UNED: ONE-SHOT UNCERTAINTY-AWARE NEURAL EXPERIMENTAL DESIGN FOR TRANSIENT PDES

Mahdi Masmoudi
Michigan State University
East Lansing, MI 48823, USA

Xuyang Li
University of North Carolina at Charlotte
Charlotte, NC 28223, USA

Rami Gharbi, Nizar Lajnef & Vishnu Boddeti
Michigan State University
East Lansing, MI 48823, USA

ABSTRACT

Partial differential equations (PDEs) are foundational tools for modeling complex physical systems. A central challenge is inferring spatially or temporally varying PDE parameters from limited observations, a task made difficult by scarce and noisy data. Since running physical experiments is often costly, experimental design (ED) is commonly performed using PDE simulations to select the most informative sensor configurations before data collection. This is particularly important under strict budgets that allow only one-shot deployment, where uncertainty from measurement noise or surrogate models can significantly affect parameter recovery. Existing ED methods often optimize sensor placement assuming a fixed model realization or noise structure, and rarely aim to explicitly minimize uncertainty or ensure robustness to model and noise variability.

We introduce UNED, a differentiable experimental design framework that enables efficient gradient-based optimization of sensor locations to minimize expected posterior uncertainty for one-shot deployment. Rather than tuning designs for a single realization, UNED accounts for randomness in model initialization and observation noise, enabling robust sensor placement in settings with spatially distributed parameters. We validate UNED through experiments based on noisy simulated data and real-world experimental data. By explicitly minimizing uncertainty using an unsupervised uncertainty quantification (UQ) model, it outperforms existing ED methods in solving inverse problems.

Our results demonstrate that explicitly minimizing expected uncertainty under model and noise randomness leads to robust sensor placements whose performance generalizes across realizations, with improved parameter recovery emerging as a direct consequence rather than a design objective.

1 INTRODUCTION

Many systems across science and engineering are governed by partial differential equations (PDEs). Given a PDE and its parameters, forward simulation enables predicting system dynamics under specified conditions. In practice, however, the governing parameters are rarely known and must be inferred from observations of system behavior. This gives rise to PDE-constrained inverse problems (IPs) Stuart (2010); Kaipio & Somersalo (2005), where the objective is to recover unknown parameters that may vary in space or time. Solving such IPs is challenging: the dependence of the PDE solution on its parameters can be highly nonlinear, and available observations are often sparse, noisy, and expensive to acquire. Since collecting extensive measurements is usually infeasible, one must carefully decide where to place sensors under strict experimental budgets. Moreover, sensor configurations that are optimal for a particular noise realization or model initialization may perform poorly under different sources of uncertainty, making robust experimental design a crucial component of reliable parameter inference.

In recent years, the rapid rise of physics-informed machine learning (PIML) has led to a surge of interest in data-driven solvers for PDE-based inverse problems Karniadakis et al. (2021). Approaches range from physics-informed neural networks (PINNs) Raissi et al. (2019) to operator-learning surrogates such as DeepONet Lu et al. (2021), the Fourier neural operator (FNO) Li et al. (2021), and related neural operators Kovachki et al. (2023), as well as physics-informed operator variants Wang et al. (2021) and other deep learning models Franco et al. (2023), which have been promoted as replacements for classical numerical solvers. While these methods show promise in settings where simulation cost is prohibitive, a growing body of evidence indicates that reported gains can be highly sensitive to weak baselines and reporting biases, potentially overstating their true performance and generalization capabilities McGreivy & Hakim (2024). In particular, when applied to parameter estimation tasks, such models often require extensive training data and careful tuning, and may still converge more slowly and less reliably than traditional numerical approaches.

An alternative line of work exemplified by the *Neptune* method Li et al. (2025) demonstrates the potential of efficient numerical solvers for inverse problems. Neptune relies on a classical forward solver embedded within an inverse optimization framework, achieving significantly faster and more accurate parameter estimation than PINN-based counterparts. This highlights the continued importance of robust numerical solvers for inverse problems, especially in experimental design contexts where efficiency, stability, and reliability are paramount.

Recent work on experimental design (ED) for parameter inference in PDE-based inverse problems has advanced along several distinct lines, yet each carries important limitations. A class of methods targets sensor placement through sparsity-inducing penalties, such as ℓ_0 relaxations or total-variation regularization, that promote a minimal subset of active sensors Haber et al. (2012); Alexanderian et al. (2014); Neitzel et al. (2018); Attia et al. (2023). While mathematically rigorous, these approaches are largely confined to linear-Gaussian inverse problems or linearized PDE models, where posterior covariance admits tractable low-rank representations. As a result, they cannot scale to highly nonlinear regimes, nor do they naturally accommodate severe data scarcity or variability in noise and model realizations that characterize real experimental deployments.

Bayesian methods represent a second major category, formulating ED as the maximization of expected information gain Bernardo (1979); Chaloner & Verdinelli (1995) or mutual information between parameters and observations Foster et al. (2019); Belghazi et al. (2018); Kleinegesse & Gutmann (2020); Kennamer et al. (2023). These frameworks are theoretically optimal in that they explicitly minimize posterior uncertainty, but they require repeated PDE solves or extensive simulation-based training to estimate information gain, making them prohibitively expensive in large-scale nonlinear settings. Neural information estimators such as MINE Belghazi et al. (2018) and amortized variational bounds Foster et al. (2019); Kennamer et al. (2023) alleviate some computational burdens, but still depend on vast synthetic datasets, contradicting the data-scarce nature of many sensing problems. More recently, physics-informed approaches such as PIED Hemachandra et al. (2025) leverage PINNs to embed PDE solvers directly in a differentiable ED framework, enabling one-shot optimization of sensor placement in nonlinear PDEs. While PIED substantially reduces computational overhead relative to classical Bayesian methods, it primarily optimizes reconstruction error proxies (e.g., FIST, MoTE) rather than explicitly minimizing posterior uncertainty.

Taken together, existing methods share a key shortcoming: they do not explicitly design sensor placements to minimize uncertainty in recovering spatially distributed parameters under model and noise randomness. These limitations underscore the need for a new class of ED methods that are uncertainty-aware, computationally efficient, and capable of robust sensor placement for PDE-constrained inverse problems.

