data-driven approaches with fundamental physical principles

URL: View paper

Prior Art Analysis

Physics Medical Imaging[55] demonstrates that comprehensive evaluation frameworks combining data-oriented and physics-oriented metrics have been established prior to the original paper. The candidate paper explicitly describes evaluation metrics spanning both data-driven measures (such as RMSE, MAE) and physics-based assessments in medical imaging contexts. Multiple sections detail the integration of physical principles with machine learning evaluation, including specific metrics for assessing model performance from both empirical accuracy and physical consistency perspectives. This prior work in physics-informed machine learning for medical imaging establishes that such dual-perspective evaluation frameworks existed before the original paper's benchmark proposal.

Evidence

Evidence 1 - **Rationale**: This evidence shows that Physics Medical Imaging[55] presents empirical evaluations that assess both data-driven performance (robustness, resolution) and physics-based consistency, establishing a precedent for the dual-metric evaluation framework claimed as novel by the original paper. - **Original**: these metrics can be categorized into data-oriented and physics-oriented metrics - **Candidate**: table 7 highlights empirical performance gains of piml methods compared to conventional approaches. across imaging modalities, piml consistently improves robustness, resolution, and physical consistency-validating its practical relevance for clinical imaging tasks. to better demonstrate the effectiv...

Evidence 2 - **Rationale**: This pair demonstrates that Physics Medical Imaging[55] explicitly describes evaluation frameworks that combine traditional data-oriented metrics (PSNR, SSIM, Dice score) with physics-specific metrics (PDE residual norms, divergence errors), establishing prior work on comprehensive dual-perspective evaluation before the original paper's benchmark. - **Original**: we provide a comprehensive set of evaluation metrics, comprising eight in total. specifically, these metrics can be categorized into data-oriented and physics-oriented metrics - **Candidate**: evaluation metrics typically combine traditional image quality indicators (e.g., psnr, ssim, dice score) with physics-specific metrics such as pde residual norms, divergence errors, or boundary violation penalties. physical priors are often computed using governing equations (e.g., diffusion, navier...

---

4. Integrating data-driven and physics-based approaches for robust wind power prediction: A comprehensive ML-PINN-Simulink framework

URL: View paper

Brief Assessment

ML-PINN-Simulink[51] focuses on wind power prediction using ML and physics-informed approaches, but does not propose a comprehensive evaluation framework with multiple data-oriented and physics-oriented metrics for benchmarking scientific ML models across diverse physical systems.

---

5. Joint physics-based and data-driven time-lapse seismic inversion: Mitigating data scarcity

URL: View paper

Brief Assessment

Joint Seismic Inversion[59] focuses on time-lapse seismic inversion for CO2 monitoring using MAE as a loss metric, not on developing a comprehensive evaluation framework with multiple data-oriented and physics-oriented metrics for benchmarking scientific ML models.

---

6. Driven by data or derived through physics? a review of hybrid physics guided machine learning techniques with cyber-physical system (cps) focus

URL: View paper

Brief Assessment

Hybrid Physics ML[57] focuses on evaluation metrics for hybrid physics-ML models in cyber-physical systems, not on benchmarking scientific ML models for PDE prediction with real-world data as in the original paper.

---

7. Data-driven, physics-based, or both: Fatigue prediction of structural adhesive joints by artificial intelligence

URL: View paper

Brief Assessment

Fatigue Prediction AI[54] focuses on fatigue lifetime prediction of adhesive joints using accuracy metrics (not spatiotemporal PDE modeling), without the dual data-oriented/physics-oriented metric framework proposed in the original paper.

---

8. A comparative study on methods for fusing data-driven and physics-based models for hybrid remaining useful life prediction of air filters

URL: View paper

Brief Assessment

Hybrid RUL Filters[58] focuses on hybrid prognostic methods for remaining useful life prediction in filtration systems, not on developing comprehensive evaluation frameworks with multiple metrics for scientific ML models across diverse physical systems.

---

9. Improving Typhoon Predictions by Integrating Dataâ□□Driven Machine Learning Model With Physics Model Based on the Spectral Nudging and Data Assimilation

URL: View paper

Brief Assessment

Typhoon Data Physics[60] focuses on typhoon prediction using hybrid ML-physics models with spectral nudging and data assimilation. It does not present a comprehensive evaluation framework with multiple data-oriented and physics-oriented metrics for benchmarking scientific ML models across diverse physical systems.

