CARE-PDE: ROBUST PDE DISCOVERY UNDER NOISY DATA–EQUATION CORRESPONDENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Data-driven partial differential equation (PDE) discovery aims to recover explicit governing equations from spatiotemporal observations. Existing methods are primarily limited to clean data or data corrupted by simple, single-source noise. In practice, however, multiple noise sources can produce heterogeneous contamination, causing some observations to deviate substantially from the underlying dynamics and weakening their correspondence with the governing equation. We formulate this problem as Noisy Data–Equation Correspondence (NDEC) and propose the Correspondence-Aware Robust Equation Discovery for PDEs (CARE-PDE) framework. Specifically, we develop Reliability-Aware Field Reconstruction (RAFR) by combining a novel Cauchy loss with adaptive reliability estimation to reduce the influence of heavily corrupted observations. Building on these estimations, Reliability-Guided Local Legendre Projection (RGLP) further exploits the estimated reliability to construct reliable local supports and recover stable derivatives through orthogonal projection. Local Integral Constraint Aggregation (LICA) is then presented to aggregate the reconstructed strong-form relations into local integral constraints, reducing sensitivity to pointwise estimation errors. Finally, Stability-Selected Sparse Regression (SSR) identifies consistently supported terms to recover sparse governing equations. Experiments on 12 PDE systems across different contamination scenarios show that CARE-PDE recovers the correct equation support in all 36 tasks and achieves lower true-term coefficient errors than the ten state-of-the-art baselines, demonstrating its effectiveness and robustness under the tested NDEC conditions.

1 INTRODUCTION

Partial differential equations (PDEs) are fundamental mathematical tools for describing the spatiotemporal evolution of natural phenomena and engineering systems (Evans, 2022), with applications in various fields, such as fluid dynamics (Xu et al., 2020; Ansari et al., 2022; Ma et al., 2024), materials science (Chen et al., 2022), biological systems (Abreu et al., 2019; Amor et al., 2020), and complex engineering applications (Karniadakis et al., 2021). However, the governing equations of many multiscale and nonlinear systems remain unknown or only partially understood (Liu et al., 2024; Champion et al., 2019). Therefore, recovering these equations from available observations can provide both interpretable models of complex dynamics and insight into the physical mechanisms that produce them.

Governing PDEs are traditionally derived from first principles, domain knowledge, and physical assumptions (Brunton et al., 2016; Rudy et al., 2017). In recent years, advances in high-fidelity numerical simulations and the increasing availability of spatiotemporal measurements have established data-driven equation discovery as a complementary modeling approach (Zanna & Bolton, 2020; Beetham et al., 2021; Xu & Zhang, 2021). By recovering explicit mathematical expressions from observations (Chen et al., 2021; Song et al., 2024), these methods reduce reliance on prior knowledge and manual derivations while providing interpretable descriptions of the physical mechanisms underlying complex systems (Lu et al., 2022). However, their applicability remains largely limited to relatively simple settings with clean observations, whereas real-world measurements are inevitably contaminated by noise. Noise degrades estimates of state variables and their derivatives, and makes genuine dynamical terms harder to distinguish from spurious terms (Raissi et al., 2019;

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Wang et al., 2021). Therefore, developing robust and accurate PDE discovery methods under complex noise conditions remains a key challenge for advancing data-driven equation discovery toward practical applications.

To address this challenge, existing studies have introduced various robust strategies, including smoothing-based reconstruction (Rudy et al., 2017; Both et al., 2021), weak-form learning (Messenger & Bortz, 2021; Kaptanoglu et al., 2021; Stephany & Earls, 2024), and integral-form modeling (Reinbold et al., 2020; Xu et al., 2021). These strategies alleviate the instability of equation identification caused by noisy derivative estimation. However, they mainly address single-source noise with similar statistical properties across observations (Chen et al., 2021; Stephany & Earls, 2024; Both et al., 2021; Gao et al., 2026; Xu et al., 2025; Chen & Xiao, 2026), leaving sample-level variation in observation quality insufficiently addressed. In practice, sensor failures, environmental disturbances, or data transmission errors may severely corrupt only a part of the observations (Song et al., 2022; Han et al., 2018), resulting in heterogeneous contamination caused by multi-source noise. The resulting sample-level heterogeneity can distort the dynamical information inferred from affected observations. We refer to this phenomenon as Noisy Data–Equation Correspondence (NDEC), which weakens the correspondence between individual observations and a shared underlying governing equation, leading to incorrect supervision and performance degradation.

Unlike conventional settings with single-source noise across samples, NDEC introduces pronounced differences in observation reliability and can render severely corrupted observations inconsistent with the underlying governing equation. The resulting data-equation mismatch poses challenges beyond conventional noise robustness, even though the physical law governing the system remains unchanged. To be specific, most existing methods jointly use all observations for equation-structure search and coefficient estimation, implicitly assuming comparable reliability across samples (Rudy et al., 2017; Both et al., 2021; Messenger & Bortz, 2021). Under NDEC, data-equation mismatch from severely corrupted samples can propagate into candidate-term selection and coefficient estimation, hindering recovery of the true equation structure. Therefore, addressing NDEC requires estimating observation-level reliability during PDE discovery and using reliable information to establish a stable data-equation correspondence.

To address the above challenges, we propose the Correspondence-Aware Robust Equation Discovery for PDEs (CARE-PDE) framework, which propagates observation-level reliability from field reconstruction to equation identification to improve robustness against NDEC, as illustrated in Figure 1. CARE-PDE consists of three complementary modules: Reliability-Aware Field Reconstruction (RAFR), Reliability-Guided Local Legendre Projection (RGLP), and Local Integral Constraint Aggregation (LICA). Specifically, RAFR couples Cauchy Robust Field Reconstruction (CRFR) with Adaptive GMM Reliability Estimation (AGRE) to reduce the influence of heavily contaminated samples on the learned physical field. RGLP then combines orthogonal basis expansion with numerical integration within reliability-guided local windows. This regularizes the field before differentiation, reducing sensitivity to reconstruction errors that would otherwise be amplified by direct high-order automatic differentiation under noise, thereby improving the stability of derivative estimation. LICA integrates the resulting strong-form quantities into local regression constraints, pooling dynamical information to improve consistency between candidate terms and the underlying evolution dynamics, and reducing sensitivity to pointwise perturbations. Finally, SSR evaluates term stability across constraint subsets at multiple surrogate checkpoints and recovers a sparse equation that balances fit and complexity, further improving the reliability of equation identification. In summary, the major contributions of this paper are as follows:

- We formulate the NDEC problem as a PDE discovery setting, where heterogeneous noise contamination weakens the correspondence between observations and a shared governing equation, impairing PDE structure identification. To address NDEC, our CARE-PDE propagates observation-level reliability estimates through physical field reconstruction, candidate-term construction, and equation identification, providing a unified mechanism for limiting the influence of corrupted samples throughout the discovery process.

- We develop RGLP, which couples reliability-adaptive local support selection with truncated Legendre orthogonal projection. By adapting support regions to observation quality and regularizing the field before differentiation, RGLP recovers stable differential representations and improves candidate-term construction under severe noise.

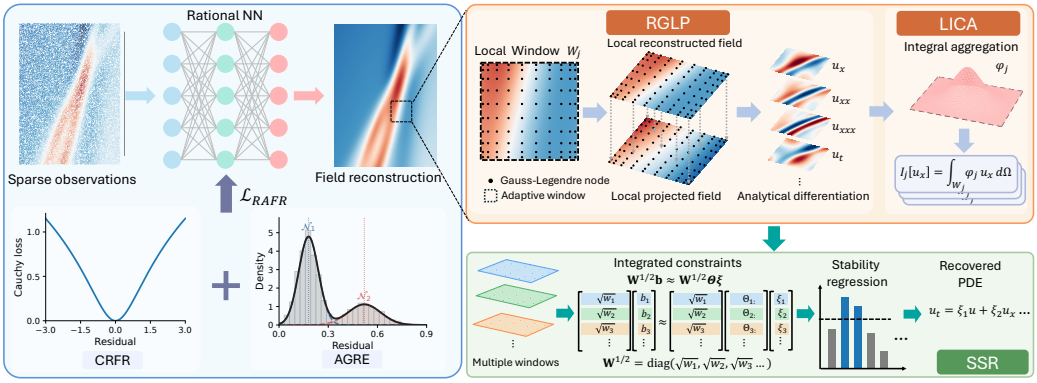

Figure 1: The framework of CARE-PDE. Given sparse noisy observations, RAFR reconstructs a continuous physical field while estimating observation reliability to reduce the influence of highly perturbed samples. The recovered field and reliability information are then used to construct adaptive local windows, within which RGLP performs local Legendre projection and analytically recovers the required spatial and temporal derivatives. Based on these local differential quantities, LICA aggregates pointwise strong-form relations into integral constraints, providing more stable equation information for subsequent identification. SSR finally uses the resulting reliability-weighted constraints to select stable sparse structures and recover the governing PDE.

- We present LICA, which aggregates strong-form dynamical information from multiple spatiotemporal locations into local integral constraints. This design preserves the differential structure of candidate terms while reducing the influence of pointwise estimation errors, yielding more stable PDE constraints for equation identification.

- We conduct extensive experiments on 12 PDE systems under various noise conditions and compare the CARE-PDE framework with ten state-of-the-art PDE discovery baselines. CARE-PDE achieves superior performance in equation structure identification, coefficient recovery, and robustness to noise, recovering the correct equation support in all 36 tasks. The results demonstrate its effectiveness under the evaluated NDEC settings.

2 Related Work

Sparse identification has established an important paradigm for interpretable equation discovery. SINDy (Brunton et al., 2016) introduced sparse representations of dynamical systems, and PDE-FIND (Rudy et al., 2017) extended this idea to PDEs by constructing candidate libraries from state variables and their derivatives. However, direct derivative estimation is sensitive to noise and generally requires sufficiently resolved observations. WSINDy (Messenger & Bortz, 2021) alleviated this issue by replacing pointwise relations with weak-form integral constraints, although such methods still rely on adequate local measurements to construct reliable constraints. Therefore, neural surrogate models provide an attractive alternative for sparse and noisy observations by reconstructing continuous physical fields before equation identification. Representative methods such as DeepMoD (Both et al., 2021) and PINN-SR (Chen et al., 2021) coupled neural field learning with sparse equation discovery, while Weak-PDE-LEARN (Stephany & Earls, 2024) further incorporated weak-form constraints to reduce dependence on high-order pointwise derivatives. More recent approaches have expanded the search for equations beyond fixed candidate libraries, including EqGPT (Xu et al., 2025) and symmetry-guided symbolic regression (Chen & Xiao, 2026). Despite these advances, sample-level observation reliability has rarely been systematically incorporated throughout the PDE discovery pipeline. CARE-PDE addresses this issue by explicitly estimating observation reliability and propagating it through field reconstruction, differential estimation, integral constraint construction, and sparse identification.

3 METHOD

3.1 PROBLEM DEFINITION

We consider a dynamical system described by a state field $\mathbf{u}(\mathbf{x}, t) \in \mathbb{R}^M$, where $\mathbf{x} \in \Omega_{\mathbf{x}} \subset \mathbb{R}^d$ denotes the spatial coordinate and $t \in (0, T)$ denotes time. Its governing equation is represented as

$$\partial_t \mathbf{u} = \boldsymbol{\Xi}^\top \boldsymbol{\psi}[\mathbf{u}], \tag{1}$$

where $\boldsymbol{\psi}[\mathbf{u}] = [\psi_1[\mathbf{u}], \ldots, \psi_F[\mathbf{u}]]^\top$ collects F candidate terms and $\boldsymbol{\Xi} \in \mathbb{R}^{F \times M}$ is a sparse coefficient matrix. The nonzero entries of $\boldsymbol{\Xi}$ determine the active terms of the governing equations, while their values specify the corresponding coefficients. PDE discovery aims to recover the sparse coefficient matrix $\boldsymbol{\Xi}$ from noisy observations $\mathcal{D} = \{(\mathbf{x}_i, t_i, \mathbf{y}_i)\}_{i=1}^N$ with $\mathbf{y}_i = \mathbf{u}(\mathbf{x}_i, t_i) + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\varepsilon}_i \in \mathbb{R}^M$ denotes observation error.

3.2 RELIABILITY-AWARE FIELD RECONSTRUCTION

Given the noisy observations, we approximate the underlying continuous state field with a neural network $\mathbf{f}_{\boldsymbol{\theta}} : \Omega_{\mathbf{x}} \times (0, T) \to \mathbb{R}^M$, which maps spatiotemporal coordinates to the corresponding state variables. For observation i, the reconstruction residual is defined as

$$\delta_i^2 = \frac{1}{M} \left\| \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_i) - \mathbf{y}_i \right\|_2^2 \qquad \mathbf{z}_i = (\mathbf{x}_i, t_i). \tag{2}$$

To improve robustness against large observation errors, we introduce Cauchy-Robust Field Reconstruction (CRFR) based on the Cauchy loss from robust statistical estimation:

$$\mathcal{L}_{\mathrm{CRFR}} = \frac{1}{N} \sum_{i=1}^{N} \frac{c^2}{2} \log\left(1 + \frac{\delta_i^2}{c^2 \sigma_{\mathrm{rob}}^2}\right), \tag{3}$$

where $c > 0$ controls the robustness of the loss, while $\sigma_{\mathrm{rob}}^2 = \max\{\mathrm{median}_i\, \delta_i^2, \sigma_{\min}^2\}$ is an adaptive residual scale, with $\sigma_{\min} > 0$ preventing numerical degeneracy.

Although CRFR reduces the influence of large observation errors during field reconstruction, it does not explicitly quantify observation reliability. We therefore introduce Adaptive GMM Reliability Estimation (AGRE) to estimate reliability from reconstruction residuals. Specifically, we model the residual distribution using one- and two-component Gaussian mixture models (GMMs) (Peel & MacLahlan, 2000) and adaptively select the number of components $G^* \in \{1, 2\}$ according to the Bayesian information criterion (BIC) and the degree of modal separation. The two-component model is selected only when it provides both a better fit and clear residual separation. Otherwise, the one-component model is retained.

When $G^* = 2$, the component with smaller reconstruction residuals is regarded as the high-reliability component, and the posterior probability of observation i belonging to this component is used as its reliability score $\rho_i \in [0, 1]$. When $G^* = 1$, no distinct reliability pattern is identified, and all observations are assigned $\rho_i = 1$.

The estimated reliability is incorporated into field reconstruction through an observation weight:

$$a_i = a_{\min} + (1 - a_{\min})\rho_i^2, \tag{4}$$

where $a_{\min} > 0$ ensures a positive contribution from every observation, while the quadratic mapping increases the contrast between observations with different reliability levels. Field reconstruction then continues with the reliability-weighted Cauchy objective

$$\mathcal{L}_{\mathrm{RAFR}} = \frac{\sum_{i=1}^{N} a_i \frac{c^2}{2} \log\left(1 + \frac{\delta_i^2}{c^2 \sigma_{\mathrm{rob}}^2}\right)}{\sum_{i=1}^{N} a_i}. \tag{5}$$

This objective combines robust field reconstruction with observation-specific reliability weighting, further reducing the influence of highly perturbed observations while retaining contributions from all observations.

3.3 Reliability-Guided Local Legendre Projection

Direct differentiation of the reconstructed field can amplify small-scale reconstruction errors, particularly for high-order derivatives. A truncated Legendre projection instead provides a finite-dimensional local approximation that suppresses rapidly varying error components before differentiation (Shen et al., 2011). Inspired by this, we incorporate local Legendre projection into PDE discovery and propose Reliability-Guided Local Legendre Projection (RGLP), which constructs reliability-adaptive local supports and performs Legendre projection before differentiation. Specifically, for each local center $\mathbf{c}_j$, RGLP constructs a spatiotemporal window:

$$W_j = \{\mathbf{z} : |z_r - c_{jr}| \le h_{jr},\ r = 1, \dots, D\}, \qquad \boldsymbol{\eta} = \frac{\mathbf{z} - \mathbf{c}_j}{\mathbf{h}_j} \in [-1, 1]^D, \tag{6}$$

where $D = d + 1$, $\mathbf{h}_j = [h_{j1}, \dots, h_{jD}]^\top$ contains the half-widths along the spatiotemporal dimensions, and $\boldsymbol{\eta}$ denotes the corresponding normalized local coordinate. A fixed projection window may contain insufficient reliable information in regions with poor observation quality, while an unnecessarily broad support can weaken locality in well-supported regions. Therefore, RGLP adapts the projection support according to the reliability distribution. For each window, we define the effective reliable support as

$$n_j^{\text{eff}} = \sum_{i:\mathbf{z}_i \in W_j} \rho_i. \tag{7}$$

If n_j^{eff} is below a prescribed target n_{tar}, the window is progressively enlarged until sufficient reliable support is obtained or the maximum allowable width is reached. If the target still cannot be satisfied, the current center is discarded, and a new center is sampled. In this way, RGLP uses broader support in regions with limited reliable observations while preserving compact local neighborhoods where reliable information is sufficient.

After determining the adaptive support W_j, we map it to the reference domain $[-1, 1]^D$ and place Q tensor-product Gauss–Legendre quadrature nodes $\{\boldsymbol{\eta}^{(q)}\}_{q=1}^{Q}$. We evaluate the reconstructed field at these nodes and project the local field onto a truncated tensor-product Legendre basis. From the resulting polynomial representation, we recover the required spatial and temporal derivatives through analytical differentiation and coordinate scaling.

3.4 Local Integral Constraint Aggregation

Although RGLP provides regularized derivative estimates, pointwise PDE constraints treat each spatiotemporal location as an independent regression relation, so errors in an individual state or derivative estimate can directly affect equation identification. To reduce this pointwise sensitivity, we introduce Local Integral Constraint Aggregation (LICA), which aggregates the pointwise strong-form relations within each accepted RGLP window into a single integral constraint. Specifically, for each accepted window W_j, we define the local integral operator

$$\mathcal{I}_j[\chi] = \int_{W_j} \varphi_j(\mathbf{z}) \chi(\mathbf{z})\, \mathrm{d}\mathbf{z}, \tag{8}$$

where χ denotes a pointwise strong-form quantity and φ_j is a nonnegative test function supported on W_j. The value of $\varphi_j(\mathbf{z})$ controls the relative contribution of each location within the window to the aggregated constraint. We evaluate the integral using the tensor-product Gauss–Legendre quadrature introduced in RGLP. For the qth quadrature node, we define the corresponding physical location as $\mathbf{z}^{(j,q)} = \mathbf{c}_j + \mathbf{h}_j \odot \boldsymbol{\eta}^{(q)}$. The integral operator is then approximated by

$$\mathcal{I}_j[\chi] \approx \left(\prod_{r=1}^{D} h_{jr} \right) \sum_{q=1}^{Q} \omega_{\text{GL}}^{(q)} \varphi_j\left(\mathbf{z}^{(j,q)}\right) \chi\left(\mathbf{z}^{(j,q)}\right), \tag{9}$$

where $\omega_{\text{GL}}^{(q)}$ denotes the tensor-product Gauss–Legendre quadrature weight and $\prod_{r=1}^{D} h_{jr}$ is the Jacobian determinant associated with the mapping from the reference domain to the physical window. For state component m, we apply $\mathcal{I}_j$ directly to the temporal derivative and candidate terms reconstructed by RGLP:

$$\mathcal{I}_j\left[\widehat{\partial_t u_m}^{(j)}\right] \approx \sum_{\ell=1}^{F} \xi_{\ell m} \mathcal{I}_j\left[\widehat{\psi}_\ell^{(j)}\right], \tag{10}$$

where $\widehat{\partial_t u_m}^{(j)}$ denotes the temporal derivative estimated by RGLP within W_j, and $\widehat{\psi}_\ell^{(j)}$ denotes the ℓth candidate term evaluated from the reconstructed field and the corresponding RGLP derivative estimates. We define

$$b_{jm} = \mathcal{I}_j\left[\widehat{\partial_t u_m}^{(j)}\right], \qquad \Theta_{j\ell} = \mathcal{I}_j\left[\widehat{\psi}_\ell^{(j)}\right]. \tag{11}$$

Each accepted window consequently contributes one integrated PDE constraint rather than multiple independent pointwise relations. Stacking the constraints from all J accepted windows yields

$$\mathbf{b}_m \approx \boldsymbol{\Theta}\boldsymbol{\xi}_m, \qquad \boldsymbol{\Theta} \in \mathbb{R}^{J\times F}, \quad \mathbf{b}_m \in \mathbb{R}^J, \quad \boldsymbol{\xi}_m = \boldsymbol{\Xi}_{:,m}. \tag{12}$$

The reliability of different integrated constraints can still vary because their underlying local supports contain observations with different reliability levels. We therefore compute a test-function-weighted average reliability for each window:

$$\overline{\rho}_j = \frac{\sum_{i:\mathbf{z}_i\in W_j} \varphi_j(\mathbf{z}_i)\rho_i}{\sum_{i:\mathbf{z}_i\in W_j} \varphi_j(\mathbf{z}_i) + \epsilon}, \tag{13}$$

where $\epsilon > 0$ prevents numerical degeneracy. When a window contains limited effective reliable support, its local field and derivative estimates can remain sensitive to individual observations. We therefore define the final constraint weight by combining $\overline{\rho}_j$ with the effective support n_j^{eff} introduced in RGLP:

$$w_j = \overline{\rho}_j\left[1 - \exp\left(-\frac{n_j^{\mathrm{eff}}}{n_0}\right)\right], \tag{14}$$

where $n_0 > 0$ controls how rapidly the support contribution saturates.

Unlike weak-form approaches that transfer derivatives from the state variables to test functions through integration by parts, LICA directly integrates the strong-form quantities recovered by RGLP. This design preserves the original semantics of the candidate library while reducing the sensitivity of equation identification to pointwise estimation errors.

3.5 Stability-Selected Sparse Regression

Sparse regression may select different candidate terms as the available constraints vary. To identify equation structures consistently supported by the data, we introduce Stability-Selected Sparse Regression (SSR), which assesses term stability by repeatedly subsampling the integrated PDE constraints. Specifically, for each state equation, we uniformly sample B fixed-size subsets from the J constraints without replacement and apply reliability-weighted STRidge using the corresponding weights w_j. Let $\widehat{\xi}_{m\ell}^{(b)}$ denote the coefficient of candidate term ℓ in equation m obtained from subset b. Its inclusion frequency and the resulting stable support are defined as

$$\widehat{\Pi}_{m\ell} = \frac{1}{B}\sum_{b=1}^{B}\mathbf{1}\left[\widehat{\xi}_{m\ell}^{(b)} \neq 0\right], \quad \mathcal{S}_m^{\mathrm{stable}} = \left\{\ell : \widehat{\Pi}_{m\ell} \geq \pi_{\mathrm{thr}}\right\}, \tag{15}$$

where $\pi_{\mathrm{thr}} \in (0, 1]$ is the inclusion threshold. The library is then restricted to $\mathcal{S}_m^{\mathrm{stable}}$, and reliability-weighted STRidge is refitted using all J constraints to estimate the coefficients.