Overall, we make the following key contributions:

- We introduce UNED a differentiable ED framework that performs continuous optimization of sensor placement for one-shot deployments.
- UNED optimizes sensor locations to minimize expected posterior uncertainty, explicitly accounting for both model and measurement noise variability.
- UNED enables robust one-shot deployment and accurately recovers spatially distributed field parameters under limited data.

Table 1: Comparison of existing methods with respect to key criteria for experimental design in PDE-based inverse problems. Representative works include Bayesian OED Foster et al. (2019); Belghazi et al. (2018), PINN-based ED Hemachandra et al. (2025), and Sparsity-based OED Neitzel et al. (2018). †Gradient-based optimization is possible with variational or surrogate-based EIG estimators, but not end-to-end differentiable with PDE solvers. ‡PINN-based methods can be applied for field parameter inference, but performance is often poor in practice.

Method	One-shot deploy	Scarce data	Field parameter	UQ-aware	Gradient-based opt.
Bayesian OED	✗	✗	✗	✓	Partial†
PINN-based ED	✓	✗	Partial‡	✗	✓
Sparsity OED	✗	✓	✗	✓	✓
Ours (UNED)	✓	✓	✓	✓	✓

- We demonstrate the effectiveness and versatility of UNED with experiments on the wave equation and real-world cell migration data.

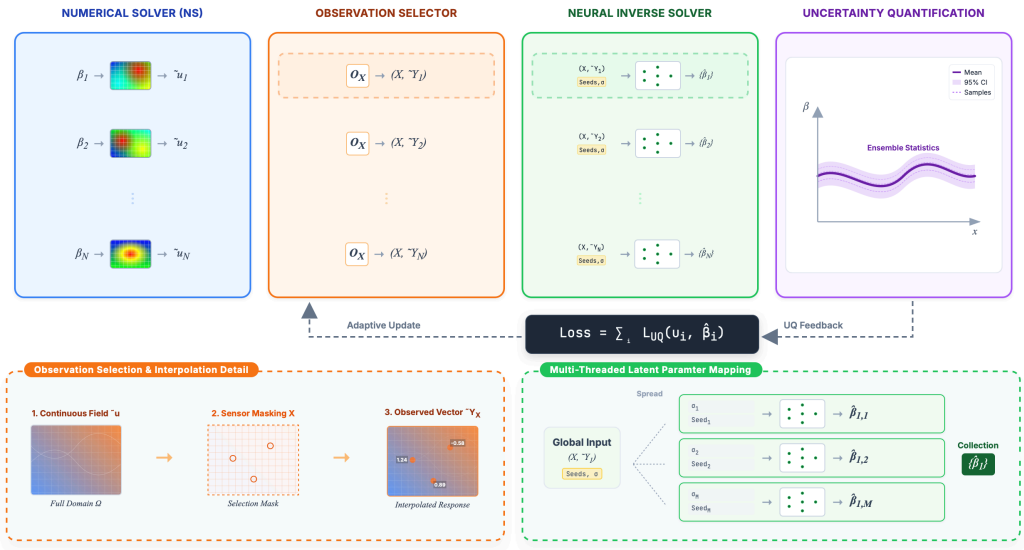

Figure 1: Overview of UNED. The differentiable ED pipeline: (1) the forward solver produces solution trajectories for candidate parameter values; (2) the sensor selector extracts measurements at candidate locations and injects noise realizations; (3) the inverse solver estimates parameters from sparse observations; (4) joint conformal prediction calibrates residual-based UQ bands; (5) the sensor placement optimizer minimizes expected uncertainty via gradient descent, with gradients flowing through the full pipeline. Calibration spans multiple seeds, noise levels, and parameter draws to ensure robust one-shot deployment.

2 Background

2.1 Inverse Problems for Transient PDEs

We focus on systems governed by time-dependent partial differential equations (PDEs) of the form

$$\mathcal{D}[u,\theta](x,t) = f(x,t), \quad (x,t) \in \Omega \times [0,T], \qquad \mathcal{C}[u,\theta](x,t) = g(x,t), \quad (x,t) \in \partial\Omega \times [0,T], \tag{1}$$

where $u : \Omega \times [0,T] \to \mathbb{R}^m$ is the observable spatio-temporal field, $x \in \Omega \subset \mathbb{R}^d$ are spatial coordinates, $t \in [0,T]$ is time, and $\theta \in \mathbb{R}^p$ are the unknown PDE parameters. Here, $\mathcal{D}$ denotes the differential operator (including time derivatives), and $\mathcal{C}$ encodes boundary and initial conditions. Different values of θ yield different transient solutions $u_\theta(x,t)$.

In the inverse setting, $\mathcal{D}, \mathcal{C}, f$, and g are known, while $\theta^\star$ is not directly observable. Instead, we acquire noisy spatio-temporal measurements of the solution:

$$Z = \{(x_j, t_j)\}_{j=1}^M \subset \Omega \times [0,T], \qquad Y = \{u_{\theta^\star}(x_j, t_j) + \xi_j\}_{j=1}^M, \tag{2}$$

with Gaussian noise $\xi_j \sim \mathcal{N}(0, \sigma^2)$. The objective is to recover $\theta^\star$ from (Z, Y).

A standard formulation is to estimate $\hat{\theta}$ by solving

$$\hat{\theta}(Z,Y) \;=\; \arg\min_{\theta \in \mathbb{R}^p} \big\| u_\theta(Z) - Y \big\|_2^2, \tag{3}$$

where $u_\theta(Z) = \big(u_\theta(x_1,t_1), \ldots, u_\theta(x_M,t_M)\big)$. This transient inverse problem formulation underpins both classical PDE-constrained optimization methods and modern neural solvers.

2.2 Uncertainty Quantification via Conformal Prediction

Instead of placing a prior over parameters as in Bayesian UQ, we adopt conformal prediction (CP) Shafer & Vovk (2008), which provides distribution-free uncertainty sets with guaranteed finite-sample coverage. The goal is not only to estimate a point solution $\hat{\theta}$ but to construct a set $\mathcal{C}_\alpha(Z,Y)$ that contains the true parameter $\theta^\star$ with high probability.