10. Global ionospheric sporadic intensity prediction from GNSS RO using a novel stacking machine learning method incorporated with physical observations

URL: View paper

Brief Assessment

Ionospheric Stacking ML[52] focuses on ionospheric sporadic E intensity prediction using machine learning with physical observations as inputs. It does not present a benchmark evaluation framework with multiple metrics for assessing scientific ML models across different physical systems.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] RealBench: A Benchmark for Complex Physical Systems with Real-World Data View paper
- [1] An operational IoT-based slope stability forecast using a digital twin View paper
- [2] Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI View paper
- [3] Sim-to-real transfer in deep reinforcement learning for robotics: a survey View paper
- [4] Application of transfer learning on physics-based models to enhance vessel shaft power predictions View paper
- [5] Sim-to-real transfer of robotic control with dynamics randomization View paper
- [6] Digital twins to embodied artificial intelligence: review and perspective View paper
- [7] Closing the sim-to-real loop: Adapting simulation randomization with real world experience View paper
- [8] A Multi-Modality Evaluation of the Reality Gap in Autonomous Driving Systems View paper
- [9] A calibration-based hybrid transfer learning framework for RUL prediction of rolling bearing across different machines View paper
- [10] Neural Fidelity Calibration for Informative Sim-to-Real Adaptation View paper
- [11] Machine learning-based digital twin for predictive modeling in wind turbines View paper
- [12] Crossing the reality gap: A survey on sim-to-real transferability of robot controllers in reinforcement learning View paper
- [13] Sampling-Based System Identification with Active Exploration for Legged Robot Sim2Real Learning View paper
- [14] Research on multi-digital twin and its application in wind power forecasting View paper
- [15] Calibrated Physics-Informed Uncertainty Quantification View paper
- [16] LEGO-Motion: Learning-Enhanced Grids with Occupancy Instance Modeling for Class-Agnostic Motion Prediction View paper
- [17] A survey of sim-to-real methods in rl: Progress, prospects and challenges with foundation models View paper
- [18] RL from Physical Feedback: Aligning Large Motion Models with Humanoid Control View paper
- [19] Tactile sim-to-real policy transfer via real-to-sim image translation View paper
- [20] A physics-based domain adaptation framework for modeling and forecasting building energy systems View paper
- [21] Transitioning from simulation to reality: applying chatter detection models to real-world machining data View paper
- [22] Towards mitigating Sim2Real gaps: A formal quantitative approach View paper
- [23] Sim-to-real via latent prediction: Transferring visual non-prehensile manipulation policies View paper
- [24] Sim-to-real transfer and reality gap modeling in model predictive control for autonomous driving View paper
- [25] Sim2real predictivity: Does evaluation in simulation predict real-world performance? View paper
- [26] Applying machine learning to electricity price forecasting in simulated energy market scenarios View paper
- [27] Bridging the Reality Gap in Digital Twins with Context-Aware, Physics-Guided Deep Learning View paper
- [28] DexNDM: Closing the Reality Gap for Dexterous In-Hand Rotation via Joint-Wise Neural Dynamics Model View paper
- [29] From Simulation to Reality: A Study of Reinforcement Learning Control in Operational Building Environments View paper
- [30] Learning Humanoid Standing-up Control across Diverse Postures View paper
- [31] Guided Model-Based Policy Search Method for Fast Motor Learning of Robots With Learned Dynamics View paper
- [32] Sim-to-real for robotic tactile sensing via physics-based simulation and learned latent projections View paper
- [33] Enhancing urban resilience: Smart city data analyses, forecasts, and digital twin techniques at the neighborhood level View paper
- [34] Reproducible and Low-cost Sim-to-Real Environment for Traffic Signal Control View paper
- [35] Sim-to-Real Transfer for a Robotics Task: Challenges and Lessons Learned View paper
- [36] GraphGarment: Learning Garment Dynamics for Bimanual Cloth Manipulation Tasks View paper
- [37] Grasp Stability Prediction with Sim-to-Real Transfer from Tactile Sensing View paper
- [38] Sim-to-real transfer in robotics: Addressing the gap between simulation and real-world performance View paper
- [39] Robot model identification and learning: A modern perspective View paper
- [40] A Simulation Pipeline to Facilitate Real-World Robotic Reinforcement Learning Applications View paper
- [41] Engineering AI systems and AI for engineering: compositionality and physics in learning View paper
- [42] Digital twin: architectures, networks, and applications View paper
- [43] Real-is-Sim: Bridging the Sim-to-Real Gap with a Dynamic Digital Twin for Real-World Robot Policy Evaluation View paper
- [44] From Simulation to Field Validation: A Digital Twin-Driven Sim2real Transfer Approach for Strawberry Fruit Detection and Sizing View paper
- [45] Graph networks as learnable physics engines for inference and control View paper
- [46] Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison View paper
- [47] Learning Agile, Vision-based Drone Flight: from Simulation to Reality View paper
- [48] From abstraction to reality: DARPA's vision for robust simâ□□toâ□□real autonomy View paper
- [49] Machine visual perception from sim-to-real transfer learning for autonomous docking maneuvers View paper
- [50] Bridging the Sim-to-Real Gap with Bayesian Inference View paper
- [51] Integrating data-driven and physics-based approaches for robust wind power prediction: A comprehensive ML-PINN-Simulink framework View paper
- [52] Global ionospheric sporadic intensity prediction from GNSS RO using a novel stacking machine learning method incorporated with physical observations View paper
- [53] Dataâ□□driven physicsâ□□based digital twins via a library of componentâ□□based reducedâ□□order models View paper