Previous studies have indicated that the surrogate checkpoint with the lowest field-reconstruction error is not necessarily the most suitable for PDE discovery (Both et al., 2021). We therefore retain the H surrogate checkpoints with the lowest reconstruction errors and repeat the above stability-selection procedure for each checkpoint. For each STRidge regression, the sparsity threshold τ is selected from a predefined candidate set $\mathcal{T}$. The optimal checkpoint and sparsity threshold are determined using the following unsupervised score:

$$\mathcal{Q}_{c,\tau}^{\mathrm{PDE}} = \frac{1}{M}\sum_{m=1}^{M}\left[\log(1 + R_{cm}(\tau)) + \alpha\frac{\log\left(1 + \|\widehat{\boldsymbol{\xi}}_{cm}(\tau)\|_0\right)}{\log(1+F)}\right], \tag{16}$$

where $R_{cm}(\tau)$ is the reliability-weighted normalized regression residual, $\|\widehat{\boldsymbol{\xi}}_{cm}(\tau)\|_0$ denotes the number of active terms, and $\alpha \geq 0$ balances regression fit and equation sparsity.

Table 1: True-term coefficient error E_{true} (%) and support F1 on the first six PDE systems. DIV. denotes an extremely large but finite coefficient error.

Method	Advection		Conv.–diff.		Burgers		KdV		KS		Allen–Cahn	
	$E_{\text{true}}\downarrow$	F1↑	$E_{\text{true}}\downarrow$	F1↑	$E_{\text{true}}\downarrow$	F1↑	$E_{\text{true}}\downarrow$	F1↑	$E_{\text{true}}\downarrow$	F1↑	$E_{\text{true}}\downarrow$	F1↑
$\beta = 0\%$												
PINN-SR (Nat. Commun., 2021)	100.00	0.000	26.020	0.333	5.8027	1.000	132.42	0.250	100.00	0.000	991.29	0.222
DeepMoD (JCP, 2021)	100.00	0.000	0.6588	1.000	0.5778	1.000	343.15	0.333	**21.140**	**0.857**	3780.1	0.500
R-DLGA (Phys. Rev. Res., 2021)	0.0504	1.000	**0.1783**	**1.000**	0.0938	0.571	1.5828	1.000	77.479	0.571	7.0049	1.000
PDE-READ (Neural Netw., 2022)	DIV.	0.200	8.9796	1.000	12.061	1.000	60.381	1.000	96.976	0.500	30.180	1.000
PIC (Research, 2023)	0.9878	1.000	2.8825	1.000	1.6464	1.000	0.5385	1.000	100.00	0.000	6.4541	1.000
R-DISCOVER (Phys. Fluids, 2024)	0.0639	1.000	1.2278	1.000	0.9165	1.000	53.470	0.364	93.734	0.333	7.6550	1.000
Weak-PDE-LEARN (JCP, 2024)	100.00	0.000	0.3872	1.000	0.3008	1.000	43.574	0.800	115.32	0.400	176.55	0.571
EqGPT (Nat. Commun., 2025)	**0.0332**	**1.000**	10.094	1.000	0.3491	1.000	1.4411	1.000	100.00	0.000	**0.7991**	**1.000**
ABL-PDE (NIPS, 2025)	0.4468	1.000	0.5899	1.000	1.8810	0.800	**0.3462**	**1.000**	89.729	0.286	3.9697	1.000
SI-SR (Phys. Fluids, 2026)	0.1720	1.000	1.3443	1.000	**0.2224**	**1.000**	100.00	0.000	100.00	0.000	100.00	0.000
CARE-PDE	**0.0037**	**1.000**	**0.0331**	**1.000**	**0.0514**	**1.000**	**0.0872**	**1.000**	**2.3265**	**1.000**	**0.0311**	**1.000**
$\beta = 25\%$												
PINN-SR (Nat. Commun., 2021)	100.00	0.000	21.351	0.400	2.3140	1.000	100.00	0.000	98.534	0.500	279.24	0.154
DeepMoD (JCP, 2021)	100.00	0.000	**0.3268**	**1.000**	2.9042	0.800	555.08	0.400	96.456	0.182	7219.6	0.500
R-DLGA (Phys. Rev. Res., 2021)	**0.0054**	**1.000**	0.5129	1.000	2.6219	0.500	27.294	1.000	90.141	0.400	80.733	0.800
PDE-READ (Neural Netw., 2022)	100.00	0.000	8.7915	0.571	14.811	1.000	60.393	1.000	**96.778**	**0.500**	28.260	1.000
PIC (Research, 2023)	0.1102	1.000	2.3061	1.000	2.8347	1.000	**0.4339**	**1.000**	100.00	0.000	**4.5413**	**1.000**
R-DISCOVER (Phys. Fluids, 2024)	0.0800	1.000	2.0889	1.000	1.4933	1.000	93.811	0.222	100.00	0.000	10.895	1.000
Weak-PDE-LEARN (JCP, 2024)	100.00	0.000	0.5806	1.000	1.0918	1.000	54.645	0.667	99.000	0.400	134.84	0.500
EqGPT (Nat. Commun., 2025)	0.3349	1.000	0.8832	1.000	**0.0777**	**1.000**	0.8927	1.000	100.00	0.000	100.00	0.000
ABL-PDE (NIPS, 2025)	0.1198	1.000	0.6376	1.000	2.6956	0.800	0.9954	1.000	97.454	0.500	10.891	1.000
SI-SR (Phys. Fluids, 2026)	0.2539	1.000	0.8092	1.000	5.0745	1.000	100.00	0.000	100.00	0.000	100.00	0.000
CARE-PDE	**0.0038**	**1.000**	**0.1165**	**1.000**	**0.0647**	**1.000**	**0.0897**	**1.000**	**1.0132**	**1.000**	**0.0185**	**1.000**
$\beta = 50\%$												
PINN-SR (Nat. Commun., 2021)	100.00	0.000	24.579	0.333	1.6205	1.000	100.00	0.000	100.00	0.000	100.00	0.000
DeepMoD (JCP, 2021)	100.00	0.000	**0.2084**	**1.000**	2.5242	0.800	100.00	0.000	94.699	0.308	9622.2	0.500
R-DLGA (Phys. Rev. Res., 2021)	0.0769	1.000	0.3427	1.000	54.422	0.250	3.0872	0.667	97.098	0.400	80.823	0.800
PDE-READ (Neural Netw., 2022)	DIV.	0.200	9.3903	1.000	15.947	1.000	59.054	1.000	**97.144**	**0.500**	31.446	1.000
PIC (Research, 2023)	0.0640	1.000	2.6201	1.000	3.0814	1.000	0.4834	1.000	100.00	0.000	**4.6892**	**1.000**
R-DISCOVER (Phys. Fluids, 2024)	**0.0263**	**1.000**	2.0163	0.800	1.3606	1.000	100.00	0.000	100.00	0.000	13.705	1.000
Weak-PDE-LEARN (JCP, 2024)	100.00	0.000	0.7608	1.000	1.9214	1.000	50.769	0.667	128.13	0.400	134.33	0.500
EqGPT (Nat. Commun., 2025)	0.2323	1.000	1.1017	1.000	**0.8731**	**1.000**	**0.4536**	**1.000**	100.00	0.000	123.57	0.750
ABL-PDE (NIPS, 2025)	0.1532	1.000	0.6071	1.000	0.9613	0.667	4.3523	0.571	97.520	0.500	9.9825	1.000
SI-SR (Phys. Fluids, 2026)	0.2489	1.000	0.2616	1.000	100.00	0.000	13.124	1.000	100.00	0.000	100.00	0.000
CARE-PDE	**0.0058**	**1.000**	**0.1734**	**1.000**	**0.0304**	**1.000**	**0.2176**	**1.000**	**7.3953**	**1.000**	**0.0312**	**1.000**

4 EXPERIMENTS

4.1 PDE SYSTEMS

We comprehensively evaluate CARE-PDE on 12 representative PDE systems: Advection, Conv.–diff., Burgers, KdV, KS, Allen–Cahn, Chaffee–Infante, Compound, Burgers2D, Adv.–diff.(2D), NLS, and λ–ω. Detailed information on these systems is provided in the Appendix E.1. In real-world applications, spatiotemporal observations are often sparse due to limitations in sensor deployment and data acquisition. Therefore, we randomly sample 20,000 spatiotemporal locations from the complete field of each PDE system, split them into training and validation sets in an 8:2 ratio, and introduce sample-level heterogeneous contamination to the training data. Specifically, the κth state component of training sample n is perturbed as

$$\widetilde{u}_{n,\kappa} = u_{n,\kappa} + \gamma_n \varsigma_\kappa \varepsilon_{n,\kappa}, \qquad \varepsilon_{n,\kappa} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \tag{17}$$

where $u_{n,\kappa}$ and $\widetilde{u}_{n,\kappa}$ denote the clean and perturbed state values, respectively, and ς_κ is the standard deviation of the κth state component computed from the clean training set. The sample-dependent perturbation intensity is defined as

$$\gamma_n = \begin{cases} \gamma_{\text{H}}, & n \in \mathcal{H}_\beta, \\ \gamma_{\text{L}}, & n \notin \mathcal{H}_\beta, \end{cases} \qquad 0 \le \gamma_{\text{L}} \le \gamma_{\text{H}}, \tag{18}$$

where $\mathcal{H}_\beta$ is a randomly selected subset containing a fraction β of the training samples. $(\gamma_{\text{L}}, \gamma_{\text{H}})$ controls the difference in perturbation intensity across samples.

Table 2: True-term coefficient error E_{true} (%) and support F1 on the remaining six PDE systems. N/A indicates unsupported tasks, and OOM denotes out-of-memory errors.

Method	Chaffee–Infante		Compound		Burgers2D		Adv.–diff. (2D)		NLS		λ–ω	
	$E_{\text{true}}\downarrow$	F1↑	$E_{\text{true}}\downarrow$	F1↑	$E_{\text{true}}\downarrow$	F1↑	$E_{\text{true}}\downarrow$	F1↑	$E_{\text{true}}\downarrow$	F1↑	$E_{\text{true}}\downarrow$	F1↑
$\beta = 0\%$												
PINN-SR (Nat. Commun., 2021)	105.78	0.364	87.791	0.154	83.498	0.222	0.7859	1.000	68.181	0.400	OOM	OOM
DeepMoD (JCP, 2021)	105.81	0.286	189.72	0.400	99.916	0.444	0.5710	1.000	78.689	0.136	41.122	0.933
R-DLGA (Phys. Rev. Res., 2021)	22.879	1.000	88.150	0.200	3.7240	1.000	**0.1493**	**1.000**	0.1567	1.000	61.137	0.741
PDE-READ (Neural Netw., 2022)	90.537	0.500	164.46	0.333	138.80	0.240	0.7422	1.000	42.218	0.800	86.819	0.250
PIC (Research, 2023)	10.318	1.000	57.849	0.500	81.414	0.400	0.4736	1.000	100.00	0.000	91.079	0.316
R-DISCOVER (Phys. Fluids, 2024)	61.036	0.333	100.00	0.000	2.0665	1.000	0.2131	1.000	30.063	0.625	53.643	0.833
Weak-PDE-LEARN (JCP, 2024)	99.583	0.250	**0.9183**	**1.000**	51.754	0.500	53.744	0.462	72.721	0.073	99.089	0.294
EqGPT (Nat. Commun., 2025)	0.2518	1.000	1.6261	1.000	**0.6983**	**1.000**	0.2391	0.889	N/A	N/A	N/A	N/A
ABL-PDE (NIPS, 2025)	71.333	0.200	67.226	0.500	43.986	0.571	0.2149	1.000	0.0923	1.000	**37.055**	**1.000**
SI-SR (Phys. Fluids, 2026)	**0.2252**	**1.000**	100.00	0.000	54.191	0.500	0.4245	1.000	**0.0674**	**1.000**	72.210	0.444
CARE-PDE	**0.0347**	**1.000**	**0.6492**	**1.000**	**0.5693**	**1.000**	**0.0361**	**1.000**	**0.0356**	**1.000**	**2.3197**	**1.000**
$\beta = 25\%$												
PINN-SR (Nat. Commun., 2021)	56.623	0.308	116.50	0.286	79.952	0.222	25.086	0.615	72.312	0.400	OOM	OOM
DeepMoD (JCP, 2021)	100.00	0.000	189.82	0.400	103.88	0.364	**0.3250**	**1.000**	66.916	0.154	35.754	0.933
R-DLGA (Phys. Rev. Res., 2021)	48.994	0.667	111.04	0.286	4.5333	1.000	79.084	0.250	16.968	1.000	64.136	0.690
PDE-READ (Neural Netw., 2022)	90.561	0.500	70.313	0.667	128.10	0.207	57.201	0.200	42.613	0.800	86.697	0.250
PIC (Research, 2023)	16.830	1.000	58.595	0.500	81.444	0.400	1.8611	1.000	100.00	0.000	83.648	0.545
R-DISCOVER (Phys. Fluids, 2024)	100.00	0.000	100.00	0.000	**3.9965**	**1.000**	0.8277	1.000	77.224	0.364	64.991	0.750
Weak-PDE-LEARN (JCP, 2024)	99.696	0.182	**1.4209**	**1.000**	51.323	0.500	54.543	0.667	77.153	0.062	87.452	0.195
EqGPT (Nat. Commun., 2025)	100.00	0.000	100.00	0.000	50.965	1.000	100.00	0.000	N/A	N/A	N/A	N/A
ABL-PDE (NIPS, 2025)	67.315	0.364	88.497	0.222	54.010	0.429	2.7265	0.667	0.8405	0.667	**39.727**	**0.966**
SI-SR (Phys. Fluids, 2026)	**0.3598**	**1.000**	100.00	0.000	53.708	0.500	1.7052	1.000	**0.5767**	**1.000**	85.714	0.235
CARE-PDE	**0.0879**	**1.000**	**0.4219**	**1.000**	**0.7345**	**1.000**	**0.0580**	**1.000**	**0.0729**	**1.000**	**4.1128**	**1.000**
$\beta = 50\%$												
PINN-SR (Nat. Commun., 2021)	85.118	0.143	114.01	0.364	63.189	0.500	50.838	0.364	67.613	0.400	OOM	OOM
DeepMoD (JCP, 2021)	100.00	0.000	190.08	0.400	81.634	0.400	62.268	0.500	53.200	0.136	35.076	0.903
R-DLGA (Phys. Rev. Res., 2021)	23.351	0.750	85.535	0.222	**3.7009**	**1.000**	82.129	0.250	1.4894	0.632	60.054	0.741
PDE-READ (Neural Netw., 2022)	90.527	0.500	100.00	0.000	100.00	0.000	59.148	0.286	42.738	0.800	86.667	0.250
PIC (Research, 2023)	47.383	0.800	60.985	0.500	54.179	0.667	2.2925	1.000	100.00	0.000	100.00	0.000
R-DISCOVER (Phys. Fluids, 2024)	85.812	0.500	81.736	0.222	6.6291	0.889	**1.2188**	**1.000**	70.717	0.400	68.016	0.750
Weak-PDE-LEARN (JCP, 2024)	99.644	0.182	**1.3024**	**1.000**	51.955	0.500	54.925	0.667	74.660	0.100	86.647	0.417
EqGPT (Nat. Commun., 2025)	100.00	0.000	100.00	0.000	100.00	0.000	100.00	0.000	N/A	N/A	N/A	N/A
ABL-PDE (NIPS, 2025)	73.373	0.200	78.777	0.286	100.00	0.000	6.5079	0.727	1.1066	0.706	**39.636**	**1.000**
SI-SR (Phys. Fluids, 2026)	**0.2835**	**1.000**	100.00	0.000	52.895	0.500	62.353	0.667	**0.2811**	**1.000**	100.00	0.000
CARE-PDE	**0.2750**	**1.000**	**0.8906**	**1.000**	**1.6862**	**1.000**	**0.3066**	**1.000**	**0.2153**	**1.000**	**10.246**	**1.000**

4.2 Implementation Details

All experiments are implemented in PyTorch and conducted on an NVIDIA GeForce RTX 4090 GPU. The surrogate model consists of five hidden layers with a width of 100 and uses rational activation functions. The robustness parameter in the Cauchy loss is set to $c = 1$. The surrogate model is trained for 50,000 epochs, with AGRE activated after an 8,500-epoch warm-up to estimate the reliability of sample-level observations. We construct $J = 5{,}000$ adaptive space–time windows for local equation identification. The Legendre projection degrees are set to 9 for the spatial coordinates and 3 for the temporal coordinate. The sparsity penalty in the equation-selection score is set to $\alpha = 0.15$. Additional implementation details and hyperparameter settings are provided in the Appendix E.3.

4.3 Comparison Methods

To evaluate the effectiveness of CARE-PDE, we select 10 representative PDE discovery methods as baselines for comparison, including PINN-SR (Chen et al., 2021), DeepMoD (Both et al., 2021), R-DLGA (Xu & Zhang, 2021), PDE-READ (Stephany & Earls, 2022), PIC (Xu et al., 2023), R-DISCOVER (Du et al., 2024), Weak-PDE-LEARN (Stephany & Earls, 2024), EqGPT (Xu et al., 2025), ABL-PDE (Gao et al., 2026), and SI-SR (Chen & Xiao, 2026). Since the baseline methods used problem-specific hyperparameter configurations in their original experiments, we tune their hyperparameters across different PDE systems to achieve optimal performance. Notably, CARE-PDE uses the same hyperparameter configuration across all PDE systems and noise settings, as described in Section 4.2, without requiring additional system-specific tuning.

4.4 Experimental Results and Analysis

We compare CARE-PDE with ten representative PDE discovery methods on 12 PDE systems. The results under NDEC with $\gamma_{\mathrm{L}} = 0$, $\gamma_{\mathrm{H}} = 25\%$, and $\beta \in \{0\%, 25\%, 50\%\}$ are reported in Tables 1 and 2. Detailed calculations of the evaluation metrics are provided in Appendix E.4, while the complete experimental results are presented in Appendix F.1. Throughout the following tables, the best and second-best results are highlighted in **red** and **blue**, respectively. The main observations are summarized as follows:

- CARE-PDE correctly identifies the complete equation structure in all 36 tasks while attaining the lowest coefficient errors among the compared methods. These results demonstrate the effectiveness of CARE-PDE in jointly recovering accurate PDE structures and coefficients under NDEC.
- As the proportion of highly perturbed observations increases from $\beta = 0\%$ to 50%, several baseline methods exhibit substantial degradation in structural or coefficient recovery. In contrast, CARE-PDE maintains perfect structural identification across all contamination ratios while preserving low coefficient errors. This indicates strong robustness to increasing sample-level heterogeneous contamination.
- For challenging systems involving high-order derivatives, two-dimensional dynamics, and multiple coupled state variables, several baseline methods exhibit structural errors, large coefficient deviations, or failed equation recovery. In contrast, CARE-PDE maintains a clear performance advantage across these complex settings. This demonstrates its applicability to PDE systems with varying dimensionality and dynamical complexity.

4.5 Ablation Study

To examine the contribution of each component in CARE-PDE, we conduct module-wise ablations over the 36 tasks considered in Section 4.4. Table 3 reports the averaged results, with detailed results provided in the Appendix F.8. Compared with the Backbone and the corresponding ablated variants, the complete CARE-PDE achieves more accurate coefficient recovery, more reliable structural identification, and a higher exact recovery rate. Removing RAFR weakens the suppression of highly perturbed observations during field reconstruction, while removing RGLP or LICA degrades the quality of local differential information and equation constraints, leading to less stable equation recovery. Removing SSR reduces the stability of structure selection and increases the risk of retaining spurious candidate terms. Overall, the results show that these components provide complementary robustness at different stages of PDE discovery and jointly contribute to the final recovery performance.

5 Conclusion

In this paper, to address PDE discovery under Noisy Data–Equation Correspondence (NDEC), we propose a Correspondence-Aware Robust Equation Discovery for PDEs (CARE-PDE) framework. To reduce the influence of highly perturbed observations, Reliability-Aware Field Reconstruction (RAFR) performs robust field reconstruction while estimating sample-level observation reliability. Building on the estimated reliability, Reliability-Guided Local Legendre Projection (RGLP) constructs reliable local supports and obtains stable differential information. The resulting strong-form information is further aggregated by Local Integral Constraint Aggregation (LICA) to reduce the influence of pointwise estimation errors on equation identification. Finally, Stability-Selected Sparse Regression (SSR) identifies stable equation structures while suppressing spurious candidate terms. Extensive experiments on 12 representative PDE systems demonstrate that CARE-PDE achieves superior structure identification, coefficient recovery, and robustness compared with ten representative PDE discovery methods under NDEC.

Table 3: Ablation study of CARE-PDE averaged over 36 tasks.

Variant	$E_{\mathrm{true}} \downarrow$	F1↑	Exact↑
w/o RAFR	13.527	0.909	30/36
w/o CRFR	6.4060	0.954	32/36
w/o AGRE	2.3713	0.991	35/36
w/o RGLP	8.9972	0.917	33/36
w/o LICA	10.216	0.915	28/36
w/o SSR	**1.0985**	**0.998**	**35/36**
w/o RGLP & LICA	12.714	0.871	27/36
Backbone	26.151	0.693	16/36
CARE-PDE	**0.9568**	**1.000**	**36/36**

AI USE STATEMENT

We declare that ChatGPT was used solely as a general-purpose assistive tool for language editing, grammar checking, and improving the clarity and readability of the manuscript. It was not used to develop the methodology, design experiments, collect or analyze data, or generate substantive scientific contributions. The authors take full responsibility for all contents of this paper, including the final text, technical claims, experimental results, and conclusions. No AI tool is listed as an author.

REPRODUCIBILITY STATEMENT

We have provided detailed information to facilitate the reproducibility of our results. The main text presents the overall CARE-PDE framework, experimental setup, and comparison protocol. The appendices provide additional methodological details, the robustness analysis of the Cauchy loss, the complete algorithmic procedure, and detailed experimental settings, including the PDE systems, hyperparameter configuration, and evaluation metrics. Complete experimental results and additional analyses are also reported in the appendices. These descriptions specify the procedures and settings used throughout the experiments and are intended to support independent reproduction of the reported results.

REFERENCES

Clare I Abreu, Jonathan Friedman, Vilhelm L Andersen Woltz, and Jeff Gore. Mortality causes universal changes in microbial community composition. *Nature communications*, 10(1):2120, 2019.

Daniel R Amor, Christoph Ratzke, and Jeff Gore. Transient invaders can induce shifts between alternative stable states of microbial communities. *Science advances*, 6(8):eaay8676, 2020.

Mehrad Ansari, Heta A Gandhi, David G Foster, and Andrew D White. Iterative symbolic regression for learning transport equations. *AIChE Journal*, 68(6):e17695, 2022.

Sarah Beetham, Rodney O Fox, and Jesse Capecelatro. Sparse identification of multiphase turbulence closures for coupled fluid–particle flows. *Journal of Fluid Mechanics*, 914:A11, 2021.

Gert-Jan Both, Subham Choudhury, Pierre Sens, and Remy Kusters. Deepmod: Deep learning for model discovery in noisy data. *Journal of Computational Physics*, 428:109985, 2021.

Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

Kathleen Champion, Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116 (45):22445–22451, 2019.

Boyuan Chen, Kuang Huang, Sunand Raghupathi, Ishaan Chandratreya, Qiang Du, and Hod Lipson. Automated discovery of fundamental variables hidden in experimental data. *Nature Computational Science*, 2(7):433–442, 2022.

Haoran Chen and Dunhui Xiao. Symmetry inspired learning of governing equations from noisy data. *Physics of Fluids*, 38(1), 2026.

Zhao Chen, Yang Liu, and Hao Sun. Physics-informed learning of governing equations from scarce data. *Nature communications*, 12(1):6136, 2021.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1): 1–22, 1977.

Mengge Du, Yuntian Chen, Longfeng Nie, Siyu Lou, and Dongxiao Zhang. Physics-constrained robust learning of open-form partial differential equations from limited and noisy data. *Physics of Fluids*, 36(5), 2024.

Lawrence C Evans. *Partial differential equations*, volume 19. American mathematical society, 2022.

En-Hao Gao, Cunjing Ge, Yuan Jiang, and Zhi-Hua Zhou. Discovering symbolic partial differential equation by abductive learning. *Advances in Neural Information Processing Systems*, 38:57359–57385, 2026.

Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31, 2018.

Alan A Kaptanoglu, Jared L Callaham, Aleksandr Aravkin, Christopher J Hansen, and Steven L Brunton. Promoting global stability in data-driven models of quadratic nonlinear dynamics. *Physical Review Fluids*, 6(9):094401, 2021.

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.

Xin-Yang Liu, Min Zhu, Lu Lu, Hao Sun, and Jian-Xun Wang. Multi-resolution partial differential equations preserved learning framework for spatiotemporal dynamics. *Communications Physics*, 7(1):31, 2024.

Peter Y Lu, Joan Ariño Bernad, and Marin Soljačić. Discovering sparse interpretable dynamics from partial observations. *Communications Physics*, 5(1):206, 2022.

Wenjun Ma, Jun Zhang, Kaikai Feng, Haoyun Xing, and Dongsheng Wen. Dimensional homogeneity constrained gene expression programming for discovering governing equations. *Journal of Fluid Mechanics*, 985:A12, 2024.

Daniel A Messenger and David M Bortz. Weak sindy for partial differential equations. *Journal of computational physics*, 443:110525, 2021.

David Peel and Geoffrey MacLahlan. Finite mixture models. *John & Sons*, 10(0471721182):10–1002, 2000.

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.

Patrick AK Reinbold, Daniel R Gurevich, and Roman O Grigoriev. Using noisy or incomplete data to discover models of spatiotemporal dynamics. *Physical Review E*, 101(1):010203, 2020.

Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science advances*, 3(4):e1602614, 2017.

Jie Shen, Tao Tang, and Li-Lian Wang. *Spectral methods: algorithms, analysis and applications*. Springer Science & Business Media, 2011.

Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin, and Jae-Gil Lee. Learning from noisy labels with deep neural networks: A survey. *IEEE transactions on neural networks and learning systems*, 34(11):8135–8153, 2022.

Wenxiang Song, Shijie Jiang, Gustau Camps-Valls, Mathew Williams, Lu Zhang, Markus Reichstein, Harry Vereecken, Leilei He, Xiaolong Hu, and Liangsheng Shi. Towards data-driven discovery of governing equations in geosciences. *Communications Earth & Environment*, 5(1):589, 2024.

Robert Stephany and Christopher Earls. Pde-read: Human-readable partial differential equation discovery using deep learning. *Neural Networks*, 154:360–382, 2022.

Robert Stephany and Christopher Earls. Weak-pde-learn: A weak form based approach to discovering pdes from noisy, limited data. *Journal of Computational Physics*, 506:112950, 2024.

Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021.

Hao Xu and Dongxiao Zhang. Robust discovery of partial differential equations in complex situations. *Physical Review Research*, 3(3):033270, 2021.

Hao Xu, Haibin Chang, and Dongxiao Zhang. Dlga-pde: Discovery of pdes with incomplete candidate library via combination of deep learning and genetic algorithm. *Journal of Computational Physics*, 418:109584, 2020.

Hao Xu, Dongxiao Zhang, and Nanzhe Wang. Deep-learning based discovery of partial differential equations in integral form from sparse and noisy data. *Journal of Computational Physics*, 445: 110592, 2021.

Hao Xu, Junsheng Zeng, and Dongxiao Zhang. Discovery of partial differential equations from highly noisy and sparse data with physics-informed information criterion. *Research*, 6:0147, 2023.

Hao Xu, Yuntian Chen, Rui Cao, Tianning Tang, Mengge Du, Jian Li, Adrian H Callaghan, and Dongxiao Zhang. Generative discovery of partial differential equations by learning from math handbooks. *Nature Communications*, 16(1):10255, 2025.

Laure Zanna and Thomas Bolton. Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, 47(17):e2020GL088376, 2020.

A Notation Summary

Table 4 summarizes the main notation used throughout the paper. Auxiliary symbols that are specific to individual procedures are defined locally when they first appear.

Table 4: Summary of the main notation used in CARE-PDE.

Symbol	Description
$\mathbf{x} \in \mathbb{R}^d$	Spatial coordinate with spatial dimension d.
t	Temporal coordinate.
$\mathbf{z} = (\mathbf{x}, t)$	Spatiotemporal coordinate.
$D = d + 1$	Dimension of the spatiotemporal domain.
$\mathbf{u}(\mathbf{x}, t) \in \mathbb{R}^M$	State field with M state components.
$\mathcal{D} = \{(\mathbf{x}_i, t_i, \mathbf{y}_i)\}_{i=1}^N$	Set of N observed spatiotemporal samples.
$\mathbf{y}_i$	Observed state variables at the ith spatiotemporal location.
$\boldsymbol{\varepsilon}_i$	Observation error associated with the ith sample.
$\boldsymbol{\psi}[\mathbf{u}] = [\psi_1[\mathbf{u}], \ldots, \psi_F[\mathbf{u}]]^\top$	Candidate library containing F candidate terms.
$\boldsymbol{\Xi} \in \mathbb{R}^{F \times M}$	Coefficient matrix of the governing PDE system.
$\boldsymbol{\xi}_m = \boldsymbol{\Xi}_{:,m}$	Coefficient vector for state component m.
$\mathbf{f}_{\boldsymbol{\theta}}$	Neural surrogate used to reconstruct the continuous state field.
δ_i	Reconstruction residual of observation i.
$\rho_i \in [0, 1]$	Reliability estimate of observation i.
a_i	Reliability-based observation weight used in field reconstruction.
W_j	jth local spatiotemporal window used by RGLP and LICA.
$\mathbf{c}_j$	Center of window W_j.
$\mathbf{h}_j = [h_{j1}, \ldots, h_{jD}]^\top$	Half-widths of window W_j along the spatiotemporal dimensions.
$\boldsymbol{\eta}$	Normalized local coordinate in the reference domain $[-1, 1]^D$.
n_j^{eff}	Effective reliable support within window W_j.
n_{tar}	Target effective reliable support used for adaptive window construction.
J	Number of accepted local windows.
Q	Number of tensor-product Gauss–Legendre quadrature nodes in each window.
$\mathcal{I}_j[\cdot]$	Local integral operator associated with window W_j.
φ_j	Nonnegative test function used for local integral aggregation.
b_{jm}	Integrated temporal-derivative term for state component m in window W_j.
$\boldsymbol{\Theta}$	Integrated candidate matrix constructed from all accepted windows.
$\overline{\rho}_j$	Test-function-weighted average observation reliability within window W_j.
w_j	Reliability weight assigned to the jth integrated PDE constraint.
β	Fraction of training observations assigned the high perturbation level in the NDEC construction.
$\gamma_{\text{L}}, \gamma_{\text{H}}$	Low and high perturbation levels, respectively, with $\gamma_{\text{L}} < \gamma_{\text{H}}$.
$\mathcal{H}_\beta$	Subset of training observations assigned the high perturbation level.
ς_κ	Standard deviation of the κth state component in the clean training data.

B Additional Method Details

B.1 Details of Adaptive GMM Reliability Estimation

This section further describes Adaptive GMM Reliability Estimation (AGRE), including residual feature construction, Gaussian mixture modeling, adaptive component selection, and observation-level reliability estimation.

Residual feature construction. For observation i, we compute the root-mean-square reconstruction residual across the M standardized state components:

$$\delta_i = \left[\frac{1}{M} \left\| \mathbf{f}_{\boldsymbol{\theta}}(\widetilde{\mathbf{z}}_i) - \widetilde{\mathbf{y}}_i \right\|_2^2 \right]^{1/2}. \tag{19}$$

Because the reconstruction residual distribution may contain a long upper tail, directly fitting a mixture model to δ_i can make the estimation sensitive to a small number of extreme residuals. We therefore apply a median-scaled logarithmic transformation:

$$\ell_i = \log\left(1 + \frac{\delta_i}{\mathrm{median}_j\, \delta_j + \epsilon}\right), \tag{20}$$

where $\epsilon > 0$ is a small constant for numerical stability. The transformed residuals are subsequently standardized as

$$s_i = \frac{\ell_i - \bar{\ell}}{\sigma_\ell}, \qquad \bar{\ell} = \frac{1}{N}\sum_{i=1}^{N} \ell_i, \qquad \sigma_\ell^2 = \frac{1}{N}\sum_{i=1}^{N} (\ell_i - \bar{\ell})^2. \tag{21}$$

If $\sigma_\ell = 0$, we set $s_i = 0$ for all observations.

Gaussian mixture modeling and adaptive selection. AGRE models the standardized residual features using Gaussian mixture models (GMMs). For a model with G components, the mixture density is

$$p(s_i \mid \mathbf{\Phi}_G) = \sum_{g=1}^{G} \pi_g \mathcal{N}(s_i; \mu_g, \sigma_g^2), \qquad G \in \{1, 2\}, \tag{22}$$

where π_g, μ_g, and σ_g^2 denote the mixing coefficient, mean, and variance of component g, respectively, and $\mathbf{\Phi}_G = \{\pi_g, \mu_g, \sigma_g^2\}_{g=1}^{G}$. The mixture parameters are estimated using expectation–maximization (Dempster et al., 1977).

Since the residual distribution does not necessarily exhibit a clear grouping structure, always adopting a two-component model may introduce artificial distinctions among observations with similar residual characteristics. AGRE therefore adaptively determines whether differentiated reliability estimation is warranted by comparing one- and two-component mixture models. For each model, the Bayesian information criterion (BIC) is computed as

$$\mathrm{BIC}_G = -2 \log \widehat{L}_G + \kappa_G \log N, \qquad G \in \{1, 2\}, \tag{23}$$

where $\widehat{L}_G$ is the maximized likelihood and κ_G denotes the number of free parameters.

Although BIC jointly accounts for both fitting quality and model complexity, the greater flexibility of a two-component model may still yield a lower BIC even when the fitted distribution does not exhibit two clearly separated modes. We therefore further evaluate the modal separation of the fitted two-component distribution. Specifically, let $p_2(s) = p(s \mid \widehat{\mathbf{\Phi}}_2)$ denote the fitted two-component density, and order the component means such that $\mu_1 < \mu_2$. Using their midpoint $m = (\mu_1 + \mu_2)/2$, we identify $\widehat{s}_1$ and $\widehat{s}_2$ as the highest local maxima of $p_2(s)$ on the left and right of m, respectively. When both modes exist, the minimum-density location between them is defined as

$$s_v = \arg\min_{s \in [\widehat{s}_1, \widehat{s}_2]} p_2(s), \tag{24}$$

and the corresponding valley-to-peak ratio is

$$V = \frac{p_2(s_v)}{\min\{p_2(\widehat{s}_1), p_2(\widehat{s}_2)\}}. \tag{25}$$

A smaller V indicates a deeper valley between the two modes and hence clearer modal separation. Combining BIC with this separation measure, AGRE selects the number of mixture components as

$$G^* = \begin{cases} 2, & \mathrm{BIC}_2 < \mathrm{BIC}_1 \text{ and } V < V_{\mathrm{th}}, \\ 1, & \text{otherwise}, \end{cases} \tag{26}$$

where V_{th} specifies the required degree of modal separation. The two-component model is therefore retained only when it yields a lower BIC and exhibits sufficient separation.

Reliability estimation. When $G^* = 2$, the component with the smaller mean, denoted by $g_{\mathrm{rel}} = \arg\min_{g\in\{1,2\}} \mu_g$, is treated as the higher-reliability component. The reliability of observation i is defined as its posterior probability of belonging to this component:

$$\rho_i = \frac{\pi_{g_{\mathrm{rel}}}\mathcal{N}(s_i;\mu_{g_{\mathrm{rel}}},\sigma^2_{g_{\mathrm{rel}}})}{\sum_{g=1}^{2}\pi_g\mathcal{N}(s_i;\mu_g,\sigma_g^2)}, \qquad \rho_i \in [0,1]. \tag{27}$$

A larger ρ_i indicates a stronger association with the lower-residual component and thus higher estimated reliability. When $G^* = 1$, the residual distribution does not provide sufficient evidence for differentiated reliability, and we set $\rho_i = 1$ for all observations. The resulting reliability estimates are converted into the bounded reconstruction weights defined in Eq. 4 and remain fixed during the subsequent reliability-weighted CRFR optimization.

B.2 Details of Reliability-Guided Local Legendre Projection

This section further describes Reliability-Guided Local Legendre Projection (RGLP), including local field evaluation, Legendre projection, Gauss–Legendre quadrature, and derivative recovery.

Local field evaluation. For each accepted window W_j, we map the local coordinates to the reference domain $[-1,1]^D$ using the normalized coordinate $\boldsymbol{\eta}$ defined in Eq. 6. We introduce Q tensor-product Gauss–Legendre quadrature nodes $\{\boldsymbol{\eta}^{(q)}\}_{q=1}^{Q}$, where $\boldsymbol{\eta}^{(q)} = (\eta_1^{(q)},\dots,\eta_D^{(q)})^\top$. For the qth quadrature node, we obtain the corresponding physical coordinate and reconstructed field value as

$$\mathbf{z}^{(j,q)} = \mathbf{c}_j + \mathbf{h}_j \odot \boldsymbol{\eta}^{(q)}, \qquad \widehat{\mathbf{u}}^{(j,q)} = \mathbf{f}_{\boldsymbol{\theta}}\left(\mathbf{z}^{(j,q)}\right), \tag{28}$$

where $\odot$ denotes elementwise multiplication. These evaluations provide the local field information used for Legendre projection.

Local Legendre projection. We approximate each state component within W_j using a truncated tensor-product Legendre expansion:

$$v_m^{(j)}(\boldsymbol{\eta}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} A_{m,\boldsymbol{\alpha}}^{(j)} \prod_{r=1}^{D} P_{\alpha_r}(\eta_r), \tag{29}$$

where $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_D)$ denotes the polynomial multi-index, $\mathcal{A}$ is the retained set of Legendre modes, P_{α_r} denotes the Legendre polynomial of degree α_r along dimension r, and $A_{m,\boldsymbol{\alpha}}^{(j)}$ is the corresponding projection coefficient. Using the orthogonality of Legendre polynomials, we compute the projection coefficients as

$$\begin{aligned} A_{m,\boldsymbol{\alpha}}^{(j)} &= \int_{[-1,1]^D} \left[\widehat{\mathbf{u}}^{(j)}(\boldsymbol{\eta})\right]_m \prod_{r=1}^{D} \frac{2\alpha_r+1}{2} P_{\alpha_r}(\eta_r)\, d\boldsymbol{\eta} \\ &\approx \sum_{q=1}^{Q} \omega_{\mathrm{GL}}^{(q)} \left[\widehat{\mathbf{u}}^{(j,q)}\right]_m \prod_{r=1}^{D} \frac{2\alpha_r+1}{2} P_{\alpha_r}\left(\eta_r^{(q)}\right), \end{aligned} \tag{30}$$

where $\omega_{\mathrm{GL}}^{(q)}$ denotes the tensor-product Gauss–Legendre quadrature weight associated with the qth quadrature node, defined as

$$\omega_{\mathrm{GL}}^{(q)} = \prod_{r=1}^{D} \omega_{\mathrm{GL},r}^{(q)}, \tag{31}$$

where $\omega_{\mathrm{GL},r}^{(q)}$ denotes the one-dimensional Gauss–Legendre weight associated with $\eta_r^{(q)}$.

Derivative recovery. After obtaining the local polynomial representation, we recover the required spatial and temporal derivatives by analytically differentiating the Legendre basis. For the kth derivative of state component u_m with respect to physical coordinate z_r, we compute

$$\widehat{\partial_{z_r}^k u_m}^{(j)}(\boldsymbol{\eta}) = h_{jr}^{-k} \sum_{\boldsymbol{\alpha}\in\mathcal{A}} A_{m,\boldsymbol{\alpha}}^{(j)} P_{\alpha_r}^{(k)}(\eta_r) \prod_{\substack{s=1\\ s\neq r}}^{D} P_{\alpha_s}(\eta_s), \tag{32}$$

where $P_{\alpha_r}^{(k)}$ denotes the kth derivative of P_{α_r} with respect to η_r. The factor h_{jr}^{-k} follows from the coordinate transformation $\eta_r = (z_r - c_{jr})/h_{jr}$ and converts derivatives with respect to the reference coordinate into derivatives with respect to the physical coordinate.

Through this local projection, we differentiate the finite-dimensional Legendre representation rather than the reconstructed field directly, which reduces the influence of rapidly varying reconstruction errors on the recovered differential information.

B.3 Details of Stability-Selected Sparse Regression

This section further describes Stability-Selected Sparse Regression (SSR), including stability-based support selection, reliability-weighted sparse regression, and checkpoint and threshold selection.

Reliability-weighted STRidge. After obtaining the stable support $\mathcal{S}_m^{\mathrm{stable}}$ defined in Eq. 15, we restrict the candidate library to the retained terms and estimate their coefficients using all J integrated constraints. To incorporate the reliability of different constraints, we solve a weighted ridge regression problem over the current active set $\mathcal{S}$:

$$\widehat{\boldsymbol{\xi}}_{m,\mathcal{S}} = \arg\min_{\boldsymbol{\xi}} \sum_{j=1}^{J} w_j \left(b_{jm} - \boldsymbol{\Theta}_{j,\mathcal{S}}\boldsymbol{\xi}\right)^2 + \lambda\|\boldsymbol{\xi}\|_2^2, \tag{33}$$

where $\boldsymbol{\Theta}_{j,\mathcal{S}}$ denotes the entries of the jth regression row associated with the active candidate set, and λ is the ridge regularization coefficient. Before regression, we normalize the candidate columns to reduce the influence of differences in term magnitude. We then remove coefficients whose magnitudes fall below the sparsity threshold τ and refit the remaining active terms. We repeat this thresholding and refitting procedure until the active support no longer changes. Finally, we restore the estimated coefficients to their original scales.

Checkpoint and threshold evaluation. For each candidate checkpoint c and sparsity threshold τ, SSR produces a coefficient vector $\widehat{\boldsymbol{\xi}}_{cm}(\tau)$ for each state component. The checkpoint-selection score defined in Eq. 16 uses the normalized reliability-weighted regression residual

$$R_{cm}(\tau) = \frac{\displaystyle\sum_{j=1}^{J} w_j \left[b_{jm} - \boldsymbol{\Theta}_{j:}\widehat{\boldsymbol{\xi}}_{cm}(\tau)\right]^2}{\displaystyle\sum_{j=1}^{J} w_j \left(b_{jm} - \bar{b}_m\right)^2 + \epsilon}, \tag{34}$$

where

$$\bar{b}_m = \frac{\sum_{j=1}^{J} w_j b_{jm}}{\sum_{j=1}^{J} w_j + \epsilon} \tag{35}$$

is the reliability-weighted mean of the integrated temporal-derivative terms. This normalization measures the PDE regression error relative to the weighted variation of the target constraints, allowing the residuals of different state equations to be compared on a common scale.

We evaluate each checkpoint–threshold pair using the fit–complexity score $\mathcal{Q}_{c,\tau}^{\mathrm{PDE}}$ and select

$$(c^*, \tau^*) = \arg\min_{c\in\mathcal{C},\,\tau\in\mathcal{T}} \mathcal{Q}_{c,\tau}^{\mathrm{PDE}}, \tag{36}$$

where $\mathcal{C}$ and $\mathcal{T}$ denote the candidate checkpoint and sparsity-threshold sets, respectively. The coefficient vectors associated with (c^*, τ^*) determine the final discovered PDE system.

C Robustness Analysis of the Cauchy Loss

In this section, we provide a theoretical analysis of the robustness of the Cauchy loss against extreme reconstruction residuals.

For the ith observation, the per-sample Cauchy loss in Eq. 3 is

$$\ell_{\mathrm{C}}(\delta_i) = \frac{c^2}{2}\log\left(1 + \frac{\delta_i^2}{c^2\sigma_{\mathrm{rob}}^2}\right), \quad \delta_i^2 = \frac{1}{M}\left\|\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_i) - \mathbf{y}_i\right\|_2^2 \tag{37}$$

where δ_i is the scalar reconstruction residual of the ith observation. Differentiating the loss with respect to δ_i gives

$$\frac{\partial \ell_{\mathrm{C}}(\delta_i)}{\partial \delta_i} = \frac{c^2\delta_i}{c^2\sigma_{\mathrm{rob}}^2 + \delta_i^2}. \tag{38}$$

Since $c^2\sigma_{\mathrm{rob}}^2 + \delta_i^2 \geq 2c\sigma_{\mathrm{rob}}\delta_i$, we obtain

$$0 \leq \frac{\partial \ell_{\mathrm{C}}(\delta_i)}{\partial \delta_i} \leq \frac{c}{2\sigma_{\mathrm{rob}}}, \qquad \lim_{\delta_i \to \infty} \frac{\partial \ell_{\mathrm{C}}(\delta_i)}{\partial \delta_i} = 0. \tag{39}$$

Let $\mathbf{e}_i = \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_i) - \mathbf{y}_i$ denote the reconstruction error vector of the ith observation. Since $\delta_i = \|\mathbf{e}_i\|_2/\sqrt{M}$, the chain rule gives

$$\nabla_{\mathbf{e}_i}\ell_{\mathrm{C}}(\delta_i) = \frac{\partial \ell_{\mathrm{C}}(\delta_i)}{\partial \delta_i}\frac{\mathbf{e}_i}{M\delta_i} = \frac{c^2\mathbf{e}_i}{Mc^2\sigma_{\mathrm{rob}}^2 + \|\mathbf{e}_i\|_2^2}. \tag{40}$$

Taking its Euclidean norm yields

$$\left\|\nabla_{\mathbf{e}_i}\ell_{\mathrm{C}}(\delta_i)\right\|_2 = \frac{1}{\sqrt{M}}\frac{\partial \ell_{\mathrm{C}}(\delta_i)}{\partial \delta_i} \leq \frac{c}{2\sqrt{M}\sigma_{\mathrm{rob}}}, \tag{41}$$

and

$$\lim_{\delta_i \to \infty}\left\|\nabla_{\mathbf{e}_i}\ell_{\mathrm{C}}(\delta_i)\right\|_2 = 0. \tag{42}$$

Let

$$\mathbf{J}_i = \frac{\partial \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_i)}{\partial \boldsymbol{\theta}} \tag{43}$$

denote the Jacobian of the surrogate model with respect to its parameters. The contribution of the ith observation to the parameter gradient is

$$\nabla_{\boldsymbol{\theta}}\ell_{\mathrm{C}}(\delta_i) = \mathbf{J}_i^{\top}\nabla_{\mathbf{e}_i}\ell_{\mathrm{C}}(\delta_i). \tag{44}$$

For a bounded network Jacobian, $\|\mathbf{J}_i\|_2 \leq C_J$, it follows that

$$\left\|\nabla_{\boldsymbol{\theta}}\ell_{\mathrm{C}}(\delta_i)\right\|_2 \leq C_J\left\|\nabla_{\mathbf{e}_i}\ell_{\mathrm{C}}(\delta_i)\right\|_2 \leq \frac{C_J c}{2\sqrt{M}\sigma_{\mathrm{rob}}}, \tag{45}$$

with

$$\lim_{\delta_i \to \infty}\left\|\nabla_{\boldsymbol{\theta}}\ell_{\mathrm{C}}(\delta_i)\right\|_2 = 0. \tag{46}$$

These results demonstrate that the optimization influence of an individual observation is bounded, preventing large reconstruction residuals from producing arbitrarily large parameter updates. Moreover, as the residual magnitude increases, the corresponding gradient contribution gradually decreases, approaching zero. Therefore, highly perturbed observations have a limited influence on the reconstruction network optimization, which explains the robustness of the Cauchy loss.