Formally, given a nonconformity score function $s(\theta; Z, Y)$ that measures the discrepancy between model predictions $u_\theta(Z)$ and observations Y, conformal prediction constructs the $(1-\alpha)$ confidence set

$$\mathcal{C}_\alpha(Z,Y) \;=\; \big\{\theta \in \mathbb{R}^p \;:\; s(\theta; Z, Y) \le q_{1-\alpha}\big\}, \tag{4}$$

where $q_{1-\alpha}$ is the $(1-\alpha)$ quantile of calibration scores. By construction, CP guarantees $P(\theta^\star \in \mathcal{C}_\alpha(Z,Y)) \ge 1-\alpha$ without distributional assumptions.

2.3 Experimental Design

In many scientific applications, observations Y are costly to obtain due to sensor deployment expenses, operational constraints, or limited experimental access. Moreover, sequential sensor placement is often impractical: relocating instruments across rounds is expensive and disruptive, and reproducing identical experimental conditions multiple times is rarely feasible. Consequently, sensor locations and time points must frequently be chosen in a *one-shot* manner, where all observation inputs are fixed in advance. This setting motivates the need for principled design strategies that maximize information gain under strict resource limitations.

In our setting, measurements are acquired at a selected set of space–time locations measurements are acquired at sensor locations $\boldsymbol{s} = \{x_j\}_{j=1}^{N_s} \subset \Omega$ over time. These sensor locations $\boldsymbol{s}$ are optimized to enable reliable recovery of the unknown PDE parameters. The design objective is to minimize uncertainty in the recovered parameters. Uncertainty is measured by the size of a $(1-\alpha)$ conformal prediction set $\mathcal{C}_\alpha(\boldsymbol{s}, Y) = \{\theta : s(\theta; \boldsymbol{s}, Y) \le q_{1-\alpha}\}$, evaluated via a monotone set functional $\mathcal{M}(\cdot)$ (e.g., diameter or volume). Furthermore, we monitor estimation accuracy, quantified by the expected risk

$$\mathcal{R}(\boldsymbol{s}) \;=\; \mathbb{E}_{\theta,Y}\big[\|\hat{\theta}(\boldsymbol{s}, Y) - \theta\|_2^2\big].$$

We therefore cast experimental design as an optimization problem over feasible sensor locations $\boldsymbol{s} \in \mathcal{S}$:

$$\min_{\boldsymbol{s} \in \mathcal{S}} \; \mathbb{E}_Y\big[\, \mathcal{M}(\mathcal{C}_\alpha(\boldsymbol{s}, Y))\,\big], \tag{5}$$

The set $\mathcal{C}_\alpha(\boldsymbol{s}, Y)$ denotes the $(1-\alpha)$ conformal prediction set measuring uncertainty via a monotone functional $\mathcal{M}(\cdot)$ (e.g., diameter or volume), and $\mathcal{S}$ encodes any physical or operational constraints on sensor placement. This formulation explicitly prioritizes minimizing expected uncertainty in the recovered parameters while guaranteeing accurate parameter estimation.

3 UNED

3.1 ED LOOP

To determine informative sensor configurations, our framework emulates the logic of experimental design (ED) by running multiple forward simulations of the PDE system across a range of reference parameter values $\{\beta_i\}$. These values can be selected based on domain knowledge or prior uncertainty. For each parameter instance β_i, we further spawn subthreads that differ in the injected noise levels and in the random seeds used for neural network initialization. This creates a diverse collection of observational scenarios that more faithfully capture both measurement variability and training stochasticity.

Within each subthread, the pipeline proceeds through three stages: (1) the numerical forward solver (Sec. 3.2) produces an approximate solution trajectory $\hat{u}_{\beta_i}$ on the discretized domain; (2) the sensor selector (Sec. 3.3) extracts measurements at candidate locations $\boldsymbol{s}$ and generates noisy observational datasets; and (3) the inverse solver (Sec. 3.4) estimates the parameters $\hat{\beta}_i$ from these measurements. The quality of a sensor configuration is then assessed by aggregating estimation accuracy and uncertainty across all subthreads.

Unlike conventional ED methods that depend on non-differentiable numerical integration schemes, our approach formulates this entire loop as a differentiable optimization problem. By combining finite-difference solvers with neural inverse models and differentiable objectives, we enable gradient-based updates of sensor placement. This design makes it possible to efficiently search over candidate configurations and identify sensor layouts that balance predictive accuracy, uncertainty control, and spatial coverage.

3.2 NUMERICAL FORWARD SOLVER.

In this framework, we use a classical *finite-difference* (FD) forward solver with the *method of lines* (MOL): we discretize space, then solve the resulting system of ODEs in time.

Spatial FD. On a uniform grid with spacing Δx, we approximate spatial operators by standard FD stencils to obtain a semi-discrete state $u_h(t) \in \mathbb{R}^{n_h}$ and a right-hand side G_h:

$$\dot{u}_h(t) \;=\; G_h\big(u_h(t); \theta, t\big), \qquad u_h(0) = u_{h,0}, \quad \text{(BCs enforced in } G_h\text{)}.$$

Time integration. We advance this ODE with a stable time integrator (e.g., RK/CN/IMEX) on steps of size Δt, yielding the discrete trajectory $\hat{u}_\theta^{\Delta x, \Delta t} = \{\hat{u}_\theta(x_j, t_k)\}$.

$$\theta \;\xrightarrow{F_{\Delta x, \Delta t}}\; \hat{u}_\theta^{\Delta x, \Delta t} \tag{6}$$

3.3 SENSORS SELECTOR

This component takes as input the forward solution u_{target} and the selected sensor positions $\boldsymbol{s} = \{x_j\}_{j=1}^{N_s}$. The ground truth field is first interpolated at these locations to produce the clean sensor measurements

$$Y = \{u_{\text{target}}(x_j)\}_{j=1}^{N_s}.$$

To model measurement noise, we generate K perturbed realizations by injecting independent Gaussian noise at each sensor, i.e.,

$$\widetilde{Y}^{(k)} = \{u_{\text{target}}(x_j) + \varepsilon_j^{(k)}\}_{j=1}^{N_s}, \quad \varepsilon_j^{(k)} \sim \mathcal{N}(0, \sigma^2),$$

for $k = 1, \ldots, K$. This results in K noisy datasets of observations associated with the same sensor placement.

