

Combining simulations and measurements in physics-informed engineering surrogates

Vladimir Shapovalov
The Polytechnic University of Turin

Abstract

Engineering surrogates often combine inexpensive simulations, a smaller set of refined calculations, and limited physical measurements. These sources differ in numerical resolution, physical assumptions, observation operators, noise, and systematic discrepancy. Pooling them without representing those differences can produce a precise approximation to the wrong target. This review develops a multifidelity formulation for physics-informed flow and thermomechanical surrogates. Analytical examples show when adding a biased source helps or harms mean-square accuracy, why spatially averaged measurements cannot be treated as point values, and how a flexible discrepancy can confound physical parameter calibration. A proposed validation workflow separates fidelity transfer from case transfer and assigns computational and measurement budgets according to the engineering quantity of interest. Published PINN, thermal, and multifidelity studies connect the formulation to engineering applications.

1. Fidelity is a relationship, not a quality label

Multifidelity methods combine sources with different costs and levels of approximation to support prediction, uncertainty propagation, inference, or optimization [1]. A coarse mesh and a fine mesh can form one such hierarchy, but mesh resolution is only one difference. A reduced physical model may omit convection, plasticity, contact, or phase change. A laboratory measurement can have finer spatial detail than a simulation while still containing calibration bias. Calling one source high fidelity does not make it exact.

PINNs combine data with differential constraints [2, 3]. Multifidelity neural models add an explicit relationship between data sources. Meng and Karniadakis construct coupled networks to learn linear and nonlinear relationships between low- and high-fidelity information and extend that construction to inverse PDE examples [4]. This provides a concrete methodology; it does not establish that every added low-cost source improves a new engineering application. The correlation and discrepancy must be evaluated over the intended operating range.

The frequency-compensated flow model of Zhang et al. and the viscoplastic model of Arora et al. supply two engineering contexts for a possible hierarchy [5, 6]. Their published studies should not be redescribed as a validated fusion of arbitrary simulations and measurements. The fluid-design patent describes a candidate-evaluation system using learned flow predictions [7]; a multifidelity extension of that system would still require independent evaluation of its data relationships and design decisions.

2. Represent what each source actually observes

Let u be a physical field, θ the physical parameters, and s a source identifier. A useful observation description is

$$y_s = H_s(u, \theta) + \delta_s + \varepsilon_s.$$

Here H_s maps the physical state to the measured or reported quantity, δ_s represents systematic discrepancy, and ε_s represents a stochastic error under an explicitly chosen model. For a simulation source, discrepancy can include discretization and omitted physics. For a sensor, H_s can include spatial averaging, temporal response, projection, and coordinate registration. The decomposition is a modeling choice and need not be identifiable from the available data.

Hybrid thermal PINN work combines partially observed infrared temperature information with physical equations [8]. Additive-manufacturing flow and thermal modeling offers a related process setting [9]. These sources motivate careful measurement maps: an infrared camera observes radiation affected by surface and instrument assumptions, and an embedded thermocouple responds over a finite volume and time. A pointwise temperature field should be compared with a measurement only after the relevant observation operation has been applied.

Consider a one-dimensional quadratic temperature field $T(x) = T_0 + a x^2$. An ideal sensor centered at x_0 averages uniformly over a width h . Direct integration gives

$$y = (1/h) \int_{x_0-h/2}^{x_0+h/2} T(x) \, dx = T(x_0) + a h^2/12.$$

The difference from the central point value is deterministic and depends on curvature and sensor width. Treating it as independent zero-mean noise can mislead parameter identification. Refining the neural output grid does not remove the difference because it is caused by the observation map. For a linear field the corresponding averaging bias vanishes, illustrating why a sensor approximation that works in a smooth region may fail near a localized source.