D Algorithmic Description

To clearly illustrate the overall procedure of CARE-PDE, we summarize the main computational steps in Algorithm 1.

Algorithm 1 The pseudo code of the proposed CARE-PDE

Input: Observation dataset $\mathcal{D} = \{(\mathbf{z}_i, \mathbf{y}_i)\}_{i=1}^N$; candidate library $\boldsymbol{\psi}$; warm-up epoch T_{w}; total training epoch $T_{\max}$; number of retained checkpoints H; number of local constraints J; number of stability subsamples B; candidate threshold set $\mathcal{T}$;
Initialize the surrogate model $\mathbf{f}_{\boldsymbol{\theta}}$;
for $epoch = 1$ to T_{w} **do**
  Compute reconstruction residuals $\{\delta_i\}_{i=1}^N$;
  Compute $\mathcal{L}_{\mathrm{CRFR}}$ using the Cauchy loss;
  Update $\boldsymbol{\theta}$ by minimizing $\mathcal{L}_{\mathrm{CRFR}}$;
end for
Apply AGRE to $\{\delta_i\}_{i=1}^N$ and obtain observation reliabilities $\{\rho_i\}_{i=1}^N$;
Compute reconstruction weights $\{a_i\}_{i=1}^N$ from $\{\rho_i\}_{i=1}^N$;
for $epoch = T_{\mathrm{w}} + 1$ to $T_{\max}$ **do**
  Compute $\mathcal{L}_{\mathrm{RAFR}}$ using $\{a_i\}_{i=1}^N$;
  Update $\boldsymbol{\theta}$ by minimizing $\mathcal{L}_{\mathrm{RAFR}}$;
  Save surrogate checkpoints and their reconstruction errors;
end for
Retain the H surrogate checkpoints with the lowest reconstruction errors;
for each retained checkpoint c **do**
  Load the surrogate model $\mathbf{f}_{\boldsymbol{\theta}_c}$;
  for $j = 1$ to J **do**
    Construct the reliability-guided local window W_j using $\{\rho_i\}_{i=1}^N$;
    Evaluate $\mathbf{f}_{\boldsymbol{\theta}_c}$ at Gauss–Legendre quadrature nodes in W_j;
    Perform local Legendre projection and recover the required derivatives;
    Construct the local integral PDE constraint by LICA;
    Compute the corresponding constraint weight w_j;
  end for
  Assemble the integrated regression system $\mathbf{b}_m \approx \boldsymbol{\Theta}\boldsymbol{\xi}_m$ for each state equation m;
  for each $\tau \in \mathcal{T}$ **do**
    for $m = 1$ to M **do**
      for $b = 1$ to B **do**
        Uniformly sample a fixed-size subset of the J constraints without replacement;
        Apply reliability-weighted STRidge with threshold τ;
        Record the selected candidate terms;
      end for
      Compute the inclusion frequencies and obtain the stable support $\mathcal{S}_m^{\mathrm{stable}}$;
      Refit reliability-weighted STRidge on all J constraints restricted to $\mathcal{S}_m^{\mathrm{stable}}$;
    end for
    Compute the unsupervised PDE score $\mathcal{Q}_{c,\tau}^{\mathrm{PDE}}$;
  end for
end for
Select $(c^*, \tau^*) = \arg\min_{c,\tau} \mathcal{Q}_{c,\tau}^{\mathrm{PDE}}$;
Output: Discovered PDE structure and coefficients $\widehat{\boldsymbol{\Xi}}$;

E Experimental Details

E.1 PDE Systems

Table 5 summarizes the 12 PDE systems used in our experiments, including their spatial–temporal resolutions, computational domains, and governing equations. Among them, the nonlinear Schrödinger (NLS) equation is complex-valued. We write the complex state as $u = a - ib$, and represent it as the coupled real-valued system

$$\begin{aligned} a_t &= 0.5b_{xx} + a^2 b + b^3, \\ b_t &= -0.5a_{xx} - a^3 - ab^2. \end{aligned} \tag{47}$$

For the λ–ω system, r denotes the local amplitude satisfying $r^2 = u^2 + v^2$. For the two-dimensional PDE systems, $\Delta u = u_{xx} + u_{yy}$ denotes the spatial Laplacian.

Table 5: Details of the PDE systems used in experiments. N_t and N_s denote the numbers of temporal snapshots and spatial grid points, respectively. Ω and T denote the spatial and temporal domains.

PDE system	N_t	N_s	Ω	T	Governing Equation(s)
Advection	201	1024	$x \in (0, 1)$	$t \in [0, 2]$	$u_t = -0.1u_x$
Conv.–diff.	100	256	$x \in [0, 2]$	$t \in [0, 1]$	$u_t = -u_x + 0.25u_{xx}$
Burgers	201	256	$x \in [-8, 8)$	$t \in [0, 10]$	$u_t = 0.1u_{xx} - uu_x$
KdV	201	512	$x \in [-1, 1)$	$t \in [0, 1]$	$u_t = -uu_x - 0.0025u_{xxx}$
KS	251	1024	$x \in (0, 100)$	$t \in [0, 100]$	$u_t = -uu_x - u_{xx} - u_{xxxx}$
Allen–Cahn	201	256	$x \in [-1, 1)$	$t \in [0, 10]$	$u_t = 0.003u_{xx} + u - u^3$
Chaffee–Infante	200	301	$x \in [0, 3]$	$t \in [0.1, 0.5)$	$u_t = u_{xx} + u^3 - u$
Compound	201	201	$x \in [0, 1]$	$t \in [0, 1]$	$u_t = 0.1(u^2)_{xx}$
Burgers2D	100	101×51	$x, y \in [-1, 1]$	$t \in [0, 2)$	$u_t = 0.01\Delta u - u(u_x + u_y)$
Adv.–diff. (2D)	61	51×51	$x, y \in [-5, 5]$	$t \in [3, 9]$	$u_t = 0.5\Delta u + 0.25u_x + 0.5u_y$
NLS	251	256	$x \in [-5, 5)$	$t \in [0, 4)$	$u_t = 0.5iu_{xx} + i\lvert u\rvert^2 u$
λ–ω	201	256×256	$x, y \in [-10, 10)$	$t \in [0, 10]$	$\begin{cases} u_t = 0.1\Delta u + u + (v-u)r^2, \\ v_t = 0.1\Delta v + v - (u+v)r^2 \end{cases}$

E.2 Candidate Library Construction

We construct the candidate library using a unified generation rule for all PDE systems. Let $\mathbf{u} = (u_1, \ldots, u_M)^\top$ denote the M state components. We first define the set of state monomials with total degree no greater than p as

$$\mathcal{P}_p = \left\{ \prod_{m=1}^{M} u_m^{\alpha_m} : \boldsymbol{\alpha} \in \mathbb{N}_0^M,\ |\boldsymbol{\alpha}| = \sum_{m=1}^{M} \alpha_m \le p \right\}, \tag{48}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$ is a multi-index of nonnegative integers. We further define the spatial derivative set up to order q as

$$\mathcal{G}_q = \left\{ \partial_{x_\ell}^s u_m : m = 1, \ldots, M,\ \ell = 1, \ldots, d,\ s = 1, \ldots, q \right\}, \tag{49}$$

which contains spatial derivatives of each state component along every spatial coordinate direction. The candidate terms underlying $\boldsymbol{\psi}[\mathbf{u}]$ are generated as

$$\{\psi_f[\mathbf{u}]\}_{f=1}^{F} = \mathcal{P}_{p_{\text{state}}} \cup (\mathcal{P}_{p_{\text{der}}} \mathcal{G}_{q_{\max}}) \cup \mathcal{G}_{q_{\max}}^{(2)}, \tag{50}$$

where

$$\mathcal{P}_{p_{\text{der}}} \mathcal{G}_{q_{\max}} = \{\phi g : \phi \in \mathcal{P}_{p_{\text{der}}},\ g \in \mathcal{G}_{q_{\max}}\}, \tag{51}$$

and

$$\mathcal{G}_{q_{\max}}^{(2)} = \left\{ g^2 : g \in \mathcal{G}_{q_{\max}} \right\}. \tag{52}$$

The resulting number of candidate terms for a PDE system with M state components and d spatial dimensions is

$$F = \binom{M+5}{5} + dMq_{\max}\binom{M+3}{3} + dMq_{\max}. \tag{53}$$

Throughout the experiments, we set

$$p_{\text{state}} = 5, \qquad p_{\text{der}} = 3, \qquad q_{\max} = 4. \tag{54}$$

These settings yield 26 candidate terms for one-dimensional scalar systems, 109 for one-dimensional two-component systems, 46 for two-dimensional scalar systems, and 197 for two-dimensional two-component systems.

E.3 Implementation Details

All experiments are implemented in PyTorch and conducted on an NVIDIA GeForce RTX 4090 GPU. Across all experiments on the 12 PDE systems, we use a single fixed parameter configuration without system- or task-specific adjustment. For clarity, we organize the implementation settings into three stages: field reconstruction and reliability estimation, local differential estimation and

Table 6: Shared implementation settings of CARE-PDE. The same configuration is used across all PDE systems and experimental tasks.

Setting	Configuration
Field reconstruction and reliability estimation	
Surrogate model $\mathbf{f}_{\boldsymbol{\theta}}$	5 hidden layers, width 100, rational activation
Optimization	Adam; $T_{\max} = 50{,}000$ epochs
Learning rate	$5 \times 10^{-4} \rightarrow 10^{-4}$ during the first $T_{\mathrm{w}} = 8{,}500$ epochs; fixed thereafter
CRFR	$c = 1$; $\sigma_{\min} = 0.1$
AGRE	Applied once at $T_{\mathrm{w}} = 8{,}500$; $V_{\mathrm{th}} = 0.9$; $a_{\min} = 0.1$;
Checkpoint retention	$H = 5$ checkpoints retained
Local differential estimation and constraint construction	
Local constraints	$J = 5{,}000$ adaptive space–time windows
Legendre projection	Maximum degree 9 along each spatial coordinate and 3 along time
Initial / maximum half-width fractions	Space: $0.0425/0.055$; time: $0.0475/0.075$, relative to the full spatial/temporal domain length
Reliable support	$n_{\mathrm{tar}} = 20$
Constraint weighting	$n_0 = 10$
Test function	$\nu_x = \nu_y = 4$, $\nu_t = 3$;
Gauss–Legendre quadrature	41 nodes per coordinate for 1D spatial systems; 12 nodes per coordinate for 2D spatial systems
Sparse equation identification	
STRidge	$\lambda = 10^{-6}$; at most 20 iterations; $\lvert\mathcal{T}\rvert = 100$ thresholds in $[10^{-4}, 1]$
Stability selection	$B = 50$ subsamples; subsampling fraction 0.8; $\pi_{\mathrm{thr}} = 0.95$
PDE selection score	$\alpha = 0.15$

constraint construction, and sparse equation identification. The shared settings are summarized in Table 6.

For local integral aggregation, we use the separable polynomial test function

$$\varphi_j(\boldsymbol{\eta}) = \prod_{r=1}^{D} \left(1 - \eta_r^2\right)^{\nu_r}, \qquad \boldsymbol{\eta} \in [-1, 1]^D, \tag{55}$$

where ν_r denotes the polynomial exponent along the rth spatiotemporal coordinate. The test function is nonnegative and symmetric within the local window, reaches its maximum at the window center, and smoothly decreases to zero as it approaches the boundary. Therefore, it assigns greater weights to the interior region and smaller weights to locations near the boundary when aggregating integrals. This weighting reduces the contribution of boundary regions, where local field and derivative estimates are more susceptible to finite-support effects, while emphasizing the more stable information in the window interior.

For STRidge, we use a common logarithmically spaced sparsity-threshold set

$$\mathcal{T} = \left\{ 10^{-4 + \frac{4(k-1)}{99}} \right\}_{k=1}^{100}, \tag{56}$$

which contains 100 candidate thresholds spanning $[10^{-4}, 1]$.

E.4 Evaluation Metrics

We evaluate the discovered equations from both structural and coefficient perspectives. Let $\mathcal{S}^\star$ and $\widehat{\mathcal{S}}$ denote the true and estimated supports, respectively. We define $\mathrm{TP} = |\mathcal{S}^\star \cap \widehat{\mathcal{S}}|$, $\mathrm{FP} = |\widehat{\mathcal{S}} \setminus \mathcal{S}^\star|$, and $\mathrm{FN} = |\mathcal{S}^\star \setminus \widehat{\mathcal{S}}|$. The structural precision, recall, and F1 score are then computed as

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \qquad \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \qquad \mathrm{F1} = \frac{2\,\mathrm{Precision}\,\mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \tag{57}$$

Undefined ratios arising from an empty estimated support are set to zero. A candidate term is regarded as active when the absolute value of its coefficient is at least 10^{-4} in physical coefficient units.

To evaluate coefficient recovery, we use the true-term coefficient error

$$E_{\mathrm{true}} = \frac{1}{|\mathcal{S}^\star|} \sum_{j \in \mathcal{S}^\star} \frac{|\widehat{\xi}_j - \xi_j^\star|}{|\xi_j^\star|} \times 100\%. \tag{58}$$

We further evaluate the complete discovered equation using

$$E_{\mathrm{full}} = \frac{\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}^\star\|_2}{\|\boldsymbol{\xi}^\star\|_2} \times 100\%. \tag{59}$$

where the coefficient vectors are aligned over the union of the true and estimated supports, with absent coefficients filled with zero. Unlike E_{true}, which evaluates coefficient recovery on the true terms, E_{full} additionally penalizes spurious terms appearing in the discovered equation. All coefficient errors are dimensionless.

For coupled PDE systems, a support element is defined by an output–term pair, and the coefficient vectors of all state equations are concatenated before computing the above metrics. An equation is considered exactly recovered only when its complete estimated support is identical to the ground-truth support. For a collection of K evaluated tasks, the success rate is defined as

$$\text{Success Rate} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{1}\left[\widehat{\mathcal{S}}_k = \mathcal{S}_k^\star\right]. \tag{60}$$

To evaluate the ability of AGRE to distinguish observations with different perturbation levels, we additionally report the area under the receiver operating characteristic curve (AUC). Let $\mathcal{H}$ and $\mathcal{L}$ denote the high- and low-perturbation observation sets, respectively, and define the unreliability score as $s_i = 1 - a_i$. The AUC can be equivalently written as

$$\mathrm{AUC} = \frac{1}{|\mathcal{H}||\mathcal{L}|} \sum_{i \in \mathcal{H}} \sum_{j \in \mathcal{L}} \left[\mathbf{1}(s_i > s_j) + \frac{1}{2}\mathbf{1}(s_i = s_j)\right]. \tag{61}$$

A larger AUC indicates better separation between high- and low-perturbation observations, with $\mathrm{AUC} = 0.5$ corresponding to random discrimination and $\mathrm{AUC} = 1$ indicating perfect separation. AUC is undefined when only one perturbation group is present.

F Additional Experimental Results

F.1 Experimental Results and Analysis

Building on the experiments reported in Section 4.4, we further evaluate CARE-PDE and the compared methods under a broader range of NDEC scenarios. In addition to the results presented in the main text, we report the complete evaluation metrics, including E_{true}, E_{full}, Precision, Recall, and F1, for all PDE systems. The detailed results are presented in Tables 7–12. From these results, we draw the following key observations:

- CARE-PDE maintains more stable structural recovery across different PDE systems and NDEC conditions, whereas the compared methods are more likely to introduce missing or spurious terms as heterogeneous contamination becomes more challenging. This robustness can be attributed to the reliability-aware design throughout the discovery process, which reduces the influence of highly perturbed observations on field reconstruction and subsequent equation identification.
- CARE-PDE generally exhibits lower coefficient errors, indicating that its advantage extends beyond structural identification to more accurate coefficient estimation. This behavior is closely related to the combination of RGLP and LICA. RGLP obtains more stable differential information through local orthogonal projection, while LICA aggregates local strong-form relations to reduce sensitivity to pointwise estimation errors. Together, they provide more stable equation constraints for subsequent sparse regression.

- CARE-PDE remains robust as the difference between low- and high-perturbation observations becomes more pronounced, and even achieves improved performance in some cases. A larger separation between the two perturbation levels can yield more distinct residual distributions, enabling AGRE to more effectively differentiate observations with different reliability levels and to provide more informative reliability estimates for subsequent stages. When the perturbation levels are less distinguishable, reliability estimation becomes more challenging, but the remaining components of CARE-PDE continue to mitigate the influence of unreliable local information.

Overall, these results demonstrate that CARE-PDE consistently outperforms the compared methods in both equation-structure identification and coefficient estimation, while maintaining robust performance across a broader range of heterogeneous contamination conditions.

Table 7: Complete comparison results of CARE-PDE and ten compared methods on Advection and Conv.–diff.. Each system reports E_{true} (%), E_{full} (%), precision (P), recall (R), and support F1. DIV. denotes an extremely large but finite coefficient error.

Method	Advection					Conv.–diff.				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 0\%)$										
PINN-SR (Nat. Commun., 2021)	100.00	100.00	0.00	0.00	0.00	26.020	91.771	0.20	1.00	0.33
DeepMoD (JCP, 2021)	100.00	100.07	0.00	0.00	0.00	0.6588	0.6214	1.00	1.00	1.00
R-DLGA (Phys. Rev. Res., 2021)	0.0504	0.0504	1.00	1.00	1.00	**0.1783**	**0.2783**	**1.00**	**1.00**	**1.00**
PDE-READ (Neural Netw., 2022)	DIV.	DIV.	0.11	1.00	0.20	8.9796	4.9487	1.00	1.00	1.00
PIC (Research, 2023)	0.9878	0.9878	1.00	1.00	1.00	2.8825	4.4999	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	0.0639	0.0639	1.00	1.00	1.00	1.2278	2.0610	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	100.00	100.35	0.00	0.00	0.00	0.3872	0.2094	1.00	1.00	1.00
EqGPT (Nat. Commun., 2025)	**0.0332**	**0.0332**	**1.00**	**1.00**	**1.00**	10.094	4.7586	1.00	1.00	1.00
ABL-PDE (NIPS, 2025)	0.4468	0.4468	1.00	1.00	1.00	0.5899	0.8175	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	0.1720	0.1720	1.00	1.00	1.00	1.3443	0.6521	1.00	1.00	1.00
CARE-PDE	**0.0037**	**0.0037**	**1.00**	**1.00**	**1.00**	**0.0331**	**0.0157**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 25\%)$										
PINN-SR (Nat. Commun., 2021)	100.00	DIV.	0.00	0.00	0.00	21.351	101.28	0.25	1.00	0.40
DeepMoD (JCP, 2021)	100.00	100.00	0.00	0.00	0.00	**0.3268**	**0.2437**	**1.00**	**1.00**	**1.00**
R-DLGA (Phys. Rev. Res., 2021)	**0.0054**	**0.0054**	**1.00**	**1.00**	**1.00**	0.5129	0.2627	1.00	1.00	1.00
PDE-READ (Neural Netw., 2022)	100.00	551.91	0.00	0.00	0.00	8.7915	DIV.	0.40	1.00	0.57
PIC (Research, 2023)	0.1102	0.1102	1.00	1.00	1.00	2.3061	2.7994	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	0.0800	0.0800	1.00	1.00	1.00	2.0889	3.6094	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	100.00	100.35	0.00	0.00	0.00	0.5806	0.6018	1.00	1.00	1.00
EqGPT (Nat. Commun., 2025)	0.3349	0.3349	1.00	1.00	1.00	0.8832	0.5861	1.00	1.00	1.00
ABL-PDE (NIPS, 2025)	0.1198	0.1198	1.00	1.00	1.00	0.6376	0.5153	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	0.2539	0.2539	1.00	1.00	1.00	0.8092	0.3925	1.00	1.00	1.00
CARE-PDE	**0.0038**	**0.0038**	**1.00**	**1.00**	**1.00**	**0.1165**	**0.1084**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	100.00	100.01	0.00	0.00	0.00	24.579	113.62	0.20	1.00	0.33
DeepMoD (JCP, 2021)	100.00	100.00	0.00	0.00	0.00	**0.2084**	**0.3356**	**1.00**	**1.00**	**1.00**
R-DLGA (Phys. Rev. Res., 2021)	0.0769	0.0769	1.00	1.00	1.00	0.3427	0.1625	1.00	1.00	1.00
PDE-READ (Neural Netw., 2022)	DIV.	DIV.	0.11	1.00	0.20	9.3903	5.1156	1.00	1.00	1.00
PIC (Research, 2023)	0.0640	0.0640	1.00	1.00	1.00	2.6201	3.5729	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	**0.0263**	**0.0263**	**1.00**	**1.00**	**1.00**	2.0163	3.9306	0.67	1.00	0.80
Weak-PDE-LEARN (JCP, 2024)	100.00	100.29	0.00	0.00	0.00	0.7608	1.0273	1.00	1.00	1.00
EqGPT (Nat. Commun., 2025)	0.2323	0.2323	1.00	1.00	1.00	1.1017	0.5370	1.00	1.00	1.00
ABL-PDE (NIPS, 2025)	0.1532	0.1532	1.00	1.00	1.00	0.6071	0.5358	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	0.2489	0.2489	1.00	1.00	1.00	0.2616	0.1269	1.00	1.00	1.00