3.4 INVERSE SOLVER

We build on the inverse solver proposed in Li et al. (2025), who introduce NEPTUNE (Neural Estimation of Parameters in mulTi-physics PDEs Under sparse obSErvations).Their approach first solves the forward model with a numerical solver and then employs a neural network to infer the

unknown parameters. This separation distinguishes NEPTUNE from physics-informed neural networks (PINNs), where the network simultaneously learns the forward solution and the parameters. By avoiding this coupling, NEPTUNE mitigates the error accumulation that often limits the reliability of PINNs in estimating parameters for complex PDEs. NEPTUNE follows a two-stage strategy: it first estimates global scalar parameters and then refines local variations through neural networks, which yields robust inference across nonlinear multiphysics systems.

In our framework, we adopt NEPTUNE directly as the inverse solver. The inverse solver is a Multi-Layer Perceptron (MLP) that takes coordinate inputs (x, t) and outputs the parameter field $\theta(x)$. The parameter optimization is performed via the adjoint method (Plessix, 2006), enabling efficient calculation of gradients through the numerical solver. The prediction loss is defined as

$$\mathcal{L}_{\text{pred}} = \sum_{t,x} (u^{(i)}_{\text{pred}}(t, x) - u^{(i)}_{\text{target}}(t, x))^2, \tag{7}$$

which measures the squared error between predicted and target observations across all sensors and time steps.

3.5 UQ Operator

Our formulation builds on the recent framework of Gopakumar et al. Gopakumar et al. (2025), which introduces physics-informed conformal prediction (PRE-CP) for neural PDE solvers. Their method calibrates uncertainty in the *residual space* by using Physics Residual Errors (PRE) as nonconformity scores, thereby achieving distribution-free coverage without access to ground-truth solution trajectories. PRE-CP computes residuals directly from the forward solver's predictions, which measures how well the surrogate respects the governing PDE but does not incorporate observed data.

In contrast, our goal is uncertainty quantification for parameter inference, where ground-truth parameters $\theta^\star$ are unavailable. We therefore adapt the residual-based calibration principle, but redefine the residual to measure the physics-consistency of candidate parameters *relative to sensor observations*. Given observations Y at space–time sample locations $Z = \{(x_j, t_j)\}_{j=1}^{M} \subset \Omega \times [0, T]$, we evaluate the PDE residual restricted to Z as

$$r(x, t; \theta) \;=\; \mathcal{D}[Y, \theta](x, t) - f(x, t), \qquad (x, t) \in Z, \tag{8}$$

so that residuals directly quantify how well candidate parameters θ explain measured dynamics under the PDE constraints.

Uncertainty sources and calibration. To reflect practically relevant uncertainties in our inverse pipeline, the calibration ensemble explicitly spans:

- **Epistemic (model) uncertainty:** variability across independently trained surrogates/inverse solvers (different random seeds), capturing model non-identifiability and optimization stochasticity.
- **Aleatoric (data) uncertainty:** variability induced by different additive noise levels in the response data Y, reflecting irreducible measurement noise.
- **Parametric/structural variability:** variability across plausible PDE parameter settings used to generate or stress-test residuals during calibration (e.g., draws from a feasible set over θ).

Operationally, we aggregate trajectory-level residual scores across seeds, noise levels, and parameter draws to estimate $q_{1-\alpha}$, ensuring that equation 9 is calibrated to joint variability and supports one-shot deployments.

Joint Uncertainty Quantification To provide reliable uncertainty quantification (UQ) across the spatial domain, we adopt a *joint conformal prediction* (joint-CP) strategy (Lei & Wasserman, 2014; Diquigiovanni et al., 2021; 2022). Unlike marginal CP, which calibrates uncertainty pointwise, joint-CP constructs coverage bands for entire spatial fields. Formally, let $\bar{r}(x) = \frac{1}{T}\int_0^T |r(x, t)|dt$ denote the time-averaged absolute residual at location x. The joint-CP bands $\mathcal{C}^\alpha$ satisfy

$$\Pr(\forall x \in \Omega : \bar{r}_{\text{true}}(x) \in \mathcal{C}^\alpha(x)) \geq 1 - \alpha, \tag{9}$$

where coverage is enforced uniformly across the spatial domain Ω.

Nonconformity scores. We define the nonconformity scores by aggregating residuals over the temporal domain and normalizing by a local variability scale. Let $\{r_i\}_{i=1}^N$ denote the residuals from the calibration set. We first compute the time-averaged absolute residuals $\bar{r}_i(x)$. To account for heteroscedasticity, we define a modulation term $m(x)$ estimated as the empirical standard deviation of these aggregated residuals over the calibration set:

$$m(x) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{r}_i(x) - \mu(x)\right)^2}, \tag{10}$$

where $\mu(x)$ is the mean aggregated residual. The normalized joint scores are then defined as the supremum of these errors across the spatial domain:

$$s_i \;=\; \sup_{x\in\Omega} \frac{\bar{r}_i(x)}{m(x)}. \tag{11}$$

These scores are calibrated against a $(1-\alpha)$ empirical quantile: $q_{1-\alpha} = \text{Quantile}_{1-\alpha}\big(\{s_i\}_{i=1}^N\big)$.

Prediction sets. The resulting joint-CP prediction region for the aggregated residuals is expressed as

$$\mathcal{C}^\alpha = \pm\, q_{1-\alpha} \cdot m(\boldsymbol{s}). \tag{12}$$

This construction yields adaptive prediction bands that reflect both the desired global coverage and the local variability of the model's physical consistency.

3.6 Sensor Placement Optimizer

The outer loop optimizes the sensor locations $\boldsymbol{s} \in \mathbb{R}^{N_s}$ by directly minimizing the expected predictive uncertainty under the joint-CP framework. Formally, the optimization problem is defined as:

$$\mathcal{L}_{\text{outer}}(\boldsymbol{s}) = q_{1-\alpha} \cdot m(\boldsymbol{s}), \tag{13}$$

where $q_{1-\alpha}$ is the calibrated $(1-\alpha)$ quantile of the nonconformity scores and $m(\boldsymbol{s})$ is the spatial modulation (residual standard deviation) across the domain.