- [54] Data-driven, physics-based, or both: Fatigue prediction of structural adhesive joints by artificial intelligence View paper
- [55] Physics-informed machine learning for advancing computational medical imaging: integrating data-driven approaches with fundamental physical principles View paper
- [56] Physics-based vs. data-driven 24-hour probabilistic forecasts of precipitation for northern tropical Africa View paper
- [57] Driven by data or derived through physics? a review of hybrid physics guided machine learning techniques with cyber-physical system (cps) focus View paper
- [58] A comparative study on methods for fusing data-driven and physics-based models for hybrid remaining useful life prediction of air filters View paper
- [59] Joint physics-based and data-driven time-lapse seismic inversion: Mitigating data scarcity View paper
- [60] Improving Typhoon Predictions by Integrating Dataâ□□Driven Machine Learning Model With Physics Model Based on the Spectral Nudging and Data Assimilation View paper
- [61] Filtered partial differential equations: a robust surrogate constraint in physics-informed deep learning framework View paper
- [62] Computational, Data-Driven, and Physics-Informed Machine Learning Approaches for Microstructure Modeling in Metal Additive Manufacturing View paper
- [63] Physics-informed deep-learning applications to experimental fluid mechanics View paper
- [64] Bulk Low-Inertia Power Systems Adaptive Fault Type Classification Method Based on Machine Learning and Phasor Measurement Units Data View paper
- [65] Predicting fusion ignition at the National Ignition Facility with physics-informed deep learning. View paper
- [66] Morpheus: Benchmarking Physical Reasoning of Video Generative Models with Real Physical Experiments View paper
- [67] On the prediction of critical heat flux using a physics-informed machine learning-aided framework View paper
- [68] Real-time Fusion of Multi-Source Monitoring Data with Geotechnical Numerical Model Results using Data-driven and Physics-informed Sparse Dictionary Learning View paper
- [69] Scientific Machine Learning (SciML) - How the Fusion of AI and Physics is Giving Rise to Promising Simulation Methodologies View paper
- [70] Evaluating Universal Machine Learning Force Fields Against Experimental Measurements View paper
- [71] Interpretable machine learning for science with PySR and SymbolicRegression. jl View paper
- [72] Combining machine learning and simulation to a hybrid modelling approach: Current and future directions View paper
- [73] Domain-adaptive neural networks improve supervised machine learning based on simulated population genetic data View paper
- [74] Next-generation deep learning based on simulators and synthetic data View paper
- [75] From realâ□□world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges View paper
- [76] MLReal: Bridging the gap between training on synthetic data and real data applications in machine learning View paper
- [77] Challenges of real-world reinforcement learning: definitions, benchmarks and analysis View paper
- [78] Merging physics-based synthetic data and machine learning for thermal monitoring of lithium-ion batteries: the role of data fidelity View paper
- [79] Physics informed synthetic image generation for deep learning-based detection of wrinkles and folds View paper
- [80] Transfer-learning: Bridging the gap between real and simulation data for machine learning in injection molding View paper