Continued on next page

Table 7 continued

Method	Advection					Conv.–diff.				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
CARE-PDE	**0.0058**	**0.0058**	**1.00**	**1.00**	**1.00**	**0.1734**	**0.1163**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 25\%)$										
PINN-SR (Nat. Commun., 2021)	100.00	606.17	0.00	0.00	0.00	20.580	81.310	0.20	1.00	0.33
DeepMoD (JCP, 2021)	100.00	1904.8	0.00	0.00	0.00	0.2262	0.4275	1.00	1.00	1.00
R-DLGA (Phys. Rev. Res., 2021)	0.0237	0.0237	1.00	1.00	1.00	0.7025	0.8923	1.00	1.00	1.00
PDE-READ (Neural Netw., 2022)	100.00	DIV.	0.00	0.00	0.00	9.4942	4.9948	1.00	1.00	1.00
PIC (Research, 2023)	0.1673	0.1673	1.00	1.00	1.00	1.9600	2.8833	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	0.2310	0.2310	1.00	1.00	1.00	4.4018	7.2548	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	100.00	100.38	0.00	0.00	0.00	0.8114	1.0095	1.00	1.00	1.00
EqGPT (Nat. Commun., 2025)	0.1648	0.1648	1.00	1.00	1.00	0.3865	0.2914	1.00	1.00	1.00
ABL-PDE (NIPS, 2025)	0.0976	0.0976	1.00	1.00	1.00	1.1167	0.5309	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	**0.0138**	**0.0138**	**1.00**	**1.00**	**1.00**	**0.1834**	**0.0890**	**1.00**	**1.00**	**1.00**
CARE-PDE	**0.0001**	**0.0001**	**1.00**	**1.00**	**1.00**	**0.1026**	**0.0953**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	100.00	100.00	0.00	0.00	0.00	25.375	112.33	0.33	1.00	0.50
DeepMoD (JCP, 2021)	100.00	100.00	0.00	0.00	0.00	**0.2832**	**0.3254**	**1.00**	**1.00**	**1.00**
R-DLGA (Phys. Rev. Res., 2021)	0.1654	0.1654	1.00	1.00	1.00	0.7641	0.5571	1.00	1.00	1.00
PDE-READ (Neural Netw., 2022)	100.00	DIV.	0.00	0.00	0.00	11.604	1141.1	0.50	1.00	0.67
PIC (Research, 2023)	32.026	32.026	1.00	1.00	1.00	2.2380	3.1872	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	**0.0380**	**0.0380**	**1.00**	**1.00**	**1.00**	2.9344	3.8280	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	100.00	100.39	0.00	0.00	0.00	1.1611	2.0904	1.00	1.00	1.00
EqGPT (Nat. Commun., 2025)	500.00	170.06	0.50	0.50	0.50	0.7856	0.4084	1.00	1.00	1.00
ABL-PDE (NIPS, 2025)	0.1159	0.1159	1.00	1.00	1.00	1.2005	0.6433	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	0.3076	0.3076	1.00	1.00	1.00	0.8173	0.3965	1.00	1.00	1.00
CARE-PDE	**0.0011**	**0.0011**	**1.00**	**1.00**	**1.00**	**0.1486**	**0.1210**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (5\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	100.00	100.03	0.00	0.00	0.00	12.085	86.279	0.18	1.00	0.31
DeepMoD (JCP, 2021)	100.00	100.00	0.00	0.00	0.00	2.2522	1.0903	1.00	1.00	1.00
R-DLGA (Phys. Rev. Res., 2021)	0.2149	0.2149	1.00	1.00	1.00	3.1951	1.5041	1.00	1.00	1.00
PDE-READ (Neural Netw., 2022)	DIV.	DIV.	0.11	1.00	0.20	8.7231	4.7580	1.00	1.00	1.00
PIC (Research, 2023)	0.8941	0.8941	1.00	1.00	1.00	3.2792	4.3768	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	**0.0691**	**0.0691**	**1.00**	**1.00**	**1.00**	2.5301	1.9898	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	100.00	100.31	0.00	0.00	0.00	2.0282	2.3509	1.00	1.00	1.00
EqGPT (Nat. Commun., 2025)	0.0788	0.0788	1.00	1.00	1.00	2.6690	4.8858	1.00	1.00	1.00
ABL-PDE (NIPS, 2025)	0.1828	0.1828	1.00	1.00	1.00	**1.9756**	**0.9421**	**1.00**	**1.00**	**1.00**
SI-SR (Phys. Fluids, 2026)	0.1215	0.1215	1.00	1.00	1.00	2.1767	1.0559	1.00	1.00	1.00
CARE-PDE	**0.0476**	**0.0476**	**1.00**	**1.00**	**1.00**	**1.9627**	**1.1780**	**1.00**	**1.00**	**1.00**

Table 8: Complete comparison results of CARE-PDE and ten compared methods on Burgers and KdV. Each system reports E_{true} (%), E_{full} (%), precision (P), recall (R), and support F1.

Method	Burgers					KdV				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 0\%)$										
PINN-SR (Nat. Commun., 2021)	5.8027	1.8701	1.00	1.00	1.00	132.42	100.64	0.17	0.50	0.25
DeepMoD (JCP, 2021)	0.5778	0.7481	1.00	1.00	1.00	343.15	100.18	0.25	0.50	0.33
R-DLGA (Phys. Rev. Res., 2021)	0.0938	0.4660	0.40	1.00	0.57	1.5828	2.0668	1.00	1.00	1.00

Continued on next page

Table 8 continued

Method	Burgers					KdV				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
PDE-READ (Neural Netw., 2022)	12.061	15.428	1.00	1.00	1.00	60.381	58.425	1.00	1.00	1.00
PIC (Research, 2023)	1.6464	3.0858	1.00	1.00	1.00	0.5385	0.5190	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	0.9165	0.6313	1.00	1.00	1.00	53.470	93.787	0.22	1.00	0.36
Weak-PDE-LEARN (JCP, 2024)	0.3008	0.2426	1.00	1.00	1.00	43.574	65.858	0.67	1.00	0.80
EqGPT (Nat. Commun., 2025)	0.3491	0.2828	1.00	1.00	1.00	1.4411	0.1515	1.00	1.00	1.00
ABL-PDE (NIPS, 2025)	1.8810	0.6040	0.67	1.00	0.80	**0.3462**	**0.2984**	**1.00**	**1.00**	**1.00**
SI-SR (Phys. Fluids, 2026)	**0.2224**	**0.0443**	**1.00**	**1.00**	**1.00**	100.00	100.03	0.00	0.00	0.00
CARE-PDE	**0.0514**	**0.0121**	**1.00**	**1.00**	**1.00**	**0.0872**	**0.0977**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 25\%)$										
PINN-SR (Nat. Commun., 2021)	2.3140	1.5165	1.00	1.00	1.00	100.00	101.18	0.00	0.00	0.00
DeepMoD (JCP, 2021)	2.9042	1.1844	0.67	1.00	0.80	555.08	100.07	0.33	0.50	0.40
R-DLGA (Phys. Rev. Res., 2021)	2.6219	2.2043	0.33	1.00	0.50	27.294	3.5356	1.00	1.00	1.00
PDE-READ (Neural Netw., 2022)	14.811	15.850	1.00	1.00	1.00	60.393	58.516	1.00	1.00	1.00
PIC (Research, 2023)	2.8347	2.0461	1.00	1.00	1.00	**0.4339**	**0.5614**	**1.00**	**1.00**	**1.00**
R-DISCOVER (Phys. Fluids, 2024)	1.4933	1.0313	1.00	1.00	1.00	93.811	135.63	0.14	0.50	0.22
Weak-PDE-LEARN (JCP, 2024)	1.0918	1.7833	1.00	1.00	1.00	54.645	80.994	0.50	1.00	0.67
EqGPT (Nat. Commun., 2025)	**0.0777**	**0.0227**	**1.00**	**1.00**	**1.00**	0.8927	1.1138	1.00	1.00	1.00
ABL-PDE (NIPS, 2025)	2.6956	0.9041	0.67	1.00	0.80	0.9954	0.9808	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	5.0745	6.5860	1.00	1.00	1.00	100.00	100.00	0.00	0.00	0.00
CARE-PDE	**0.0647**	**0.0418**	**1.00**	**1.00**	**1.00**	**0.0897**	**0.0761**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	1.6205	2.1146	1.00	1.00	1.00	100.00	101.49	0.00	0.00	0.00
DeepMoD (JCP, 2021)	2.5242	1.3200	0.67	1.00	0.80	100.00	100.16	0.00	0.00	0.00
R-DLGA (Phys. Rev. Res., 2021)	54.422	14.715	0.17	0.50	0.25	3.0872	2.2626	0.50	1.00	0.67
PDE-READ (Neural Netw., 2022)	15.947	16.471	1.00	1.00	1.00	59.054	57.184	1.00	1.00	1.00
PIC (Research, 2023)	3.0814	2.8178	1.00	1.00	1.00	0.4834	0.4640	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	1.3606	1.1051	1.00	1.00	1.00	100.00	151.20	0.00	0.00	0.00
Weak-PDE-LEARN (JCP, 2024)	1.9214	2.4752	1.00	1.00	1.00	50.769	75.550	0.50	1.00	0.67
EqGPT (Nat. Commun., 2025)	**0.8731**	**0.7292**	**1.00**	**1.00**	**1.00**	**0.4536**	**0.2663**	**1.00**	**1.00**	**1.00**
ABL-PDE (NIPS, 2025)	0.9613	8.9364	0.50	1.00	0.67	4.3523	5.6402	0.40	1.00	0.57
SI-SR (Phys. Fluids, 2026)	100.00	100.00	0.00	0.00	0.00	13.124	12.593	1.00	1.00	1.00
CARE-PDE	**0.0304**	**0.0235**	**1.00**	**1.00**	**1.00**	**0.2176**	**0.2048**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 25\%)$										
PINN-SR (Nat. Commun., 2021)	6.1682	3.2043	1.00	1.00	1.00	100.00	111.59	0.00	0.00	0.00
DeepMoD (JCP, 2021)	2.2342	1.5184	0.67	1.00	0.80	100.00	100.14	0.00	0.00	0.00
R-DLGA (Phys. Rev. Res., 2021)	56.296	17.306	0.33	0.50	0.40	26.921	0.6305	0.50	1.00	0.67
PDE-READ (Neural Netw., 2022)	14.758	16.579	1.00	1.00	1.00	60.717	58.770	1.00	1.00	1.00
PIC (Research, 2023)	5.8957	2.7332	1.00	1.00	1.00	**0.6626**	**0.8380**	**1.00**	**1.00**	**1.00**
R-DISCOVER (Phys. Fluids, 2024)	2.0889	1.6986	1.00	1.00	1.00	100.00	148.74	0.00	0.00	0.00
Weak-PDE-LEARN (JCP, 2024)	1.9233	3.2756	0.67	1.00	0.80	50.226	74.625	0.50	1.00	0.67
EqGPT (Nat. Commun., 2025)	**0.7992**	**0.7413**	**1.00**	**1.00**	**1.00**	100.00	229.82	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	4.6753	2.0883	0.67	1.00	0.80	2.5436	1.4696	0.67	1.00	0.80
SI-SR (Phys. Fluids, 2026)	100.00	100.00	0.00	0.00	0.00	20.908	19.867	1.00	1.00	1.00
CARE-PDE	**0.0485**	**0.0341**	**1.00**	**1.00**	**1.00**	**0.0274**	**0.0095**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	13.904	4.1904	1.00	1.00	1.00	74.631	101.48	0.10	0.50	0.17
DeepMoD (JCP, 2021)	1.8570	1.6931	0.67	1.00	0.80	696.06	106.44	0.25	0.50	0.33
R-DLGA (Phys. Rev. Res., 2021)	60.432	25.804	0.11	0.50	0.18	7.8838	12.792	0.29	1.00	0.44
PDE-READ (Neural Netw., 2022)	15.076	16.683	1.00	1.00	1.00	62.628	60.731	1.00	1.00	1.00

Continued on next page

Table 8 continued

Method	Burgers					KdV				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
PIC (Research, 2023)	**6.8981**	**5.4871**	**1.00**	**1.00**	**1.00**	**0.9990**	**0.8543**	**1.00**	**1.00**	**1.00**
R-DISCOVER (Phys. Fluids, 2024)	2.9269	2.3903	0.50	1.00	0.67	100.00	156.35	0.00	0.00	0.00
Weak-PDE-LEARN (JCP, 2024)	9.3428	8.9996	0.67	1.00	0.80	48.335	73.596	0.67	1.00	0.80
EqGPT (Nat. Commun., 2025)	100.00	100.00	0.00	0.00	0.00	100.00	226.91	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	1.0657	16.689	0.50	1.00	0.67	6.5856	8.3284	0.40	1.00	0.57
SI-SR (Phys. Fluids, 2026)	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
CARE-PDE	**0.0766**	**0.0408**	**1.00**	**1.00**	**1.00**	**0.1440**	**0.0969**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (5\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	4.3420	1.7479	1.00	1.00	1.00	100.00	101.16	0.00	0.00	0.00
DeepMoD (JCP, 2021)	3.3700	1.5952	0.67	1.00	0.80	625.77	100.29	0.33	0.50	0.40
R-DLGA (Phys. Rev. Res., 2021)	56.902	23.396	0.13	0.50	0.20	2.9789	5.2959	1.00	1.00	1.00
PDE-READ (Neural Netw., 2022)	11.381	15.159	1.00	1.00	1.00	60.242	58.023	1.00	1.00	1.00
PIC (Research, 2023)	7.6832	8.0517	1.00	1.00	1.00	**1.2894**	**0.0990**	**1.00**	**1.00**	**1.00**
R-DISCOVER (Phys. Fluids, 2024)	2.3555	3.1618	0.33	1.00	0.50	100.00	129.27	0.00	0.00	0.00
Weak-PDE-LEARN (JCP, 2024)	6.5351	5.7065	0.50	1.00	0.67	60.608	85.278	0.29	1.00	0.44
EqGPT (Nat. Commun., 2025)	**1.8853**	**3.6824**	**1.00**	**1.00**	**1.00**	5.7016	1.2962	1.00	1.00	1.00
ABL-PDE (NIPS, 2025)	3.5375	3.9205	0.50	1.00	0.67	1.3655	1.4394	0.67	1.00	0.80
SI-SR (Phys. Fluids, 2026)	8.1852	8.8362	1.00	1.00	1.00	1.8238	0.0091	1.00	1.00	1.00
CARE-PDE	**0.9920**	**0.4603**	**1.00**	**1.00**	**1.00**	**0.7479**	**0.6931**	**1.00**	**1.00**	**1.00**

Table 9: Complete comparison results of CARE-PDE and ten compared methods on KS and Allen–Cahn. Each system reports E_{true} (%), E_{full} (%), precision (P), recall (R), and support F1. DIV. denotes an extremely large but finite coefficient error.

Method	KS					Allen–Cahn				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 0\%)$										
PINN-SR (Nat. Commun., 2021)	100.00	100.00	0.00	0.00	0.00	991.29	100.84	0.17	0.33	0.22
DeepMoD (JCP, 2021)	**21.140**	**21.159**	**0.75**	**1.00**	**0.86**	3780.1	275.96	0.40	0.67	0.50
R-DLGA (Phys. Rev. Res., 2021)	77.479	83.633	0.50	0.67	0.57	7.0049	8.4446	1.00	1.00	1.00
PDE-READ (Neural Netw., 2022)	96.976	97.070	1.00	0.33	0.50	30.180	32.221	1.00	1.00	1.00
PIC (Research, 2023)	100.00	100.00	0.00	0.00	0.00	6.4541	1.3974	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	93.734	94.824	0.33	0.33	0.33	7.6550	7.5031	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	115.32	122.58	0.50	0.33	0.40	176.55	83.028	0.50	0.67	0.57
EqGPT (Nat. Commun., 2025)	100.00	731.34	0.00	0.00	0.00	**0.7991**	**0.8058**	**1.00**	**1.00**	**1.00**
ABL-PDE (NIPS, 2025)	89.729	95.664	0.25	0.33	0.29	3.9697	3.9152	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	100.00	100.00	0.00	0.00	0.00	100.00	100.01	0.00	0.00	0.00
CARE-PDE	**2.3265**	**2.4772**	**1.00**	**1.00**	**1.00**	**0.0311**	**0.0402**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 25\%)$										
PINN-SR (Nat. Commun., 2021)	98.534	98.556	1.00	0.33	0.50	279.24	102.14	0.10	0.33	0.15
DeepMoD (JCP, 2021)	96.456	98.387	0.13	0.33	0.18	7219.6	264.46	0.40	0.67	0.50
R-DLGA (Phys. Rev. Res., 2021)	90.141	92.414	0.50	0.33	0.40	80.733	91.713	1.00	0.67	0.80
PDE-READ (Neural Netw., 2022)	**96.778**	**96.885**	**1.00**	**0.33**	**0.50**	28.260	30.670	1.00	1.00	1.00
PIC (Research, 2023)	100.00	100.01	0.00	0.00	0.00	**4.5413**	**5.0877**	**1.00**	**1.00**	**1.00**
R-DISCOVER (Phys. Fluids, 2024)	100.00	100.00	0.00	0.00	0.00	10.895	11.195	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	99.000	119.76	0.25	1.00	0.40	134.84	80.928	0.40	0.67	0.50
EqGPT (Nat. Commun., 2025)	100.00	489.91	0.00	0.00	0.00	100.00	115.47	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	97.454	97.521	1.00	0.33	0.50	10.891	11.170	1.00	1.00	1.00

Continued on next page

Table 9 continued

Method	KS					Allen–Cahn				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
SI-SR (Phys. Fluids, 2026)	100.00	100.00	0.00	0.00	0.00	100.00	100.01	0.00	0.00	0.00
CARE-PDE	**1.0132**	**1.0168**	**1.00**	**1.00**	**1.00**	**0.0185**	**0.0241**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
DeepMoD (JCP, 2021)	94.699	95.093	0.20	0.67	0.31	9622.2	304.90	0.40	0.67	0.50
R-DLGA (Phys. Rev. Res., 2021)	97.098	97.185	0.50	0.33	0.40	80.823	91.730	1.00	0.67	0.80
PDE-READ (Neural Netw., 2022)	**97.144**	**97.228**	**1.00**	**0.33**	**0.50**	31.446	33.192	1.00	1.00	1.00
PIC (Research, 2023)	100.00	100.01	0.00	0.00	0.00	**4.6892**	**4.5672**	**1.00**	**1.00**	**1.00**
R-DISCOVER (Phys. Fluids, 2024)	100.00	100.01	0.00	0.00	0.00	13.705	13.782	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	128.13	130.44	0.29	0.67	0.40	134.33	81.021	0.40	0.67	0.50
EqGPT (Nat. Commun., 2025)	100.00	100.00	0.00	0.00	0.00	123.57	135.00	0.75	0.75	0.75
ABL-PDE (NIPS, 2025)	97.520	97.583	1.00	0.33	0.50	9.9825	9.9059	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	100.00	100.00	0.00	0.00	0.00	100.00	100.01	0.00	0.00	0.00
CARE-PDE	**7.3953**	**7.4444**	**1.00**	**1.00**	**1.00**	**0.0312**	**0.0366**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 25\%)$										
PINN-SR (Nat. Commun., 2021)	95.744	98.669	0.20	0.33	0.25	100.00	100.00	0.00	0.00	0.00
DeepMoD (JCP, 2021)	96.860	97.509	0.29	0.67	0.40	DIV.	329.00	0.40	0.67	0.50
R-DLGA (Phys. Rev. Res., 2021)	97.171	97.254	0.50	0.33	0.40	80.632	91.697	1.00	0.67	0.80
PDE-READ (Neural Netw., 2022)	97.265	97.342	1.00	0.33	0.50	31.433	34.208	1.00	1.00	1.00
PIC (Research, 2023)	100.00	100.01	0.00	0.00	0.00	**1.0350**	**0.6527**	**1.00**	**1.00**	**1.00**
R-DISCOVER (Phys. Fluids, 2024)	97.663	97.719	1.00	0.33	0.50	20.166	20.583	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	**121.49**	**122.91**	**0.50**	**1.00**	**0.67**	131.58	81.299	0.40	0.67	0.50
EqGPT (Nat. Commun., 2025)	100.00	DIV.	0.00	0.00	0.00	100.00	115.47	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	97.602	97.661	1.00	0.33	0.50	28.068	28.408	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	100.00	100.00	0.00	0.00	0.00	100.00	100.01	0.00	0.00	0.00
CARE-PDE	**1.0893**	**1.2260**	**1.00**	**1.00**	**1.00**	**0.0193**	**0.0163**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	100.00	101.43	0.00	0.00	0.00	1429.4	101.31	0.17	0.33	0.22
DeepMoD (JCP, 2021)	80.232	83.568	0.25	0.67	0.36	1223.5	104.49	0.17	0.33	0.22
R-DLGA (Phys. Rev. Res., 2021)	97.299	97.374	1.00	0.33	0.50	80.876	91.746	1.00	0.67	0.80
PDE-READ (Neural Netw., 2022)	**97.185**	**97.266**	**1.00**	**0.33**	**0.50**	**30.503**	**32.112**	**1.00**	**1.00**	**1.00**
PIC (Research, 2023)	100.00	100.00	0.00	0.00	0.00	97.762	96.702	1.00	0.33	0.50
R-DISCOVER (Phys. Fluids, 2024)	100.00	100.00	0.00	0.00	0.00	96.792	238.73	0.22	0.67	0.33
Weak-PDE-LEARN (JCP, 2024)	106.09	119.68	0.23	1.00	0.38	128.65	82.123	0.40	0.67	0.50
EqGPT (Nat. Commun., 2025)	100.00	141.42	0.00	0.00	0.00	100.00	173.20	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	97.880	97.925	1.00	0.33	0.50	38.798	38.448	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	100.00	100.00	0.00	0.00	0.00	100.00	100.01	0.00	0.00	0.00
CARE-PDE	**13.859**	**13.891**	**1.00**	**1.00**	**1.00**	**0.0221**	**0.0284**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (5\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	**13.566**	**13.611**	**1.00**	**1.00**	**1.00**	96.945	95.528	1.00	0.33	0.50
DeepMoD (JCP, 2021)	98.575	99.155	0.25	0.67	0.36	324.07	158.72	0.50	0.67	0.57
R-DLGA (Phys. Rev. Res., 2021)	97.120	97.207	0.50	0.33	0.40	81.075	91.768	1.00	0.67	0.80
PDE-READ (Neural Netw., 2022)	97.033	97.123	1.00	0.33	0.50	30.267	31.660	1.00	1.00	1.00
PIC (Research, 2023)	100.00	100.01	0.00	0.00	0.00	**4.8244**	**6.2245**	**1.00**	**1.00**	**1.00**
R-DISCOVER (Phys. Fluids, 2024)	100.00	100.01	0.00	0.00	0.00	79.360	77.492	0.75	1.00	0.86
Weak-PDE-LEARN (JCP, 2024)	101.26	118.37	0.43	1.00	0.60	118.44	80.417	0.40	0.67	0.50
EqGPT (Nat. Commun., 2025)	100.00	111.94	0.00	0.00	0.00	100.00	DIV.	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	97.454	97.521	1.00	0.33	0.50	13.654	12.950	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	100.00	100.00	0.00	0.00	0.00	100.00	100.01	0.00	0.00	0.00

Continued on next page

Table 9 continued

Method	KS					Allen–Cahn				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
CARE-PDE	**4.2859**	**4.4348**	**1.00**	**1.00**	**1.00**	**1.6595**	**1.4377**	**1.00**	**1.00**	**1.00**

Table 10: Complete comparison results of CARE-PDE and ten compared methods on Chaffee–Infante and Compound. Each system reports E_{true} (%), E_{full} (%), precision (P), recall (R), and support F1. DIV. denotes an extremely large but finite coefficient error.