Source difference	Mathematical location	Independent check
Coarse spatial or temporal discretization	Simulator discrepancy	Mesh and time-step refinement
Omitted physical mechanism	Model discrepancy	Regime-specific comparison or experiment
Sensor averaging and response	Observation operator	Instrument characterization
Calibration offset or drift	Systematic measurement component	Reference measurement and repeated calibration
Random acquisition variation	Noise model	Replicate measurements and covariance checks

3. How biased information changes an estimator

Suppose a high-quality estimator H of a scalar engineering quantity has mean μ and variance v_H . A cheaper estimator L has mean $\mu + b$ and variance v_L . Let their covariance be c , and combine them as $T_w = wH + (1 - w)L$. The mean-square error relative to μ is

$$\text{MSE}(T_w) = w^2 v_H + (1-w)^2 (v_L + b^2) + 2w(1-w)c.$$

If the quadratic coefficient $v_H + v_L + b^2 - 2c$ is positive, differentiation gives the unconstrained minimizing weight

$$w^* = (v_L + b^2 - c) / (v_H + v_L + b^2 - 2c).$$

If the intended combination is convex, restrict w to the interval from zero to one. The formula assumes that bias, variances, and covariance are known. They usually must be estimated, so the theoretical optimum is not an immediately available engineering recipe. Its value is explanatory: a low-variance source may deserve little influence when its bias is large, and shared errors affect the value of combining sources.

For a hypothetical independent example, let $v_H = 1$, $v_L = 0.04$, and $b = 2$. Equal weighting gives $\text{MSE} = 1.26$, worse than using H alone. The unconstrained optimal weight on H is $4.04/5.04$, about 0.802 , giving MSE about 0.802 . If the bias were incorrectly assumed zero, the resulting weight on H would be about 0.0385 , producing actual MSE about 3.737 . These direct calculations demonstrate how the assumed bias changes the optimal weighting and resulting mean-square error.

Increasing the number of cheap samples can reduce v_L without removing b . In the limit of vanishing sampling variance, a systematic discrepancy remains. This is why an enormous low-fidelity dataset should not dominate automatically through raw sample count. The same issue appears when summed losses give a source greater influence simply because it contains more observations. A mean reduction avoids that particular count effect but does not decide the appropriate physical or statistical weight.

4. Discrepancy and calibration can compensate each other

Computer-model calibration distinguishes uncertain physical parameters from discrepancy between model and reality [10]. In a physics-informed setting, the same distinction must be maintained. Suppose a scalar response is modeled as $y = \theta x + \delta(x)$. If δ is unrestricted, any candidate θ' can reproduce the same observations by choosing $\delta'(x) = \delta(x) + (\theta - \theta')x$. Data fit alone then cannot identify θ .

This exact reparameterization is a simple example of confounding. Restricting the discrepancy, adding independent physical measurements, or using additional experiments can help, but each introduces assumptions or information. A smoothness penalty on δ selects among explanations; it does not prove which explanation is physically true. Identifiability analysis provides a useful

conceptual framework [11], while conditional inverse-problem stability results apply to their specified equations and observations [12]. Neither justifies arbitrary nonlinear parameter recovery.

In steady conduction, source magnitude and conductivity can compensate each other through their ratio. In a flow model, inlet or pressure uncertainty can compensate viscosity. In viscoplastic calibration, temperature, loading rate, and material strength can have interacting effects. Mixed-variable flow and solid formulations affect numerical representation but do not create independent parameter information [13, 14]. Thermodynamic structure can restrict admissible constitutive behavior without eliminating every calibration ambiguity [15].

A useful study therefore reports which parameters are estimated, which discrepancies are permitted, and which quantities are independently known. Compare predictions on experiments that change the relative influence of those components. A model that interpolates the calibration experiment but fails under a new excitation has not established a transferable physical parameter estimate.

5. A loss that preserves the meaning of each source

For source residual $e_s = H_s(u_\theta) - y_s$ and a specified positive-definite covariance Σ_s , a Gaussian observation model contributes a quadratic term $e_s^T \Sigma_s^{-1} e_s$, along with the covariance normalization when covariance is inferred. A PDE residual contributes a different object: a physical consistency penalty over a stated measure. Its multiplier is not automatically an inverse sensor variance. Boundary uncertainties and model discrepancy require separate treatment.

Normalize variables and differential operators consistently before combining sources. Flow momentum, heat balance, displacement, stress, and internal-state residuals have different dimensions and characteristic scales. The viscoplastic study explicitly addresses the balance among its mechanical loss terms [6]. Gradient-pathology and neural-tangent-kernel analyses show why the numerical gradients can remain imbalanced even when residual magnitudes appear comparable [16, 17]. Failure-mode studies caution that a favorable aggregate loss can conceal unresolved components [18].