Differentiable Uncertainty Loss. A primary challenge in this formulation is that the quantile operator $q_{1-\alpha}$ is non-differentiable with respect to its input scores, which typically results in sparse or zero gradients during backpropagation. To ensure effective end-to-end training of the sensor coordinates, we replace the standard quantile with a custom differentiable surrogate $\hat{q}_{1-\alpha}$ that distributes importance-weighted gradients across all calibration samples. This allows the optimizer to reduce the total area of the coverage band by simultaneously minimizing the local variability (m) and the global calibration threshold $(q_{1-\alpha})$. We provide a detailed derivation and discussion of this custom gradient weighting scheme in Appendix B.1.

Regional Placement Constraints To prevent the common issue of sensor clustering, in which multiple sensors aggregate at the same optimal point, we partition the spatial domain Ω into N_s disjoint sub-regions $\mathcal{R}_i$. We then constrain each sensor to find its optimal location within its designated region:

$$\boldsymbol{s}_i \in \mathcal{R}_i \quad \forall i \in \{1, \ldots, N_s\}. \tag{14}$$

This regional strategy ensures that the sensor network maintains comprehensive spatial coverage rather than collapsing onto a single global singleton. Empirically, this approach has been shown to improve results significantly as it forces the optimizer to explore the most informative locations within each spatial segment. Furthermore, it effectively resolves the "stacking" issue, where many sensors are placed nearly on top of one another, a frequent failure mode observed in other baseline sensor placement models.

The regional constraint is implemented through a projection operator $\mathcal{P}_{\mathcal{R}_i}$ that maps any proposed sensor update during gradient descent back to the nearest valid position within its sub-region $\mathcal{R}_i$:

$$\boldsymbol{s}_i^{(k+1)} = \mathcal{P}_{\mathcal{R}_i}\left(\boldsymbol{s}_i^{(k)} - \eta \nabla_{\boldsymbol{s}_i} \mathcal{L}_{\text{outer}}\right) \tag{15}$$

Table 2: Uncertainty quantification (UQ) bounds comparison across different problem classes. Reported values are mean absolute residual bounds (lower values indicate tighter uncertainty bounds). Best performing method in each application is shown in bold.

	Linear PDEs	Real-world datasets
Method	Wave ($\times 10^{-2}$)	Cell Growth ($\times 10^{-5}$)
Grid	8.91	1.08
Random	9.72	1.59
MI	**8.41**	1.044
MINE	9.34	1.27
FIST	9.46	1.36
UNED	**8.41**	**1.03**

where η is the learning rate and k denotes the optimization iteration. This guarantees a well-distributed sensor configuration that maximizes the reliability and physical consistency of the inverse problem solution.

4 RESULTS

We evaluate UNED on benchmark PDE inverse problems and real-world experimental datasets. We compare against baseline sensor placement strategies: *Grid* (uniform grid), *Random* (random placement), *MI*, *MINE*,(based on Frazier (2018)) and *FIST* (PIED-based method Hemachandra et al. (2025)). The primary metric is UQ residual bounds (mean absolute residual, lower is better), computed over 300 test simulations that vary additive noise levels and neural network initialization seeds to assess calibration under epistemic and aleatoric uncertainty. Table 2 reports UQ residual results across problem classes.

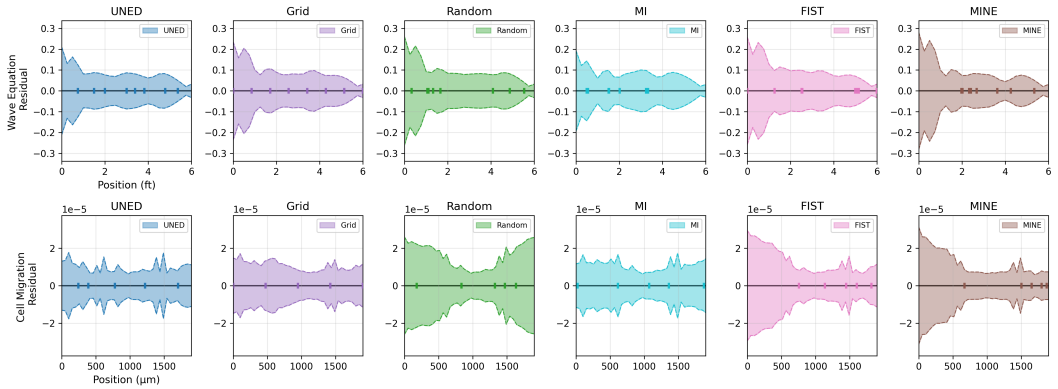

Figure 2: UQ residual for the Wave Equation (top row) and Cell Migration (bottom row). Columns from left to right correspond to UNED, Grid, Random, MI, FIST, and MINE. The plots visualize the calibration of residuals, where tighter bands around zero indicate better uncertainty quantification.

4.1 WAVE EQUATION

We consider a one-dimensional wave equation where the inverse task is to recover the wave velocity field from sparse observations of the solution. We define the propagation speed v as a *field parameter* $v(x)$ to be estimated. Full governing equations and problem setup are provided in Appendix C.

UNED achieves the lowest UQ residual bounds (8.41), tied with MI, among all baselines (Table 2). Parameter UQ and parameter estimation performance are reported in Table 3. The wave equation uses synthetic data with known ground-truth parameters, enabling evaluation of parameter-space

metrics in addition to residual bounds. UNED minimizes residual-space uncertainty (Table 2); it does not directly optimize parameter estimation error or parameter-space UQ. Nevertheless, Table 3 shows that minimizing residual uncertainty yields beneficial downstream effects: both parameter error and parameter-space UQ improve compared to all baselines. This indicates that sensor placements that tighten residual bands also produce more accurate and better-calibrated parameter estimates.