Method	Chaffee–Infante					Compound				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
$(\gamma_{\text{L}},\gamma_{\text{H}},\beta)=(0\%,25\%,0\%)$										
PINN-SR (Nat. Commun., 2021)	105.78	130.72	0.25	0.67	0.36	87.791	777.32	0.09	0.50	0.15
DeepMoD (JCP, 2021)	105.81	110.16	0.25	0.33	0.29	189.72	282.33	0.33	0.50	0.40
R-DLGA (Phys. Rev. Res., 2021)	22.879	25.815	1.00	1.00	1.00	88.150	198.90	0.13	0.50	0.20
PDE-READ (Neural Netw., 2022)	90.537	91.520	1.00	0.33	0.50	164.46	712.06	0.25	0.50	0.33
PIC (Research, 2023)	10.318	11.393	1.00	1.00	1.00	57.849	71.618	0.50	0.50	0.50
R-DISCOVER (Phys. Fluids, 2024)	61.036	81.786	0.22	0.67	0.33	100.00	388.23	0.00	0.00	0.00
Weak-PDE-LEARN (JCP, 2024)	99.583	99.595	0.20	0.33	0.25	**0.9183**	**0.9183**	**1.00**	**1.00**	**1.00**
EqGPT (Nat. Commun., 2025)	0.2518	0.3658	1.00	1.00	1.00	1.6261	1.6261	1.00	1.00	1.00
ABL-PDE (NIPS, 2025)	71.333	90.243	0.14	0.33	0.20	67.226	351.67	0.33	1.00	0.50
SI-SR (Phys. Fluids, 2026)	**0.2252**	**0.3901**	**1.00**	**1.00**	**1.00**	100.00	376.29	0.00	0.00	0.00
CARE-PDE	**0.0347**	**0.0408**	**1.00**	**1.00**	**1.00**	**0.6492**	**0.7259**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}},\gamma_{\text{H}},\beta)=(0\%,25\%,25\%)$										
PINN-SR (Nat. Commun., 2021)	56.623	114.63	0.20	0.67	0.31	116.50	1025.3	0.20	0.50	0.29
DeepMoD (JCP, 2021)	100.00	103.20	0.00	0.00	0.00	189.82	285.67	0.33	0.50	0.40
R-DLGA (Phys. Rev. Res., 2021)	48.994	60.978	0.67	0.67	0.67	111.04	232.09	0.17	1.00	0.29
PDE-READ (Neural Netw., 2022)	90.561	91.540	1.00	0.33	0.50	70.313	76.323	1.00	0.50	0.67
PIC (Research, 2023)	16.830	19.038	1.00	1.00	1.00	58.595	71.793	0.50	0.50	0.50
R-DISCOVER (Phys. Fluids, 2024)	100.00	148.42	0.00	0.00	0.00	100.00	200.22	0.00	0.00	0.00
Weak-PDE-LEARN (JCP, 2024)	99.696	99.705	0.13	0.33	0.18	**1.4209**	**1.4209**	**1.00**	**1.00**	**1.00**
EqGPT (Nat. Commun., 2025)	100.00	156.67	0.00	0.00	0.00	100.00	151.24	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	67.315	84.467	0.25	0.67	0.36	88.497	472.10	0.14	0.50	0.22
SI-SR (Phys. Fluids, 2026)	**0.3598**	**0.6232**	**1.00**	**1.00**	**1.00**	100.00	376.19	0.00	0.00	0.00
CARE-PDE	**0.0879**	**0.1167**	**1.00**	**1.00**	**1.00**	**0.4219**	**0.5722**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}},\gamma_{\text{H}},\beta)=(0\%,25\%,50\%)$										
PINN-SR (Nat. Commun., 2021)	85.118	113.53	0.09	0.33	0.14	114.01	267.53	0.22	1.00	0.36
DeepMoD (JCP, 2021)	100.00	102.80	0.00	0.00	0.00	190.08	283.52	0.33	0.50	0.40
R-DLGA (Phys. Rev. Res., 2021)	23.351	32.756	0.60	1.00	0.75	85.535	311.73	0.14	0.50	0.22
PDE-READ (Neural Netw., 2022)	90.527	91.513	1.00	0.33	0.50	100.00	353.18	0.00	0.00	0.00
PIC (Research, 2023)	47.383	60.400	1.00	0.67	0.80	60.985	72.425	0.50	0.50	0.50
R-DISCOVER (Phys. Fluids, 2024)	85.812	182.42	0.40	0.67	0.50	81.736	4410.7	0.14	0.50	0.22
Weak-PDE-LEARN (JCP, 2024)	99.644	99.655	0.13	0.33	0.18	**1.3024**	**1.3024**	**1.00**	**1.00**	**1.00**
EqGPT (Nat. Commun., 2025)	100.00	157.07	0.00	0.00	0.00	100.00	149.92	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	73.373	92.137	0.14	0.33	0.20	78.777	241.14	0.20	0.50	0.29
SI-SR (Phys. Fluids, 2026)	**0.2835**	**0.4910**	**1.00**	**1.00**	**1.00**	100.00	375.99	0.00	0.00	0.00
CARE-PDE	**0.2750**	**0.3146**	**1.00**	**1.00**	**1.00**	**0.8906**	**0.9441**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}},\gamma_{\text{H}},\beta)=(0\%,50\%,25\%)$										
PINN-SR (Nat. Commun., 2021)	87.436	106.19	0.06	0.33	0.11	100.00	650.46	0.00	0.00	0.00
DeepMoD (JCP, 2021)	100.00	0.0000	0.00	0.00	0.00	190.83	286.56	0.33	0.50	0.40

Continued on next page

Table 10 continued

Method	Chaffee–Infante					Compound				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
R-DLGA (Phys. Rev. Res., 2021)	21.130	33.586	0.43	1.00	0.60	100.00	386.66	0.00	0.00	0.00
PDE-READ (Neural Netw., 2022)	90.517	91.505	1.00	0.33	0.50	100.00	116.53	0.00	0.00	0.00
PIC (Research, 2023)	79.136	85.567	0.50	0.33	0.40	59.083	71.885	0.50	0.50	0.50
R-DISCOVER (Phys. Fluids, 2024)	181.39	277.77	0.14	0.33	0.20	100.00	270.84	0.00	0.00	0.00
Weak-PDE-LEARN (JCP, 2024)	99.793	99.798	0.14	0.33	0.20	**1.4329**	**1.4329**	**1.00**	**1.00**	**1.00**
EqGPT (Nat. Commun., 2025)	100.00	157.93	0.00	0.00	0.00	100.00	151.24	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	147.67	151.50	1.00	1.00	1.00	93.734	850.25	0.17	0.50	0.25
SI-SR (Phys. Fluids, 2026)	**0.9834**	**1.7032**	**1.00**	**1.00**	**1.00**	100.00	375.90	0.00	0.00	0.00
CARE-PDE	**0.1216**	**0.1628**	**1.00**	**1.00**	**1.00**	**0.4229**	**0.5852**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	72.553	111.07	0.14	0.67	0.24	100.00	102.46	0.00	0.00	0.00
DeepMoD (JCP, 2021)	100.00	100.44	0.00	0.00	0.00	189.81	286.70	0.33	0.50	0.40
R-DLGA (Phys. Rev. Res., 2021)	100.00	103.99	0.00	0.00	0.00	100.00	372.75	0.00	0.00	0.00
PDE-READ (Neural Netw., 2022)	100.00	108.67	0.00	0.00	0.00	100.00	353.91	0.00	0.00	0.00
PIC (Research, 2023)	70.212	87.445	0.50	0.33	0.40	60.259	72.216	0.50	0.50	0.50
R-DISCOVER (Phys. Fluids, 2024)	90.836	157.58	0.33	0.67	0.44	83.722	1980.3	0.13	0.50	0.20
Weak-PDE-LEARN (JCP, 2024)	100.00	100.01	0.00	0.00	0.00	**0.6493**	**0.6493**	**1.00**	**1.00**	**1.00**
EqGPT (Nat. Commun., 2025)	100.00	157.98	0.00	0.00	0.00	100.00	164.29	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	64.498	78.978	0.25	0.67	0.36	83.954	5897.5	0.20	0.50	0.29
SI-SR (Phys. Fluids, 2026)	**45.033**	**70.148**	**0.67**	**0.67**	**0.67**	100.00	375.73	0.00	0.00	0.00
CARE-PDE	**0.2003**	**0.2652**	**1.00**	**1.00**	**1.00**	**0.5267**	**0.6555**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (5\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	96.252	112.91	0.10	0.33	0.15	100.00	174.85	0.00	0.00	0.00
DeepMoD (JCP, 2021)	100.00	109.01	0.00	0.00	0.00	191.23	290.42	0.33	0.50	0.40
R-DLGA (Phys. Rev. Res., 2021)	**24.504**	**26.999**	**1.00**	**1.00**	**1.00**	87.614	521.96	0.20	0.50	0.29
PDE-READ (Neural Netw., 2022)	90.544	91.526	1.00	0.33	0.50	100.00	8944.6	0.00	0.00	0.00
PIC (Research, 2023)	**48.450**	**60.736**	**1.00**	**0.67**	**0.80**	58.276	71.737	0.50	0.50	0.50
R-DISCOVER (Phys. Fluids, 2024)	100.00	106.10	0.00	0.00	0.00	96.382	DIV.	0.08	0.50	0.14
Weak-PDE-LEARN (JCP, 2024)	99.725	99.733	0.13	0.33	0.18	**3.2909**	**3.2909**	**1.00**	**1.00**	**1.00**
EqGPT (Nat. Commun., 2025)	100.00	157.62	0.00	0.00	0.00	100.00	150.95	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	73.172	90.084	0.13	0.33	0.18	73.068	623.70	0.17	0.50	0.25
SI-SR (Phys. Fluids, 2026)	97.374	149.49	0.50	0.67	0.57	100.00	375.83	0.00	0.00	0.00
CARE-PDE	41.886	80.209	0.67	0.67	0.67	**2.3212**	**2.4361**	**1.00**	**1.00**	**1.00**

Table 11: Complete comparison results of CARE-PDE and ten compared methods on Burgers2D and Adv.–diff. (2D). Each system reports E_{true} (%), E_{full} (%), precision (P), recall (R), and support F1. DIV. denotes an extremely large but finite coefficient error.

Method	Burgers2D					Adv.–diff. (2D)				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 0\%)$										
PINN-SR (Nat. Commun., 2021)	83.498	100.05	0.20	0.25	0.22	0.7859	0.7940	1.00	1.00	1.00
DeepMoD (JCP, 2021)	99.916	100.04	0.40	0.50	0.44	0.5710	0.5968	1.00	1.00	1.00
R-DLGA (Phys. Rev. Res., 2021)	3.7240	5.7652	1.00	1.00	1.00	**0.1493**	**0.2449**	**1.00**	**1.00**	**1.00**
PDE-READ (Neural Netw., 2022)	138.80	DIV.	0.14	0.75	0.24	0.7422	0.9660	1.00	1.00	1.00
PIC (Research, 2023)	81.414	99.998	1.00	0.25	0.40	0.4736	0.5157	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	2.0665	0.9510	1.00	1.00	1.00	0.2131	0.2657	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	51.754	100.10	0.50	0.50	0.50	53.744	71.879	0.33	0.75	0.46

Continued on next page

Table 11 continued

Method	Burgers (2D)					Adv.–diff. (2D)				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
EqGPT (Nat. Commun., 2025)	**0.6983**	**0.3537**	**1.00**	**1.00**	**1.00**	0.2391	0.2574	0.80	1.00	0.89
ABL-PDE (NIPS, 2025)	43.986	950.32	0.40	1.00	0.57	0.2149	0.2715	1.00	1.00	1.00
SI-SR (Phys. Fluids, 2026)	54.191	99.995	0.50	0.50	0.50	0.4245	0.3751	1.00	1.00	1.00
CARE-PDE	**0.5693**	**0.6960**	**1.00**	**1.00**	**1.00**	**0.0361**	**0.0379**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 25\%)$										
PINN-SR (Nat. Commun., 2021)	79.952	100.07	0.20	0.25	0.22	25.086	73.751	0.44	1.00	0.62
DeepMoD (JCP, 2021)	103.88	100.03	0.29	0.50	0.36	**0.3250**	**0.5420**	**1.00**	**1.00**	**1.00**
R-DLGA (Phys. Rev. Res., 2021)	4.5333	6.3504	1.00	1.00	1.00	79.084	106.42	0.25	0.25	0.25
PDE-READ (Neural Netw., 2022)	128.10	DIV.	0.12	0.75	0.21	57.201	255.17	0.13	0.50	0.20
PIC (Research, 2023)	81.444	99.998	1.00	0.25	0.40	1.8611	2.1536	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	**3.9965**	**3.0292**	**1.00**	**1.00**	**1.00**	0.8277	1.0669	1.00	1.00	1.00
Weak-PDE-LEARN (JCP, 2024)	51.323	100.10	0.50	0.50	0.50	54.543	72.504	0.60	0.75	0.67
EqGPT (Nat. Commun., 2025)	50.965	2.4091	1.00	1.00	1.00	100.00	145.87	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	54.010	553.37	0.30	0.75	0.43	2.7265	40.301	0.50	1.00	0.67
SI-SR (Phys. Fluids, 2026)	53.708	99.995	0.50	0.50	0.50	1.7052	1.4686	1.00	1.00	1.00
CARE-PDE	**0.7345**	**1.0170**	**1.00**	**1.00**	**1.00**	**0.0580**	**0.0538**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	63.189	100.06	0.50	0.50	0.50	50.838	151.23	0.29	0.50	0.36
DeepMoD (JCP, 2021)	81.634	100.07	0.33	0.50	0.40	62.268	84.229	0.50	0.50	0.50
R-DLGA (Phys. Rev. Res., 2021)	**3.7009**	**5.4001**	**1.00**	**1.00**	**1.00**	82.129	85.958	0.25	0.25	0.25
PDE-READ (Neural Netw., 2022)	100.00	271.63	0.00	0.00	0.00	59.148	144.93	0.20	0.50	0.29
PIC (Research, 2023)	54.179	99.995	1.00	0.50	0.67	2.2925	3.7311	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	6.6291	3.7893	0.80	1.00	0.89	**1.2188**	**1.4809**	**1.00**	**1.00**	**1.00**
Weak-PDE-LEARN (JCP, 2024)	51.955	100.12	0.50	0.50	0.50	54.925	72.915	0.60	0.75	0.67
EqGPT (Nat. Commun., 2025)	100.00	100.00	0.00	0.00	0.00	100.00	145.85	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	100.00	DIV.	0.00	0.00	0.00	6.5079	15.295	0.57	1.00	0.73
SI-SR (Phys. Fluids, 2026)	52.895	99.995	0.50	0.50	0.50	62.353	64.975	1.00	0.50	0.67
CARE-PDE	**1.6862**	**2.7276**	**1.00**	**1.00**	**1.00**	**0.3066**	**0.3552**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 25\%)$										
PINN-SR (Nat. Commun., 2021)	64.497	100.05	0.50	0.50	0.50	58.347	124.49	0.22	0.50	0.31
DeepMoD (JCP, 2021)	68.216	100.03	0.40	0.50	0.44	73.991	93.948	0.29	0.50	0.36
R-DLGA (Phys. Rev. Res., 2021)	**5.1076**	**7.7782**	**1.00**	**1.00**	**1.00**	82.203	97.598	0.50	0.25	0.33
PDE-READ (Neural Netw., 2022)	118.30	DIV.	0.15	1.00	0.26	100.00	189.77	0.00	0.00	0.00
PIC (Research, 2023)	100.00	100.03	0.00	0.00	0.00	81.532	84.458	1.00	0.25	0.40
R-DISCOVER (Phys. Fluids, 2024)	100.00	DIV.	0.00	0.00	0.00	**1.1646**	**1.5810**	**1.00**	**1.00**	**1.00**
Weak-PDE-LEARN (JCP, 2024)	52.634	100.10	0.50	0.50	0.50	55.874	73.844	0.60	0.75	0.67
EqGPT (Nat. Commun., 2025)	100.00	152.75	0.00	0.00	0.00	100.00	146.22	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	84.789	1037.4	0.18	0.50	0.27	18.245	37.910	0.50	1.00	0.67
SI-SR (Phys. Fluids, 2026)	54.576	99.995	0.50	0.50	0.50	75.454	289.37	0.20	0.25	0.22
CARE-PDE	**0.5390**	**0.6706**	**1.00**	**1.00**	**1.00**	**0.0575**	**0.0659**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	85.114	100.07	0.20	0.25	0.22	100.00	198.75	0.00	0.00	0.00
DeepMoD (JCP, 2021)	119.24	100.17	0.25	0.50	0.33	90.512	109.16	0.20	0.25	0.22
R-DLGA (Phys. Rev. Res., 2021)	27.799	70.864	0.75	0.75	0.75	79.923	101.00	0.20	0.25	0.22
PDE-READ (Neural Netw., 2022)	168.92	DIV.	0.13	0.75	0.21	100.00	186.46	0.00	0.00	0.00
PIC (Research, 2023)	81.776	99.998	1.00	0.25	0.40	81.332	84.383	1.00	0.25	0.40
R-DISCOVER (Phys. Fluids, 2024)	**15.245**	**83.159**	**0.80**	**1.00**	**0.89**	**6.5125**	**9.7190**	**0.80**	**1.00**	**0.89**
Weak-PDE-LEARN (JCP, 2024)	51.957	100.13	0.50	0.50	0.50	68.034	85.308	0.60	0.75	0.67
EqGPT (Nat. Commun., 2025)	100.00	152.75	0.00	0.00	0.00	100.00	147.12	0.00	0.00	0.00

Continued on next page

Table 11 continued

Method	Burgers (2D)					Adv.–diff. (2D)				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
ABL-PDE (NIPS, 2025)	63.043	6485.8	0.18	0.50	0.27	39.704	173.20	0.38	0.75	0.50
SI-SR (Phys. Fluids, 2026)	100.00	100.00	0.00	0.00	0.00	100.00	100.16	0.00	0.00	0.00
CARE-PDE	**1.4761**	**1.8718**	**1.00**	**1.00**	**1.00**	**0.1995**	**0.2783**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (5\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	80.319	100.07	0.20	0.25	0.22	55.881	170.21	0.17	0.50	0.25
DeepMoD (JCP, 2021)	93.282	100.04	0.50	0.50	0.50	61.523	84.128	0.40	0.50	0.44
R-DLGA (Phys. Rev. Res., 2021)	**3.6530**	**4.5149**	**1.00**	**1.00**	**1.00**	87.966	99.270	0.25	0.25	0.25
PDE-READ (Neural Netw., 2022)	100.11	DIV.	0.17	1.00	0.29	100.00	195.47	0.00	0.00	0.00
PIC (Research, 2023)	78.473	99.998	1.00	0.25	0.40	2.7554	3.7569	1.00	1.00	1.00
R-DISCOVER (Phys. Fluids, 2024)	6.3467	5.0404	1.00	1.00	1.00	**2.0569**	**2.4757**	**1.00**	**1.00**	**1.00**
Weak-PDE-LEARN (JCP, 2024)	53.708	100.12	0.50	0.50	0.50	57.776	75.053	0.60	0.75	0.67
EqGPT (Nat. Commun., 2025)	100.00	152.75	0.00	0.00	0.00	100.00	146.24	0.00	0.00	0.00
ABL-PDE (NIPS, 2025)	41.265	2135.3	0.44	1.00	0.62	6.5260	14.502	0.57	1.00	0.73
SI-SR (Phys. Fluids, 2026)	54.027	99.995	0.50	0.50	0.50	59.650	63.839	1.00	0.50	0.67
CARE-PDE	**2.2609**	**2.0870**	**1.00**	**1.00**	**1.00**	**0.6181**	**0.6627**	**1.00**	**1.00**	**1.00**

Table 12: Complete comparison results of CARE-PDE and ten compared methods on NLS and λ–ω. Each system reports E_{true} (%), E_{full} (%), precision (P), recall (R), and support F1. N/A indicates unsupported tasks, OOM denotes out-of-memory errors, and Err. indicates errors encountered during optimization.