Adaptive weighting can change optimization priorities [19, 20]. It should not be interpreted automatically as learning a physical noise covariance or detecting simulator bias. Likewise, residual-adaptive and importance sampling alter where equations are tested or how their norms are estimated [21, 22]. A paper should report whether source weights are fixed, statistically estimated, or chosen for optimization, and whether the sampling measure changes during training.

Some physical restrictions can be enforced structurally, such as selected boundary values [23]. Weak-form formulations provide another way to test equations and manage derivative requirements [24]. These approaches change the relationship between a physical model and its approximation. They do not resolve a conflict caused by the wrong PDE, an unmodeled boundary flux, or an inaccurate measurement operator. Such conflicts require model diagnosis rather than merely stronger penalties.

6. Source dependence and leakage across physical cases

Low- and high-fidelity simulations of the same case are related observations. If the high-fidelity field is placed in a nominal test set but its low-fidelity counterpart was used to train the predictor, the model has case-specific information. This can be exactly the intended deployment workflow, but it tests fidelity correction rather than completely unseen-case prediction. Both evaluations are useful and should be named separately.

Maintain identifiers for geometry, operating history, material batch, solver family, mesh, time step, sensor configuration, and acquisition session. Split at the physical-case level before generating many coordinate samples. A million points from one thermal movie do not constitute a million independent experiments. Shared preprocessing statistics and jointly fitted discrepancy models must also be restricted to the development data.

For transient problems, document synchronization and temporal filtering. A small time shift between measurement and simulation can resemble a temperature or phase discrepancy, particularly around rapid heating or oscillatory wakes. Frequency-aware representations can help fit periodic dynamics [5, 25, 26], but cannot establish that the measurement clock is correct. Causal training changes temporal optimization and should be reported separately from correcting timing errors [27].

A source-provenance table should accompany a reproducible release. It should state which observations are paired, which cases are unseen, which sources are available at prediction time, and whether target-case refinement occurs. The final assessment must preserve the same information pattern as the proposed engineering use.

7. Allocate the expensive information to a decision

The next acquisition can be a high-fidelity simulation, an additional physical measurement, a replicated calibration, or a new low-cost case. Each reduces a different uncertainty. Sparse-sensor methods provide tools for selecting informative observations [28], while multifidelity methods address broader allocation among models and costs [1]. A low-fidelity sample in a poorly covered region may be useful for exploration but insufficient to validate a mechanism absent from the low-fidelity physics.

As a proposed decision rule, estimate the expected reduction in error or uncertainty of an engineering quantity per unit acquisition cost. Compare that rule with simpler space-filling and fixed-budget allocations. If the desired quantity is peak stress, acquisition should not be optimized only for average temperature reconstruction. Thermal-stress modeling illustrates how downstream mechanical quantities can require different information from the thermal field alone [29].

Bayesian PINNs and ensembles offer different ways to represent predictive uncertainty [30, 31]. Their uncertainty should be evaluated on held-out cases and separated by source and operating regime. A confident agreement among models sharing the same biased simulator can understate

reality discrepancy. Replicate measurements and independent reference conditions remain important even when ensemble spread is small.

Operator learning can amortize repeated predictions across cases [32, 33, 34]. However, a larger operator does not establish a valid fidelity relationship. The choice between a neural field, an operator, and a simpler discrepancy regression should follow the amount and structure of available data. Include the computational cost of generating data, training, acquisition selection, and target-case adaptation in the comparison.

8. A concrete proposed thermal validation study

Consider a moving-source heating process with an inexpensive conduction model, a refined thermal reference, and surface-temperature measurements. The proposed study would first specify which physical effects differ between the simulators and which instrument assumptions enter the measurements. It would then hold out complete source-speed and power combinations, rather than random pixels, and reserve a final group acquired in an independent session.

Compare four baselines: high-quality data only, low-fidelity data only, direct pooling, and explicit discrepancy learning. Keep the expensive-data budget fixed. Evaluate temperature histories, peak temperature, cooling rate, and a downstream quantity such as thermal strain or a specified temperature-threshold duration. A secondary comparison can hold total computational cost fixed to reveal whether the fusion procedure is economical as well as accurate.

Introduce controlled stress tests one at a time: a low-fidelity source-power bias, a sensor offset, a temporal lag, and an omitted heat-loss mechanism. These perturbations have different signatures and should not be collapsed into generic noise. Measure when the model detects disagreement, when it compensates through an incorrect parameter, and when a new measurement restores prediction. Report failures as well as average errors.