Table 3: Wave equation: parameter UQ and parameter estimation error. UNED optimizes neither metric directly, yet achieves the best performance on both. Param. UQ: Median (Interquartile Range [IQR]: 25th, 75th percentiles). Param. error: median absolute error. Best in bold.

Method	Param. UQ ($\times 10^{-3}$)	Param. Error ($\times 10^{-2}$)
Grid	1.27 (0.33, 5.06)	1.31
Random	1.35 (0.24, 6.71)	1.36
MI	1.27 (0.52,4.46)	1.49
MINE	1.28 (0.42,4.89)	1.40
FIST	1.39 (0.26, 6.11)	1.08
UNED	**0.95 (0.32, 4.29)**	**0.92**

The improvement over Grid, Random, MINE, and FIST demonstrates that optimizing sensor placement to minimize expected posterior uncertainty yields tighter calibrated bands and more accurate parameter recovery. By accounting for model and noise variability during design, UNED identifies sensor configurations that generalize across realizations.

4.2 Real-World Datasets: Cell Migration and Proliferation

We validate UNED on real-world experimental data to assess generalization beyond simulated settings. Cell migration and proliferation describes the diffusion-reaction dynamics of cell density under a Fisher-KPP-type equation, with unknown diffusivity and proliferation parameters estimated from in vitro scratch assay data. For this application, we use synthetic data to train UNED and optimize sensor locations; the experimental data are held out for testing only, ensuring that evaluation reflects performance on real observations unseen during design. The setup is extremely data-scarce (152 observations over 38 spatial points and 4 timesteps). Governing equations, boundary conditions, and detailed problem setup are provided in Appendix D. UQ residual results are reported in Table 2; parameter UQ and parameter estimation metrics are not available for this setting since ground-truth parameters are unknown for experimental data. On this real-world task, UNED attains the lowest UQ bounds (1.03×10^{-5}), while offering a fully differentiable, uncertainty-aware design that generalizes to held-out experimental data without access to ground-truth parameters during training.

5 Conclusion

We have introduced UNED, a differentiable experimental design framework for PDE-constrained inverse problems that optimizes sensor placement to minimize expected posterior uncertainty under one-shot deployment and scarce data constraints. By combining finite-difference forward solvers with neural inverse models and joint conformal prediction, UNED enables gradient-based optimization of sensor locations while explicitly accounting for variability in model initialization and measurement noise. Our experiments on the wave equation and on real-world cell migration data show that UNED achieves the tightest residual-space UQ bounds on the wave equation (tied with MI) and competitive bounds on cell migration (Table 2), alongside baselines Grid, Random, MI, MINE, and FIST. For the wave equation, where ground-truth parameters are available, minimizing residual uncertainty also yields the best parameter estimation accuracy and parameter-space UQ, even though those metrics are not optimized directly. These results indicate that explicitly minimizing uncertainty under model and noise randomness yields robust sensor configurations whose performance generalizes across realizations. We view UNED as a promising step toward practical experimental design for scientific sensing applications where data are expensive to acquire and deployment is one-shot.

REFERENCES

Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. Optimal experimental design for infinite-dimensional bayesian inverse problems with linear pdes. *SIAM Journal on Scientific Computing*, 36(5):A2122–A2148, 2014.

Anas Attia, Sven Leyffer, and Todd Munson. Robust a-optimal design of experiments for bayesian linear inverse problems. *arXiv preprint arXiv:2306.09246*, 2023.

Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and R Devon Hjelm. Mutual information neural estimation. In *International Conference on Machine Learning*, pp. 531–540. PMLR, 2018.

José M Bernardo. Expected information as expected utility. *The Annals of Statistics*, 7(3):686–690, 1979.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

Yuhan Chen, Lu Lu, George Em Karniadakis, and Luca Dal Negro. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 116(44):22445–22451, 2021.

Jacopo Diquigiovanni, Matteo Fontana, and Simone Vantini. The importance of being a band: Finite-sample exact distribution-free prediction sets for functional data. *arXiv preprint arXiv:2102.06746*, 2021.

Jacopo Diquigiovanni, Matteo Fontana, and Simone Vantini. Conformal prediction bands for multivariate functional data. *Journal of Multivariate Analysis*, 189:104879, 2022.

Ronald Aylmer Fisher. The wave of advance of advantageous genes. *Annals of Eugenics*, 7:355–369, 1937.

Adam Foster, Martin Jankowiak, Eli Bingham, Paul Horsfall, Yee Whye Teh, Tom Rainforth, and Noah Goodman. Variational bayesian optimal experimental design. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Nicolas R Franco, Carlos Orozco, and Hector Gómez. Deep learning methods for pdes and related parameter identification problems. *Computers & Mathematics with Applications*, 145:130–147, 2023.

Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

Vignesh Gopakumar, Ander Gray, Lorenzo Zanisi, Timothy Nunn, Daniel Giles, Matt J Kusner, Stanislas Pamela, and Marc Peter Deisenroth. Calibrated physics-informed uncertainty quantification. *arXiv preprint arXiv:2502.04406*, 2025.

Eldad Haber, Jason Magnant, Carlos Lucero, and Luis Tenorio. Sparsity constrained a-optimal design of experiments. *Computational Optimization and Applications*, 52(1):293–314, 2012.

Ayesh Hemachandra, Gary K.R. Lau, Samuel K. Ng, and Boon Khai H. Low. Pied: Physics-informed experimental design for inverse problems. *arXiv preprint arXiv:2503.07070*, 2025.

Wenrui Jin, Eshan T Shah, Catherine J Penington, Scott W McCue, Lisa K Chopin, and Matthew J Simpson. Reproducibility of scratch assays is affected by the initial degree of confluence: experiments, modelling and model selection. *Journal of theoretical biology*, 390:136–145, 2016.

Jari Kaipio and Erkki Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer, 2005.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Noble Kennamer, Martin Jankowiak, Fritz Obermeyer, Eli Bingham, Melanie F Pradier, Brooks Paige, Suchi Saria, Yee Whye Teh, and Tom Rainforth. Amortized variational bounds for bayesian optimal experimental design. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 9267–9275, 2023.