Method	NLS					λ–ω				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 0\%)$										
PINN-SR (Nat. Commun., 2021)	68.181	74.648	0.50	0.33	0.40	OOM	OOM	OOM	OOM	OOM
DeepMoD (JCP, 2021)	78.689	95.912	0.08	0.50	0.14	41.122	45.062	0.88	1.00	0.93
R-DLGA (Phys. Rev. Res., 2021)	0.1567	0.2577	1.00	1.00	1.00	61.137	73.114	0.77	0.71	0.74
PDE-READ (Neural Netw., 2022)	42.218	54.035	1.00	0.67	0.80	86.819	89.555	1.00	0.14	0.25
PIC (Research, 2023)	100.00	335.62	0.00	0.00	0.00	91.079	100.52	0.60	0.21	0.32
R-DISCOVER (Phys. Fluids, 2024)	30.063	68.181	0.50	0.83	0.63	53.643	70.480	1.00	0.71	0.83
Weak-PDE-LEARN (JCP, 2024)	72.721	86.481	0.04	0.67	0.07	99.089	112.13	0.31	0.28	0.29
EqGPT (Nat. Commun., 2025)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ABL-PDE (NIPS, 2025)	0.0923	0.1033	1.00	1.00	1.00	**37.055**	**41.868**	**1.00**	**1.00**	**1.00**
SI-SR (Phys. Fluids, 2026)	**0.0674**	**0.0847**	**1.00**	**1.00**	**1.00**	72.210	77.639	1.00	0.29	0.44
CARE-PDE	**0.0356**	**0.0482**	**1.00**	**1.00**	**1.00**	**2.3197**	**2.7207**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 25\%)$										
PINN-SR (Nat. Commun., 2021)	72.312	75.510	0.50	0.33	0.40	OOM	OOM	OOM	OOM	OOM
DeepMoD (JCP, 2021)	66.916	185.09	0.09	0.50	0.15	35.754	39.894	0.88	1.00	0.93
R-DLGA (Phys. Rev. Res., 2021)	16.968	16.729	1.00	1.00	1.00	64.136	74.069	0.67	0.71	0.69
PDE-READ (Neural Netw., 2022)	42.613	54.102	1.00	0.67	0.80	86.697	89.541	1.00	0.14	0.25
PIC (Research, 2023)	100.00	336.01	0.00	0.00	0.00	83.648	98.310	0.75	0.43	0.55
R-DISCOVER (Phys. Fluids, 2024)	77.224	279.45	0.40	0.33	0.36	64.991	79.513	0.90	0.64	0.75
Weak-PDE-LEARN (JCP, 2024)	77.153	83.251	0.03	0.50	0.06	87.452	90.789	0.12	0.57	0.20
EqGPT (Nat. Commun., 2025)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ABL-PDE (NIPS, 2025)	0.8405	0.8611	0.50	1.00	0.67	**39.727**	**43.902**	**0.93**	**1.00**	**0.97**
SI-SR (Phys. Fluids, 2026)	**0.5767**	**0.5768**	**1.00**	**1.00**	**1.00**	85.714	89.487	0.67	0.14	0.24
CARE-PDE	**0.0729**	**0.0853**	**1.00**	**1.00**	**1.00**	**4.1128**	**5.2903**	**1.00**	**1.00**	**1.00**

Continued on next page

Table 12 continued

Method	NLS					λ–ω				
	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	67.613	74.567	0.50	0.33	0.40	OOM	OOM	OOM	OOM	OOM
DeepMoD (JCP, 2021)	53.200	88.704	0.08	0.50	0.14	35.076	40.285	0.82	1.00	0.90
R-DLGA (Phys. Rev. Res., 2021)	1.4894	1.9164	0.46	1.00	0.63	60.054	72.767	0.77	0.71	0.74
PDE-READ (Neural Netw., 2022)	42.738	54.157	1.00	0.67	0.80	86.667	89.537	1.00	0.14	0.25
PIC (Research, 2023)	100.00	336.45	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
R-DISCOVER (Phys. Fluids, 2024)	70.717	145.89	0.33	0.50	0.40	68.016	80.353	0.90	0.64	0.75
Weak-PDE-LEARN (JCP, 2024)	74.660	83.026	0.06	0.50	0.10	86.647	89.228	0.50	0.36	0.42
EqGPT (Nat. Commun., 2025)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ABL-PDE (NIPS, 2025)	1.1066	1.0686	0.55	1.00	0.71	**39.636**	**44.559**	**1.00**	**1.00**	**1.00**
SI-SR (Phys. Fluids, 2026)	**0.2811**	**0.3005**	**1.00**	**1.00**	**1.00**	100.00	100.00	0.00	0.00	0.00
CARE-PDE	**0.2153**	**0.3012**	**1.00**	**1.00**	**1.00**	**10.246**	**12.614**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 25\%)$										
PINN-SR (Nat. Commun., 2021)	67.114	74.542	1.00	0.33	0.50	OOM	OOM	OOM	OOM	OOM
DeepMoD (JCP, 2021)	86.482	105.94	0.08	0.50	0.14	**36.649**	**40.683**	**0.88**	**1.00**	**0.93**
R-DLGA (Phys. Rev. Res., 2021)	1.3374	1.2455	0.43	1.00	0.60	66.282	76.973	0.59	0.71	0.65
PDE-READ (Neural Netw., 2022)	42.756	54.166	1.00	0.67	0.80	86.287	89.505	1.00	0.14	0.25
PIC (Research, 2023)	100.00	336.56	0.00	0.00	0.00	100.00	103.01	0.00	0.00	0.00
R-DISCOVER (Phys. Fluids, 2024)	65.246	162.40	0.44	0.67	0.53	92.267	92.547	0.50	0.14	0.22
Weak-PDE-LEARN (JCP, 2024)	75.810	84.565	0.12	0.67	0.20	93.908	93.607	0.50	0.21	0.30
EqGPT (Nat. Commun., 2025)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ABL-PDE (NIPS, 2025)	2.0488	1.9765	0.55	1.00	0.71	45.966	49.669	0.88	1.00	0.93
SI-SR (Phys. Fluids, 2026)	**0.3428**	**0.3972**	**1.00**	**1.00**	**1.00**	92.857	101.04	0.33	0.07	0.12
CARE-PDE	**0.0504**	**0.0650**	**1.00**	**1.00**	**1.00**	**7.2301**	**8.7295**	**1.00**	**1.00**	**1.00**
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 50\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	292.31	53.124	0.67	0.67	0.67	OOM	OOM	OOM	OOM	OOM
DeepMoD (JCP, 2021)	90.468	185.17	0.06	0.33	0.10	**37.304**	**41.475**	**0.88**	**1.00**	**0.93**
R-DLGA (Phys. Rev. Res., 2021)	0.8671	1.9056	0.60	1.00	0.75	62.180	74.506	0.71	0.71	0.71
PDE-READ (Neural Netw., 2022)	43.253	54.287	1.00	0.67	0.80	86.085	89.499	1.00	0.14	0.25
PIC (Research, 2023)	100.00	337.37	0.00	0.00	0.00	100.00	101.51	0.00	0.00	0.00
R-DISCOVER (Phys. Fluids, 2024)	59.152	164.46	0.23	0.83	0.36	90.286	91.147	0.33	0.14	0.20
Weak-PDE-LEARN (JCP, 2024)	Err.	Err.	Err.	Err.	Err.	92.769	91.831	0.50	0.29	0.36
EqGPT (Nat. Commun., 2025)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ABL-PDE (NIPS, 2025)	3.2715	3.7047	0.50	1.00	0.67	**47.284**	**50.848**	**0.88**	**1.00**	**0.93**
SI-SR (Phys. Fluids, 2026)	**0.2181**	**0.2609**	**1.00**	**1.00**	**1.00**	100.00	100.00	0.00	0.00	0.00
CARE-PDE	**0.1975**	**0.2872**	**1.00**	**1.00**	**1.00**	40.306	122.04	0.54	0.93	0.68
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (5\%, 25\%, 50\%)$										
PINN-SR (Nat. Commun., 2021)	70.740	75.073	0.50	0.33	0.40	OOM	OOM	OOM	OOM	OOM
DeepMoD (JCP, 2021)	54.890	83.136	0.07	0.50	0.13	38.949	43.159	0.88	1.00	0.93
R-DLGA (Phys. Rev. Res., 2021)	8.5652	11.891	0.60	1.00	0.75	61.069	73.028	0.71	0.71	0.71
PDE-READ (Neural Netw., 2022)	42.091	53.981	1.00	0.67	0.80	86.760	89.547	1.00	0.14	0.25
PIC (Research, 2023)	100.00	335.98	0.00	0.00	0.00	92.576	99.257	0.60	0.21	0.32
R-DISCOVER (Phys. Fluids, 2024)	83.034	228.37	0.27	0.50	0.35	74.360	77.865	1.00	0.29	0.44
Weak-PDE-LEARN (JCP, 2024)	76.150	83.669	0.12	0.50	0.19	88.480	94.199	0.63	0.36	0.45
EqGPT (Nat. Commun., 2025)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ABL-PDE (NIPS, 2025)	1.5339	1.4819	0.50	1.00	0.67	**41.246**	**45.336**	**0.93**	**1.00**	**0.97**
SI-SR (Phys. Fluids, 2026)	**0.7333**	**0.7763**	**1.00**	**1.00**	**1.00**	87.697	90.998	0.67	0.14	0.24
CARE-PDE	**0.2604**	**0.3777**	**1.00**	**1.00**	**1.00**	**14.990**	**17.917**	**1.00**	**1.00**	**1.00**

F.2 Discovered Equations

To complement the quantitative results, we present the explicit equations discovered by CARE-PDE under the experimental settings described in Section 4.4. As shown in Table 13, CARE-PDE recovers the correct governing structures in all 36 cases, although the coefficient accuracy varies across systems and corruption rates.

Table 13: Equations discovered by CARE-PDE for the 12 PDE systems under the three corruption rates considered in Section 4.4.

PDE system	Discovered equation
$(\gamma_{\mathrm{L}}, \gamma_{\mathrm{H}}, \beta) = (0\%, 25\%, 0\%)$	
Advection	$u_t = -0.1000u_x$
Conv.–diff.	$u_t = -1.0000u_x + 0.2498u_{xx}$
Burgers	$u_t = -1.0001uu_x + 0.1001u_{xx}$
KdV	$u_t = -0.9990uu_x - 0.0025u_{xxx}$
KS	$u_t = -1.0308uu_x - 1.0277u_{xxxx} - 1.0114u_{xx}$
Allen–Cahn	$u_t = 0.9997u - 0.9995u^3 + 0.0030u_{xx}$
Chaffee–Infante	$u_t = -0.9999u + 0.9997u^3 + 0.9994u_{xx}$
PDE-compound	$u_t = 0.1994u_x^2 + 0.1981uu_{xx}$
Burgers2D	$u_t = -1.0021uu_x - 0.9904uu_y + 0.0100u_{xx} + 0.0099u_{yy}$
Adv.–diff. (2D)	$u_t = 0.5002u_y + 0.4999u_{xx} + 0.4998u_{yy} + 0.2501u_x$
NLS	$\begin{cases} a_t = 1.0001a^2b + 0.9999b^3 + 0.4998b_{xx}, \\ b_t = -1.0003a^3 - 0.9991ab^2 - 0.4998a_{xx} \end{cases}$
λ–ω	$\begin{cases} u_t = 0.9987v^3 + 0.9723u + 0.9716u^2v - 0.9695u^3 \\ \qquad - 0.9640uv^2 + 0.0977u_{xx} + 0.0976u_{yy}, \\ v_t = -0.9999u^3 - 0.9766v^3 + 0.9764v - 0.9719uv^2 \\ \qquad - 0.9586u^2v + 0.0984v_{xx} + 0.0979v_{yy} \end{cases}$
$(\gamma_{\mathrm{L}}, \gamma_{\mathrm{H}}, \beta) = (0\%, 25\%, 25\%)$	
Advection	$u_t = -0.1000u_x$
Conv.–diff.	$u_t = -1.0011u_x + 0.2497u_{xx}$
Burgers	$u_t = -1.0004uu_x + 0.1001u_{xx}$
KdV	$u_t = -0.9992uu_x - 0.0025u_{xxx}$
KS	$u_t = -1.0111uu_x - 1.0103u_{xxxx} - 0.9910u_{xx}$
Allen–Cahn	$u_t = 0.9998u - 0.9997u^3 + 0.0030u_{xx}$
Chaffee–Infante	$u_t = 0.9997u^3 + 0.9996u_{xx} - 0.9980u$
PDE-compound	$u_t = 0.1999u_x^2 + 0.1984uu_{xx}$
Burgers2D	$u_t = -1.0056uu_x - 0.9867uu_y + 0.0100u_{xx} + 0.0099u_{yy}$
Adv.–diff. (2D)	$u_t = 0.5002u_y + 0.4999u_{yy} + 0.4997u_{xx} + 0.2503u_x$
NLS	$\begin{cases} a_t = 1.0009a^2b + 0.9994b^3 + 0.5005b_{xx}, \\ b_t = -0.9994a^3 - 0.9988ab^2 - 0.5000a_{xx} \end{cases}$
λ–ω	$\begin{cases} u_t = 0.9952v^3 + 0.9627u - 0.9380uv^2 - 0.9272u^3 \\ \qquad + 0.9064u^2v + 0.0983u_{xx} + 0.0967u_{yy}, \\ v_t = -1.0252uv^2 + 0.9671v - 0.9596u^3 - 0.9595v^3 \\ \qquad - 0.9386u^2v + 0.0974v_{yy} + 0.0971v_{xx} \end{cases}$
$(\gamma_{\mathrm{L}}, \gamma_{\mathrm{H}}, \beta) = (0\%, 25\%, 50\%)$	
Advection	$u_t = -0.1000u_x$
Conv.–diff.	$u_t = -1.0010u_x + 0.2494u_{xx}$

Continued on next page

Table 13 continued

PDE system	Discovered equation
Burgers	$u_t = -0.9998uu_x + 0.1000u_{xx}$
KdV	$u_t = -0.9980uu_x - 0.0025u_{xxx}$
KS	$u_t = -0.9358uu_x - 0.9274u_{xxxx} - 0.9150u_{xx}$
Allen–Cahn	$u_t = -0.9998u^3 + 0.9995u + 0.0030u_{xx}$
Chaffee–Infante	$u_t = -1.0049u + 1.0019u^3 + 1.0015u_{xx}$
PDE-compound	$u_t = 0.1988u_x^2 + 0.1976uu_{xx}$
Burgers2D	$u_t = -1.0385uu_x - 1.0014uu_y + 0.0102u_{xx} + 0.0099u_{yy}$
Adv.–diff. (2D)	$u_t = 0.5027u_{yy} + 0.5014u_y + 0.5010u_{xx} + 0.2505u_x$
NLS	$\begin{cases} a_t = 1.0053a^2b + 0.9982b^3 + 0.5000b_{xx}, \\ b_t = -1.0028ab^2 - 0.9993a^3 - 0.4989a_{xx} \end{cases}$
λ–ω	$\begin{cases} u_t = 1.0006v^3 - 0.9605uv^2 + 0.8839u - 0.7869u^3 \\ \qquad + 0.7692u^2v + 0.0895u_{xx} + 0.0885u_{yy}, \\ v_t = -1.0098u^3 - 0.9288u^2v + 0.9118v - 0.8994uv^2 \\ \qquad - 0.8498v^3 + 0.0914v_{yy} + 0.0892v_{xx} \end{cases}$

F.3 Adaptive GMM Reliability Analysis

To examine AGRE's adaptability across different NDEC settings and the effectiveness of the resulting reliability estimates, we analyze residual distributions and GMM partitioning results. We visualize the residual distributions and fitted GMMs for the experiments reported in Section 4.4, as shown in Figures 2 and 3. We also summarize the statistical results of all experiments in Section F.1 in Table 14.

The results show that AGRE consistently produces partitions that agree with the underlying perturbation configurations across all evaluated PDE systems. A single component is selected when no high-perturbation group is present, whereas two components are consistently identified under heterogeneous perturbations. Among these heterogeneous settings, the separation between the two residual populations becomes more pronounced as the difference between the perturbation levels increases, resulting in clearer GMM partitioning. This separation is facilitated by the Cauchy-robust reconstruction, which limits the influence of highly perturbed observations on the reconstructed field and allows their larger discrepancies to remain reflected in the residual distribution. Consequently, the reliability weights of the two groups are clearly separated, accompanied by generally high AUC values. When both groups are noisy, as in $(5\%, 25\%, 50\%)$, the residual distribution of the low-perturbation group becomes broader and overlaps more with that of the high-perturbation group, making the partition more difficult. Nevertheless, the two groups remain distinguishable in their estimated reliability weights, indicating that the AGRE can still capture the underlying reliability heterogeneity even when the residual populations exhibit substantial overlap.

F.4 Local Feature Quality Analysis

To examine the quality of local differential features under strong heterogeneous perturbations, we conduct a controlled diagnostic experiment on the KS equation under $(\gamma_{\mathrm{L}}, \gamma_{\mathrm{H}}, \beta) = (0\%, 25\%, 50\%)$. Specifically, we compare AD and RGLP using both pointwise and LICA-integrated features, with clean numerical references for evaluation. The results are shown in Fig. 4.

The results show that the pointwise derivatives obtained by AD become increasingly inaccurate as the derivative order increases, with severe error amplification observed for the higher-order terms. RGLP substantially alleviates this deterioration, although noticeable errors remain for the highest-order derivatives. This improvement arises from the local Legendre projection, which provides a smoother local representation and reduces the sensitivity of derivative estimation to perturbations in the reconstructed field. The integral results further show that LICA cannot compensate for severely inaccurate derivatives produced by AD, whereas applying LICA to the more stable RGLP-derived

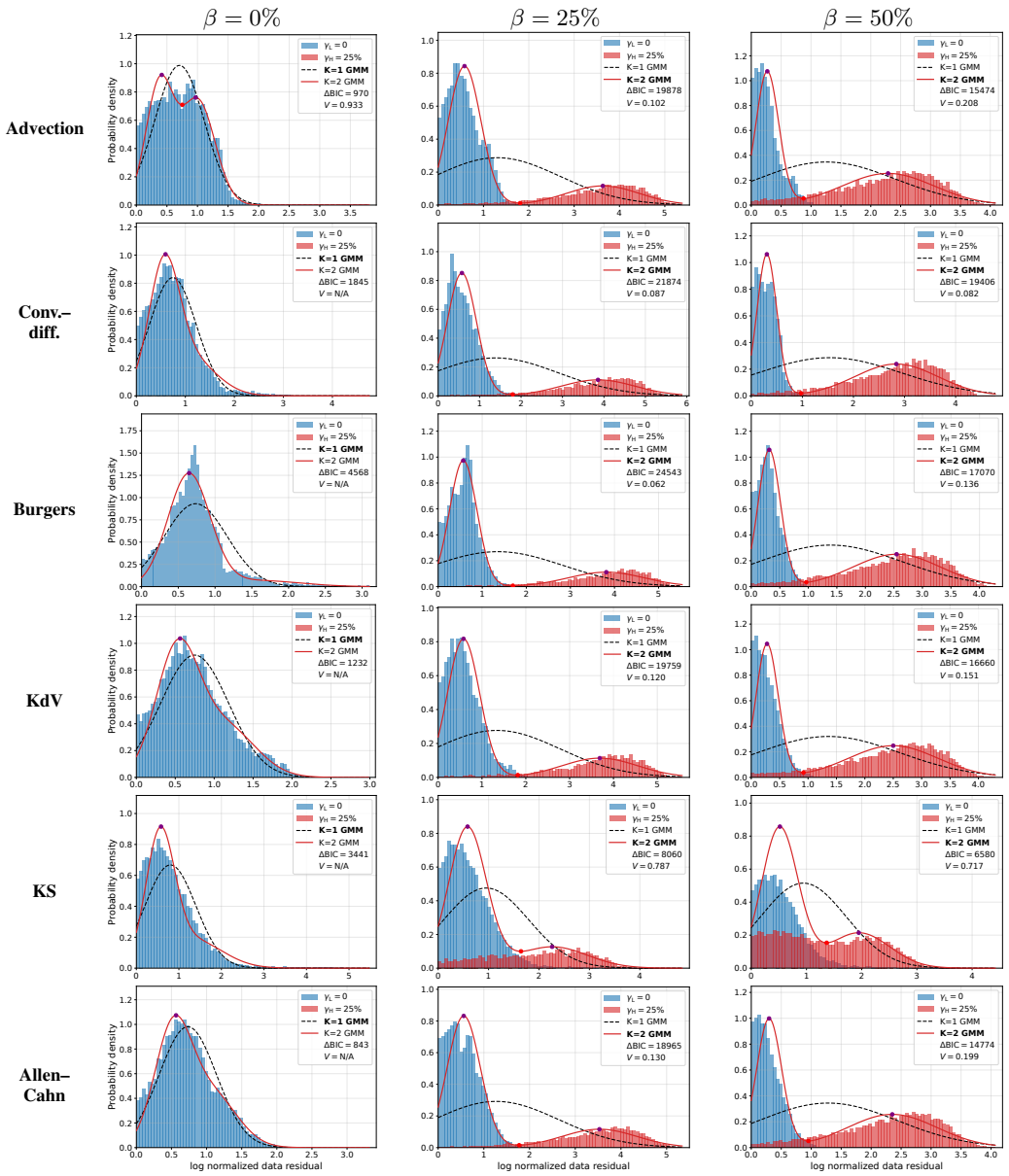

Figure 2: GMM partitioning results for Advection, Conv.–Diff., Burgers, KdV, KS, and Allen–Cahn under $(\gamma_{\mathrm{L}}, \gamma_{\mathrm{H}}) = (0\%, 25\%)$. The dashed and solid curves denote the one- and two-component GMM fits, respectively, with the bold entry in the legend indicating the final model selected. $\Delta\mathrm{BIC}$ denotes the BIC difference used for model selection, and V denotes the peak-to-valley ratio used to characterize the separation between two modes. N/A indicates that no valid bimodal structure is formed.

quantities yields considerably more accurate integral features. These results demonstrate that RGLP primarily suppresses error amplification during high-order derivative estimation, while LICA further stabilizes the representation of the resulting local differential quantities.

F.5 Analysis of the Cauchy Robustness

To investigate the influence of the Cauchy robustness parameter c on the performance of CARE-PDE, we conduct parameter sensitivity experiments under the NDEC scenarios considered in Section 4.4. Specifically, we vary c from 0.5 to 5 and evaluate the resulting equation-recovery errors

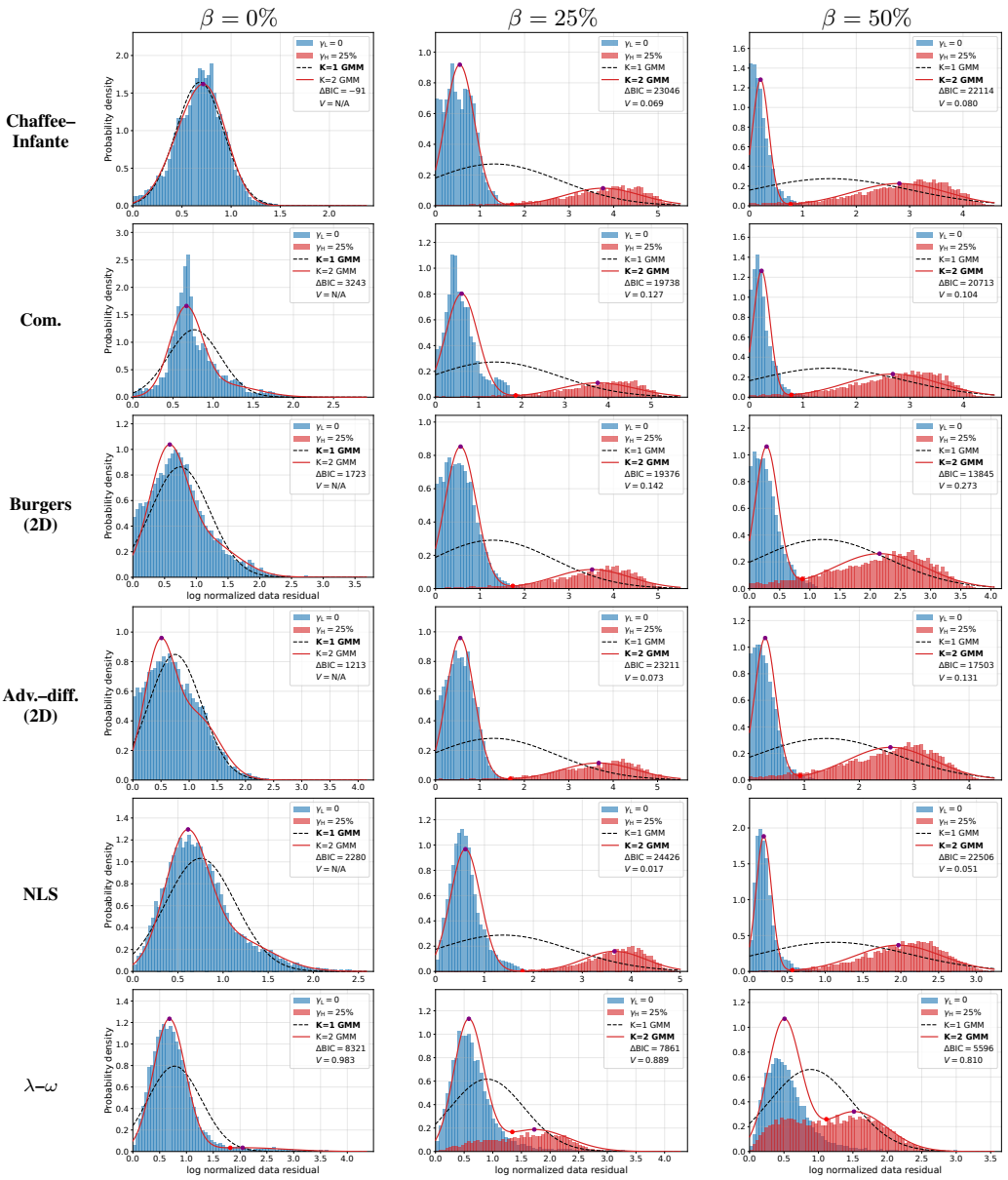

Figure 3: GMM partitioning results for Chaffee–Infante, Compound, Burgers2D, Adv.–diff. (2D), NLS, and the λ–ω system under $(\gamma_{\mathrm{L}}, \gamma_{\mathrm{H}}) = (0\%, 25\%)$. The dashed and solid curves denote the one- and two-component GMM fits, respectively, with the bold entry in the legend indicating the final model selected. ΔBIC denotes the BIC difference used for model selection, and V denotes the peak-to-valley ratio used to characterize the separation between two modes. N/A indicates that no valid bimodal structure is formed.

and identification metrics. The averaged results over the 12 PDE systems are presented in Fig. 5, while the detailed results for each individual PDE system are summarized in Table 15.