For an energy check, integrate the thermal balance over the domain and compare input energy, stored energy, and boundary losses under the chosen model. A good pointwise fit can coexist with an incorrect integrated balance if data are sparse or loss weights permit compensation. Conversely, an exact global balance does not guarantee a resolved hot spot. Both local and integral diagnostics are required for the intended engineering decision.

The conventional reference should include discretization checks and stated tolerances. Published FEM comparisons warn against assuming that a PINN automatically improves total solution cost [35]. Reusable software and benchmark suites support implementation and reporting [36, 37, 38], while training guidance helps make numerical choices reproducible [39]. Application-specific validation must still establish whether the additional source improves the quantities that matter.

9. Conclusions

Combining simulations and measurements is beneficial only when their relationship is modeled and tested. Observation averaging, systematic bias, source dependence, and discrepancy-parameter confounding can dominate the effect of adding more samples. The analytical examples show that those issues persist even with perfect optimization of a flexible predictor.

A defensible physics-informed multifidelity workflow keeps the observation maps and physical residuals distinct, validates case transfer separately from fidelity correction, and allocates expensive information according to the engineering decision. Its evidence should be an improvement at a stated accuracy and information budget, with uncertainty and failure modes assessed on independent physical cases.

References

- [1] Benjamin Peherstorfer, Karen Willcox, Max Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. arXiv:1806.10761 (2018). arXiv.
- [2] Maziar Raissi, Paris Perdikaris, George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics* 378, 686-707 (2019). DOI.
- [3] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, et al.. Physics-informed machine learning. *Nature Reviews Physics* 3, 422-440 (2021). DOI.
- [4] Xuhui Meng, George Em Karniadakis. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems. arXiv:1903.00104 (2019). arXiv.
- [5] Tongtao Zhang, Biswadip Dey, Pratik Kakkar, Arindam Dasgupta, Amit Chakraborty. Frequency-compensated PINNs for Fluid-dynamic Design Problems. *ML4Eng Workshop, NeurIPS 2020*. arXiv:2011.01456 (2020). arXiv.
- [6] Rajat Arora, Pratik Kakkar, Amit Chakraborty, Biswadip Dey. Physics-informed neural networks for modeling rate- and temperature-dependent plasticity. *Machine Learning and the Physical Sciences Workshop, NeurIPS 2022*. arXiv:2201.08363v3 (2022). arXiv.
- [7] Biswadip Dey, Pratik Kakkar, Ankur Kaushik, Arindam Dasgupta, Amit Chakraborty, Tongtao Zhang. Frequency-compensated neural networks for fluid dynamics design problems. WO2021163309A1. PCT/US2021/017635. Published 19 August 2021 (2021). Source.
- [8] Shuheng Liao, Tianju Xue, Jihoon Jeong, et al.. Hybrid thermal modeling of additive manufacturing processes using physics-informed neural networks for temperature prediction and parameter identification. arXiv:2206.07756 (2022). arXiv.
- [9] Qiming Zhu, Zeliang Liu, Jinhui Yan. Machine learning for metal additive manufacturing: Predicting temperature and melt pool fluid dynamics using physics-informed neural networks. arXiv:2008.13547 (2020). arXiv.
- [10] M. C. Kennedy, A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2001). DOI.
- [11] Alejandro F. Villaverde. Observability and Structural Identifiability of Nonlinear Biological Systems. arXiv:1812.04525 (2018). arXiv.
- [12] Siddhartha Mishra, Roberto Molinaro. Estimates on the generalization error of Physics Informed Neural Networks (PINNs) for approximating a class of inverse problems for PDEs. arXiv:2007.01138 (2020). arXiv.
- [13] Chengping Rao, Hao Sun, Yang Liu. Physics-informed deep learning for incompressible laminar flows. arXiv:2002.10558 (2020). arXiv.
- [14] Ehsan Haghighat, Maziar Raissi, Adrian Moure, et al.. A deep learning framework for solution and discovery in solid mechanics. arXiv:2003.02751 (2020). arXiv.