Steven Kleinegesse and Michael U Gutmann. Bayesian experimental design for implicit models by mutual information neural estimation. In *International Conference on Machine Learning*, pp. 5311–5320. PMLR, 2020.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Jing Lei and Larry Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):71–96, 2014.

Xuyang Li, Mahdi Masmoudi, Rami Gharbi, Nizar Lajnef, and Vishnu Naresh Boddeti. Estimating parameter fields in multi-physics pdes from scarce measurements. *arXiv preprint arXiv:2509.00203*, 2025.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Philip K Maini, DLS McElwain, and DI Leavesley. Traveling waves in biology. *Mathematical medicine and biology*, 21(1):1–17, 2004.

Nick McGreivy and Ammar Hakim. Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nature machine intelligence*, 6 (10):1256–1269, 2024.

Ilse Neitzel, Kathrin Pieper, Boris Vexler, and Daniel Walter. Optimal experimental design for inverse problems by objective-based sparse regularization. *Inverse Problems*, 34(6):065001, 2018.

René-Edouard Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, 2006.

Maziar Raissi, Paris Perdikaris, and George E 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.

BG Sengers, CP Please, and ROC Oreffo. Experimental characterisation of hydrogel modulus. *Biotechnology and bioengineering*, 98(1):123–131, 2007.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(12):371–421, 2008.

Andrew M Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

Bao Ngu Vo, Christopher C Drovandi, Anthony N Pettitt, and Matthew J Simpson. Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate bayesian computation. *Mathematical Biosciences*, 263:133, 2015.

Sifan Wang, Hanwen Wang, and Paris Perdikaris. Learning the solution operator of parametric partial differential equations with physics-informed deeponets. *Science advances*, 7(40):eabi8605, 2021.

A NOTATIONS

Table 4 summarizes the main symbols used throughout the paper.

Table 4: List of notations used throughout the paper.

Symbol	Meaning	Example
$\mathcal{D}$	PDE differential operator	equation 1
$\mathcal{C}$	Boundary and initial condition operator	equation 1
Ω	Spatial domain	equation 1
$[0, T]$	Time interval	equation 1
u	Observable spatio-temporal field	equation 1
θ	PDE parameters	equation 1
u_θ	Solution of the PDE with parameter θ	equation 1
Z	Observation locations (space–time)	equation 3
Y	Noisy observations	equation 3
$\hat{\theta}$	Parameter estimate from inverse solver	equation 3
$\theta^\star$	True unknown PDE parameter	—
$\mathcal{C}_\alpha$	Conformal prediction set (coverage $1-\alpha$)	equation 4
$q_{1-\alpha}$	$(1-\alpha)$ quantile of calibration scores	equation 4
$\mathcal{S}$	Set of feasible sensor locations	equation 5
$\mathcal{M}$	Monotone set functional (e.g., diameter, volume)	equation 5
$\mathcal{R}(Z)$	Expected estimation risk	equation 5
β_i	Reference parameter values in ED pipeline	—
$\hat{u}_\theta^{\Delta x, \Delta t}$	Discrete forward solution	equation 6
$\boldsymbol{s}$	Sensor positions	—
$\widetilde{Y}$	Noisy observation realizations	—
r	PDE residual at observation locations	equation 8
$\bar{r}(x)$	Time-averaged absolute residual at x	equation 9
s_i	Nonconformity score for calibration sample i	equation 11
$m(x)$	Spatial modulation (residual std. deviation)	equation 12
$\mathcal{C}^\alpha$	Joint-CP prediction band for residuals	equation 12
$\boldsymbol{s}$	Sensor location vector	equation 13
N_s	Number of sensors	—
$\mathcal{R}_i$	Sub-region for sensor i (placement constraint)	—
$\mathcal{L}_{\text{outer}}$	Outer-loop UQ loss	equation 13
$\mathcal{L}_{\text{pred}}$	Inverse solver prediction loss	equation 7
$\nabla_{\boldsymbol{s}}$	Gradient w.r.t. sensor locations	—

B METHOD DETAILS

B.1 DIFFERENTIABLE QUANTILE AND GRADIENT BALANCING

The joint-CP objective requires minimizing the width of the prediction set $\mathcal{C}^\alpha = \pm q_{1-\alpha} \cdot m$. Applying the product rule with respect to sensor locations $\boldsymbol{s}$, the total gradient is:

$$\nabla_{\boldsymbol{s}}(q_{1-\alpha} \cdot m) = q_{1-\alpha} \cdot \nabla_{\boldsymbol{s}} m + m \cdot \nabla_{\boldsymbol{s}} q_{1-\alpha}. \tag{16}$$

In a standard implementation, the quantile operator $q_{1-\alpha}$ is non-differentiable, causing the second term in equation 16 to vanish. This leads to the approximation:

$$\nabla_{\boldsymbol{s}}(q_{1-\alpha} \cdot m) \approx q_{1-\alpha} \cdot \nabla_{\boldsymbol{s}} m. \tag{17}$$

Because the modulation m represents the standard deviation of residuals, optimizing based solely on this approximation minimizes the average residual variance rather than the actual calibrated coverage width.

To enable the optimizer to actively reduce the coverage threshold $q_{1-\alpha}$, we define a custom Vector-Jacobian Product (VJP) that distributes gradients across all calibration samples s_i satisfying $s_i \leq q_{1-\alpha}$. We first compute raw importance weights w_i^{raw} based on the score's proximity to the threshold:

$$\frac{\partial \hat{q}_{1-\alpha}}{\partial s_i} = \begin{cases} w_i = 0.1 + 0.9 \cdot \frac{s_i - \min(s)}{q_{1-\alpha} - \min(s)}, & \text{if } s_i \leq q_{1-\alpha} \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

Gradient Balancing via Normalization. To ensure that neither force in equation 16 dominates the optimization, we apply a normalization constraint to the weights such that $\sum w_i = 1.0$. This ensures that $\nabla_{\boldsymbol{s}} q_{1-\alpha}$ remains on the same order of magnitude as a standard quantile update. This normalization balances the two competing objectives:

$$\nabla_{\boldsymbol{s}} \mathcal{L}_{\text{outer}} = \underbrace{q_{1-\alpha} \cdot \nabla_{\boldsymbol{s}} m}_{\text{Residual Variance}} + \underbrace{m \cdot \nabla_{\boldsymbol{s}} q_{1-\alpha}}_{\text{Calibration Gain}}. \tag{19}$$

By explicitly calculating $\nabla_{\boldsymbol{s}} q_{1-\alpha}$ through the nonconformity scores, the sensor placement optimizer can identify coordinates that simultaneously improve the model's average physical consistency and its worst-case calibration performance, resulting in significantly tighter and more reliable UQ bands.