- [15] Filippo Masi, Ioannis Stefanou, Paolo Vannucci, Victor Maffi-Berthier. Thermodynamics-based Artificial Neural Networks for constitutive modeling. arXiv:2005.12183 (2020). arXiv.
- [16] Sifan Wang, Yujun Teng, Paris Perdikaris. Understanding and mitigating gradient pathologies in physics-informed neural networks. arXiv:2001.04536 (2020). arXiv.
- [17] Sifan Wang, Xinling Yu, Paris Perdikaris. When and why PINNs fail to train: A neural tangent kernel perspective. arXiv:2007.14527 (2020). arXiv.
- [18] Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, et al.. Characterizing possible failure modes in physics-informed neural networks. arXiv:2109.01050 (2021). arXiv.
- [19] Levi McClenny, Ulisses Braga-Neto. Self-Adaptive Physics-Informed Neural Networks using a Soft Attention Mechanism. arXiv:2009.04544 (2020). arXiv.
- [20] Rafael Bischof, Michael Kraus. Multi-Objective Loss Balancing for Physics-Informed Deep Learning. arXiv:2110.09813 (2021). arXiv.
- [21] Chenxi Wu, Min Zhu, Qinyang Tan, et al.. A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. arXiv:2207.10289 (2022). arXiv.
- [22] Mohammad Amin Nabian, Rini Jasmine Gladstone, Hadi Meidani. Efficient training of physics-informed neural networks via importance sampling. arXiv:2104.12325 (2021). arXiv.
- [23] N. Sukumar, Ankit Srivastava. Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks. arXiv:2104.08426 (2021). arXiv.
- [24] E. Kharazmi, Z. Zhang, G. E. Karniadakis. Variational Physics-Informed Neural Networks For Solving Partial Differential Equations. arXiv:1912.00873 (2019). arXiv.
- [25] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, et al.. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. arXiv:2006.10739 (2020). arXiv.
- [26] Sifan Wang, Hanwen Wang, Paris Perdikaris. On the eigenvector bias of Fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks. arXiv:2012.10047 (2020). arXiv.
- [27] Sifan Wang, Shyam Sankaran, Paris Perdikaris. Respecting causality is all you need for training physics-informed neural networks. arXiv:2203.07404 (2022). arXiv.
- [28] Krithika Manohar, Bingni W. Brunton, J. Nathan Kutz, Steven L. Brunton. Data-Driven Sparse Sensor Placement for Reconstruction. arXiv:1701.07569 (2017). arXiv.
- [29] R. Sharma, Y. B. Guo. Thermal-Mechanical Physics Informed Deep Learning For Fast Prediction of Thermal Stress Evolution in Laser Metal Deposition. arXiv:2412.18786 (2024). arXiv.
- [30] Liu Yang, Xuhui Meng, George Em Karniadakis. B-PINNs: Bayesian Physics-Informed Neural Networks for Forward and Inverse PDE Problems with Noisy Data. arXiv:2003.06097 (2020). arXiv.
- [31] Balaji Lakshminarayanan, Alexander Pritzel, Charles Blundell. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. arXiv:1612.01474 (2016). arXiv.
- [32] Lu Lu, Pengzhan Jin, George Em Karniadakis. DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. arXiv:1910.03193 (2019). arXiv.
- [33] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, et al.. Fourier Neural Operator for Parametric Partial Differential Equations. arXiv:2010.08895 (2020). arXiv.
- [34] Zongyi Li, Hongkai Zheng, Nikola Kovachki, et al.. Physics-Informed Neural Operator for Learning Partial Differential Equations. arXiv:2111.03794 (2021). arXiv.
- [35] Tamara G. Grossmann, Urszula Julia Komorowska, Jonas Latz, Carola-Bibiane Schönlieb. Can Physics-Informed Neural Networks beat the Finite Element Method?. arXiv:2302.04107 (2023). arXiv.
- [36] Lu Lu, Xuhui Meng, Zhiping Mao, George E. Karniadakis. DeepXDE: A deep learning library for solving differential equations. arXiv:1907.04502 (2019). arXiv.

- [37] Zhongkai Hao, Jiachen Yao, Chang Su, et al.. PINNacle: A Comprehensive Benchmark of Physics-Informed Neural Networks for Solving PDEs. arXiv:2306.08827 (2023). arXiv.
- [38] Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, et al.. PDEBENCH: An Extensive Benchmark for Scientific Machine Learning. arXiv:2210.07182 (2022). arXiv.
- [39] Sifan Wang, Shyam Sankaran, Hanwen Wang, Paris Perdikaris. An Expert's Guide to Training Physics-informed Neural Networks. arXiv:2308.08468 (2023). arXiv.