B.2 State-dependent parameterization in the inverse solver

In the cell migration setting, the parameters are modeled as density-dependent fields. We therefore adjust the neural inverse solver so that its input is not only the spatial coordinate x but the coupled pair $(x, \rho(x))$—i.e., the coordinate and the observed response (density) at that coordinate. This allows the network to learn inverse parameter fields that vary with the local density. In the outer optimization loop, sensor positions are updated via gradient descent, and the gradient flows through the inverse solver to the x-coordinate component of each input; the corresponding density values are induced by the forward solution at those locations. This end-to-end differentiable construction enables gradient-based optimization of sensor placement for density-dependent parameter inference.

C Wave Equation

Application and Known Physics: For simplicity, we consider the one-dimensional wave equation

$$\frac{\partial^2 u(x,t)}{\partial t^2} = v(x)^2 \frac{\partial^2 u(x,t)}{\partial x^2}, \qquad (x,t) \in [0,L] \times [0,T], \tag{20}$$

where $u(x,t)$ denotes the wavefield and $v(x)$ is the space-varying wave speed. In our inverse setup, the objective is to recover $v(x)$ from partial observations of $u(x,t)$.

Initial and Boundary Conditions: We impose homogeneous Dirichlet boundary conditions

$$u(0,t) = 0, \qquad u(L,t) = 0, \qquad t \in [0,T], \tag{21}$$

and use a single-mode sinusoidal initial displacement together with zero initial velocity,

$$u(x,0) = \sin\Big(\frac{\pi x}{L}\Big), \qquad \frac{\partial u(x,0)}{\partial t} = 0, \qquad x \in [0,L]. \tag{22}$$

Problem Setup and Forward Solver: We take $L = 6$ and $T = 5$ and simulate the forward dynamics on a uniform grid with $N_x = 25$ spatial points and $N_t = 150$ time steps (with $\Delta x = L/(N_x - 1)$ and $\Delta t = T/N_t$) using a standard second-order explicit finite-difference scheme for u_{tt} and u_{xx}, enforcing the Dirichlet boundary values at every time step.

D Cell Migration and Proliferation

Application and Known Physics: This section presents an experimental study of the diffusion (cell migration) and reaction (cell proliferation) process, which is governed by a known form of PDE Fisher (1937); Maini et al. (2004); Chen et al. (2021).

$$\frac{\partial \rho}{\partial t} = \gamma \frac{\partial^2 \rho}{\partial x^2} + \lambda_1 \rho - \lambda_2 \rho^2 \tag{23}$$

where parameters such as cell diffusivity γ and proliferation rates λ_1 and λ_2 define the characteristics of cell migration and proliferation.

Unknown Parameters: These parameters are unknown and require estimation from cell density ρ observations. According to the Porous-Fisher model Sengers et al. (2007); Vo et al. (2015), cell diffusivity γ varies with density, increasing with higher cell density, while proliferation rates λ_1 and

λ_2 are treated as constants. However, to demonstrate UNED's generalizability, all three parameters are modeled as density-dependent fields.

Problem Setup: The experimental data were originally extracted from high-resolution imaging collected in previous research on in vitro cell migration (scratch) assays Jin et al. (2016), and were further analyzed Chen et al. (2021) aimed at discovering the governing equation. The preprocessed data describes the cell density distributions at 38 evenly distributed spatial points in one dimension and only at 5 time steps (0h, 12h, 24h, 36h, 48h). The spatial domain ranges from 0 to 1900 μm. A detailed description of the experiment setup and data preprocessing can be found in the previous research Jin et al. (2016); Chen et al. (2021). This spatial-temporal experimental data is extremely sparse and noisy.

Measurements and train-test split: The experimental data provide cell density at 38 spatial points and 5 time steps (152 observations total over 4 timesteps excluding the initial condition). Given the initial condition at $t = 0$h from the experimental data and Neumann boundary conditions $\partial\rho(x = 0)/\partial x = 0$ and $\partial\rho(x = 1900)/\partial x = 0$, we use *synthetic* data (generated from the forward solver) to train UNED and find optimal sensor locations; the experimental observations are reserved exclusively for testing.

E On mapping residual bounds to parameter space.

For PDEs where parameters enter linearly or multiplicatively, residual inequalities (e.g., $|u_t - \theta u_{xx}| \leq q_{1-\alpha}$) can *in principle* be projected into interval/region constraints on θ. In practice, however, numerical differentiation amplifies noise, near-degenerate denominators cause ill-conditioning, and multi-parameter models yield complex feasible geometries that can undermine finite-sample validity. Accordingly, we anchor guarantees in residual space via equation 9, and use the resulting calibrated sets to inform parameter uncertainty in the inverse problem.

F Limitations

We acknowledge two limitations that constrain the scope of our results. First, UNED relies on a differentiable forward solver; gradient-based sensor optimization requires that the map $\theta \mapsto \hat{u}_\theta^{\Delta x, \Delta t}$ be differentiable with respect to both θ and the observation locations. While finite-difference methods with the method of lines satisfy this for many PDEs, problems such as the Navier–Stokes equations with turbulence or strongly nonlinear convection-dominated flows may lack stable, accurate differentiable numerical solvers. In such settings, differentiable surrogate models (e.g., neural operators or learned solvers) could be a solution to retain end-to-end optimization, though this would introduce additional approximation error and training overhead.

Second, our empirical validation is currently confined to linear and moderately nonlinear PDE benchmarks together with real-world datasets (cell migration); comprehensive evaluation on strongly coupled systems (e.g., thermal runaway, porous media flow) remains to be completed. Although the framework is designed for coupled settings, generalization to stiff or multiscale systems warrants further study.