The results show that the selection of c determines the balance between robustness and reconstruction fidelity. When c is relatively small, the Cauchy loss imposes strong suppression on large residuals. Although this improves robustness to abnormal observations, excessive robustness may also suppress informative regions that inherently exhibit higher reconstruction errors, leading to underfitting of complex physical patterns. As c increases, the loss becomes less robust and gradually allows larger residual contributions, thereby improving fitting flexibility but making the reconstruc-

Table 14: GMM reliability estimation under different NDEC settings. $G^\star$ denotes the selected number of Gaussian components, $\widehat{\beta}$ denotes the estimated proportion of the high-perturbation group, and $\bar{a}_{\mathrm{L}}$ and $\bar{a}_{\mathrm{H}}$ denote the mean reliability weights of the low- and high-perturbation groups, respectively. AUC measures the discrimination between the two groups. The mean is computed over all 12 PDE systems.

NDEC setting		Adv.	Conv.–diff.	Burgers	KdV	KS	Allen–Cahn	Chaffee–Infante	PDE-compound	Burgers2D	Adv.–diff. (2D)	NLS	λ–ω	Mean
$(\gamma_{\mathrm{L}}, \gamma_{\mathrm{H}}, \beta)$	Metric													
$(0\%, 25\%, 0\%)$	$G^\star$	$G=1$	$G=1$	$G=1$	$G=1$	$G=1$	$G=1$	$G=1$	$G=1$	$G=1$	$G=1$	$G=1$	$G=1$	**G=1**
	$\widehat{\beta}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0000**
	$\bar{a}_{\mathrm{L}}$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	**1.0000**
	$\bar{a}_{\mathrm{H}}$	–	–	–	–	–	–	–	–	–	–	–	–	**–**
	AUC	–	–	–	–	–	–	–	–	–	–	–	–	**–**
$(0\%, 25\%, 25\%)$	$G^\star$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	**G=2**
	$\widehat{\beta}$	0.2358	0.2482	0.2422	0.2464	0.2264	0.2401	0.2423	0.2476	0.2448	0.2403	0.2535	0.2870	**0.2462**
	$\bar{a}_{\mathrm{L}}$	0.9955	0.9870	0.9944	0.9823	0.9071	0.9893	0.9943	0.9790	0.9816	0.9957	0.9913	0.8297	**0.9689**
	$\bar{a}_{\mathrm{H}}$	0.1532	0.1320	0.1357	0.1441	0.3488	0.1502	0.1362	0.1435	0.1529	0.1396	0.1058	0.3071	**0.1708**
	AUC	0.9864	0.9922	0.9903	0.9889	0.8770	0.9884	0.9907	0.9890	0.9855	0.9885	0.9992	0.8657	**0.9702**
$(0\%, 25\%, 50\%)$	$G^\star$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	**G=2**
	$\widehat{\beta}$	0.4912	0.4859	0.4810	0.4883	0.3041	0.4814	0.5048	0.5057	0.4976	0.4882	0.5187	0.3866	**0.4695**
	$\bar{a}_{\mathrm{L}}$	0.9350	0.9823	0.9715	0.9564	0.9022	0.9519	0.9545	0.9481	0.9114	0.9633	0.9513	0.8189	**0.9372**
	$\bar{a}_{\mathrm{H}}$	0.1516	0.1336	0.1462	0.1452	0.4879	0.1567	0.1236	0.1260	0.1578	0.1409	0.1053	0.3948	**0.1891**
	AUC	0.9789	0.9890	0.9826	0.9823	0.7814	0.9800	0.9914	0.9894	0.9730	0.9865	0.9971	0.7996	**0.9526**
$(0\%, 50\%, 25\%)$	$G^\star$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	**G=2**
	$\widehat{\beta}$	0.2435	0.2501	0.2514	0.2398	0.2234	0.2447	0.2468	0.2456	0.2442	0.2472	0.2496	0.2646	**0.2459**
	$\bar{a}_{\mathrm{L}}$	0.9969	0.9942	0.9881	0.9986	0.9616	0.9951	0.9960	0.9886	0.9940	0.9957	0.9989	0.9160	**0.9853**
	$\bar{a}_{\mathrm{H}}$	0.1265	0.1131	0.1211	0.1355	0.2639	0.1264	0.1190	0.1339	0.1309	0.1177	0.1034	0.2407	**0.1443**
	AUC	0.9946	0.9962	0.9954	0.9913	0.9349	0.9944	0.9962	0.9925	0.9931	0.9956	0.9998	0.9210	**0.9838**
$(0\%, 50\%, 50\%)$	$G^\star$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	**G=2**
	$\widehat{\beta}$	0.4814	0.4930	0.4920	0.4941	0.3625	0.4929	0.5033	0.4929	0.5019	0.4964	0.5088	0.3923	**0.4760**
	$\bar{a}_{\mathrm{L}}$	0.9875	0.9881	0.9831	0.9734	0.9195	0.9725	0.9738	0.9869	0.9468	0.9759	0.9783	0.8418	**0.9606**
	$\bar{a}_{\mathrm{H}}$	0.1372	0.1191	0.1238	0.1263	0.3899	0.1285	0.1131	0.1198	0.1324	0.1214	0.1022	0.4104	**0.1687**
	AUC	0.9890	0.9939	0.9926	0.9917	0.8437	0.9912	0.9955	0.9942	0.9862	0.9927	0.9986	0.7762	**0.9621**
$(5\%, 25\%, 50\%)$	$G^\star$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	$G=2$	**G=2**
	$\widehat{\beta}$	0.3427	0.3447	0.3441	0.3414	0.3062	0.3479	0.3451	0.3402	0.3330	0.3479	0.4234	0.3867	**0.3503**
	$\bar{a}_{\mathrm{L}}$	0.9218	0.9201	0.9278	0.9222	0.8548	0.9191	0.9284	0.9263	0.9288	0.9234	0.9520	0.7539	**0.9066**
	$\bar{a}_{\mathrm{H}}$	0.3827	0.3787	0.3791	0.3842	0.5141	0.3766	0.3752	0.3828	0.3966	0.3744	0.2336	0.4460	**0.3853**
	AUC	0.8732	0.8722	0.8714	0.8701	0.7363	0.8671	0.8773	0.8723	0.8654	0.8727	0.9608	0.7254	**0.8554**

tion more sensitive to highly perturbed observations. Therefore, excessively small or large values of c may degrade the final equation recovery performance. A moderate value of c achieves a better balance between suppressing unreliable residuals and preserving meaningful physical information. Based on this observation, we set $c = 1$ throughout all experiments.

Table 15: Sensitivity analysis of the Cauchy robustness parameter c. The table reports detailed results for individual PDE systems.

Dataset	Metric	c									
		0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
$\beta = 0\%$											
	E_{true}	0.0065	0.0037	0.0001	0.0045	0.0015	0.0032	0.0014	0.0021	0.0035	0.0040
	E_{full}	0.0065	0.0037	0.0001	0.0045	0.0015	0.0032	0.0014	0.0021	0.0035	0.0040
Advection	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0631	0.0331	0.0647	0.0491	0.0449	0.0340	0.0209	0.0390	0.0556	0.0628
	E_{full}	0.0357	0.0157	0.0521	0.0306	0.0211	0.0232	0.0101	0.0216	0.0402	0.0296
Conv.–diff.	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Continued on next page

Table 15 continued

Dataset	Metric	c									
		0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Burgers	E_{true}	0.0669	0.0514	0.0579	0.0604	0.0615	0.0586	0.0592	0.0609	0.0678	0.0637
	E_{full}	0.0184	0.0121	0.0197	0.0215	0.0196	0.0197	0.0197	0.0199	0.0278	0.0187
	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
KdV	E_{true}	0.0185	0.0872	0.0289	0.0151	0.0188	0.0566	0.0169	0.0449	0.0242	0.0219
	E_{full}	0.0041	0.0977	0.0169	0.0198	0.0184	0.0523	0.0095	0.0481	0.0233	0.0110
	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
KS	E_{true}	4.9807	2.3265	2.5652	2.0477	2.2625	2.1557	2.6715	3.1214	2.5573	2.1860
	E_{full}	5.0231	2.4772	2.7175	2.2306	2.4341	2.3058	2.7801	3.2417	2.6945	2.3403
	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Allen–Cahn	E_{true}	0.0092	0.0311	0.0596	0.0391	0.0697	0.0106	0.0137	0.0430	0.0171	0.0309
	E_{full}	0.0145	0.0402	0.0641	0.0451	0.0813	0.0103	0.0257	0.0382	0.0111	0.0244
	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Chaffee–Infante	E_{true}	0.3227	0.0347	0.1696	0.0625	0.0999	0.0128	0.2525	0.1708	0.1707	0.1600
	E_{full}	0.3833	0.0408	0.2058	0.0750	0.1188	0.0134	0.3155	0.2015	0.2017	0.1891
	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Compound	E_{true}	0.6245	0.6492	0.5705	0.4950	0.5444	0.6190	0.5745	0.6123	0.6087	0.6152
	E_{full}	0.7099	0.7259	0.6641	0.6111	0.6479	0.6956	0.6654	0.6936	0.6985	0.7037
	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Burgers2D	E_{true}	0.4568	0.5693	0.5029	0.4916	0.4502	0.4674	0.5070	0.4598	0.5384	0.4976
	E_{full}	0.4260	0.6960	0.5659	0.4977	0.4565	0.5157	0.5434	0.4394	0.5688	0.4837
	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Adv.–diff. (2D)	E_{true}	0.0235	0.0361	0.0439	0.0487	0.0485	0.0341	0.0361	0.0334	0.0373	0.0390
	E_{full}	0.0270	0.0379	0.0416	0.0529	0.0414	0.0454	0.0439	0.0388	0.0475	0.0494
	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
NLS	E_{true}	0.0560	0.0356	0.0227	0.0352	0.0284	0.0605	0.0477	0.0375	0.0500	0.0210
	E_{full}	0.0557	0.0482	0.0398	0.0494	0.0528	0.0930	0.0688	0.0600	0.0844	0.0317
	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
λ–ω	E_{true}	2.5331	2.3197	2.2789	2.5707	1.7548	1.8844	2.5219	2.5334	3.1069	1.9520
	E_{full}	2.8120	2.7207	3.0168	3.2707	2.2494	2.2339	2.9937	3.0352	4.3596	2.5018
	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Continued on next page

Table 15 continued

Dataset	Metric	c									
		0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\beta = 25\%$											
	E_{true}	0.0041	0.0038	0.0041	0.0061	0.0049	0.0037	0.0205	0.0355	0.0280	0.0328
	E_{full}	0.0041	0.0038	0.0041	0.0061	0.0049	0.0037	0.0205	0.0355	0.0280	0.0328
Advection	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0866	0.1165	0.1450	0.1290	0.1050	0.1460	0.1229	0.1657	0.1261	0.0988
	E_{full}	0.0665	0.1084	0.1163	0.1032	0.1160	0.1610	0.1633	0.1759	0.1612	0.1124
Conv.–diff.	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0565	0.0647	0.0260	0.0501	0.1074	0.0451	0.0551	0.0741	0.0247	0.0446
	E_{full}	0.0171	0.0418	0.0079	0.0144	0.0213	0.0275	0.0448	0.0863	0.0361	0.0817
Burgers	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0481	0.0897	0.0186	0.2091	0.0313	0.2364	0.4187	0.3132	0.4104	0.4570
	E_{full}	0.0219	0.0761	0.0131	0.1814	0.0349	0.2252	0.4201	0.2679	0.3976	0.4091
KdV	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.7896	1.0132	0.8188	0.6542	0.8243	2.7953	0.7080	1.5192	0.8706	1.6481
	E_{full}	0.9007	1.0168	0.8328	0.7688	1.0273	2.9408	1.0171	1.8237	1.1977	1.9023
KS	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0093	0.0185	0.0324	0.0468	0.0312	0.1060	0.1646	0.1959	0.1678	0.0353
	E_{full}	0.0042	0.0241	0.0481	0.0633	0.0337	0.1405	0.1778	0.2119	0.1931	0.0347
Allen–Cahn	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0342	0.0879	0.0166	0.1204	0.1949	0.0310	0.2318	0.2278	0.2529	0.1992
	E_{full}	0.0356	0.1167	0.0225	0.1396	0.2914	0.0363	0.3307	0.2877	0.3187	0.2823
Chaffee–Infante	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.5312	0.4219	0.4826	0.6482	0.6324	0.7583	0.8917	0.9580	0.9671	0.8071
	E_{full}	0.6270	0.5722	0.5745	0.7345	0.6939	0.8177	0.9220	0.9852	0.9802	0.8583
Compound	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.6554	0.7345	0.9140	0.8547	0.5797	0.9259	1.6364	2.1064	1.6847	1.4976
	E_{full}	0.9452	1.0170	1.4671	1.5153	1.0770	1.0802	2.4347	3.5347	2.7995	2.3577
Burgers2D	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Continued on next page

Table 15 continued

Dataset	Metric	c									
		0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
	E_{true}	0.0567	0.0580	0.0748	0.1160	0.1131	0.1470	0.0946	0.0695	0.1406	0.1074
	E_{full}	0.0675	0.0538	0.0784	0.1278	0.1460	0.1452	0.1033	0.0820	0.1619	0.1693
Adv.–diff. (2D)	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0505	0.0729	0.1486	0.1440	0.2547	0.2538	0.2855	0.4108	0.4800	0.4262
	E_{full}	0.0669	0.0853	0.2020	0.1762	0.3156	0.3145	0.3372	0.5301	0.5741	0.5268
NLS	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	6.7396	4.1128	9.1328	8.6770	9.5095	11.037	17.067	18.168	15.785	15.279
	E_{full}	8.2016	5.2903	10.950	10.090	11.269	12.757	24.157	32.928	18.500	18.573
λ–ω	P	1.00	1.00	1.00	1.00	1.00	1.00	0.93	0.88	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.93	1.00	1.00
$\beta = 50\%$											
	E_{true}	0.0002	0.0058	0.0105	0.0051	0.0411	0.0267	0.0465	0.0521	0.0656	0.0727
	E_{full}	0.0002	0.0058	0.0105	0.0051	0.0411	0.0267	0.0465	0.0521	0.0656	0.0727
Advection	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.1387	0.1734	0.2108	0.1275	0.3518	0.4861	0.5271	0.4501	0.6052	0.6252
	E_{full}	0.0995	0.1163	0.1421	0.0714	0.1715	0.2450	0.2500	0.2119	0.2929	0.2943
Conv.–diff.	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0464	0.0304	0.0786	0.1140	0.2330	0.3261	0.3703	0.4274	0.5729	0.6208
	E_{full}	0.0326	0.0235	0.0995	0.2004	0.3346	0.4464	0.4650	0.5669	0.6332	0.6256
Burgers	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0324	0.2176	0.6777	0.9113	1.2694	1.5619	2.3157	2.2007	2.2367	3.1175
	E_{full}	0.0271	0.2048	0.6275	0.8189	1.1880	1.4662	2.1679	1.9830	2.0746	2.8591
KdV	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	6.0939	7.3953	5.9386	6.4691	9.6918	9.7827	7.1341	10.441	8.1215	10.233
	E_{full}	6.2054	7.4444	5.9876	6.5074	9.7379	9.8463	7.2151	10.478	8.2161	10.356
KS	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0544	0.0312	0.1270	0.0674	0.2830	0.4196	0.4664	0.5274	0.4261	0.6530
	E_{full}	0.0452	0.0366	0.1327	0.0783	0.3135	0.4501	0.4996	0.6090	0.4276	0.6789
Allen–Cahn	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0713	0.2750	0.2780	43.958	0.3750	0.3962	0.3718	43.555	69.986	106.31
	E_{full}	0.0723	0.3146	0.3861	59.329	0.4591	0.4254	0.4379	59.224	86.344	206.95

Continued on next page

Table 15 continued

Dataset	Metric	c									
		0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Chaffee–Infante	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.40
	R	1.00	1.00	1.00	0.67	1.00	1.00	1.00	0.67	0.33	0.67
	F1	1.00	1.00	1.00	0.80	1.00	1.00	1.00	0.80	0.33	0.50
	E_{true}	0.8077	0.8906	0.8671	56.863	56.929	57.788	57.128	57.092	1.8247	1.8649
	E_{full}	0.8755	0.9441	0.9022	268.30	267.70	265.51	265.40	264.75	1.8350	1.8723
Compound	P	1.00	1.00	1.00	0.25	0.25	0.14	0.25	0.25	1.00	1.00
	R	1.00	1.00	1.00	0.50	0.50	0.50	0.50	0.50	1.00	1.00
	F1	1.00	1.00	1.00	0.33	0.33	0.22	0.33	0.33	1.00	1.00
	E_{true}	0.4150	1.6862	2.6833	34.734	34.925	35.969	10.725	67.217	133.41	62.561
	E_{full}	0.7786	2.7276	6.3219	72.235	72.990	71.875	DIV.	DIV.	DIV.	DIV.
Burgers2D	P	1.00	1.00	1.00	1.00	1.00	1.00	0.57	0.22	0.33	0.30
	R	1.00	1.00	1.00	0.75	0.75	0.75	1.00	0.50	1.00	0.75
	F1	1.00	1.00	1.00	0.86	0.86	0.86	0.73	0.31	0.50	0.43
	E_{true}	0.0733	0.3066	0.3776	0.4091	0.7296	1.2214	0.9750	1.2106	0.9420	0.7292
	E_{full}	0.0926	0.3552	0.4481	0.5425	0.7384	1.3519	1.1280	1.1795	0.8835	0.6750
Adv.–diff. (2D)	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.1716	0.2153	0.3358	0.6085	0.5673	0.8124	0.8351	0.8591	1.0284	1.0253
	E_{full}	0.2516	0.3012	0.4303	0.6854	0.7207	0.9665	1.0085	0.9335	1.2577	1.3117
NLS	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	12.298	10.246	17.455	17.426	19.047	22.500	24.614	33.247	21.545	45.136
	E_{full}	25.299	12.614	20.075	19.047	21.333	35.545	81.644	48.833	79.444	62.672
λ–ω	P	0.93	1.00	1.00	1.00	1.00	1.00	0.93	1.00	0.88	0.92
	R	1.00	1.00	1.00	1.00	1.00	0.93	0.93	0.86	1.00	0.79
	F1	0.97	1.00	1.00	1.00	1.00	0.96	0.93	0.92	0.93	0.85

F.6 Complexity Weight Analysis

To investigate the effect of the complexity weight α in the equation-scoring function, we conduct parameter sensitivity experiments under the NDEC scenarios used in Section 4.4. The averaged results over 12 PDE systems are presented in Fig. 6, while the detailed results for each PDE system are summarized in Table 16.

The results show that α consistently affects the trade-off between equation complexity and recovery accuracy. When α is small, the complexity constraint is insufficient, which may lead to the selection of overly complex candidate equations and unstable equation recovery. As α increases, CARE-PDE gradually improves its performance and maintains stable results within a certain range of α values, indicating that an appropriate complexity penalty can balance equation compactness and recovery accuracy. However, excessively large α values impose overly strong penalties on complex terms and may suppress necessary components of the governing equations, resulting in degraded recovery performance, especially under severe perturbations. Based on these observations, we set $\alpha = 0.15$ in all experiments.

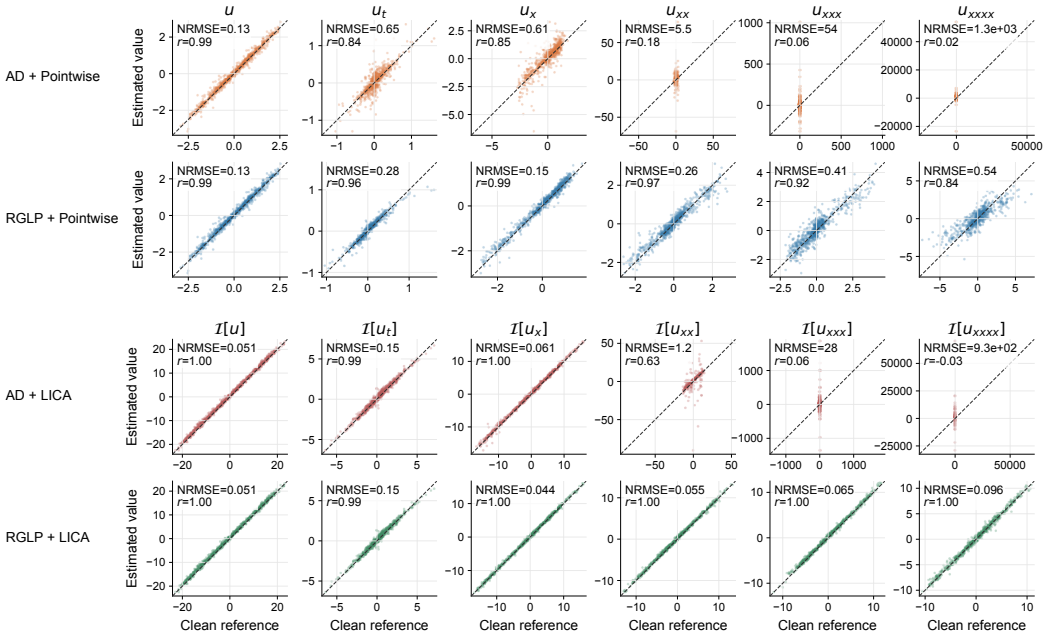

Figure 4: Comparison of derivative quality under different differentiation and constraint constructions. The four rows correspond to AD + Pointwise, RGLP + Pointwise, AD + LICA, and RGLP + LICA, respectively. Each column compares the estimated quantity against its clean reference, with NRMSE and correlation coefficient r reported in each panel.

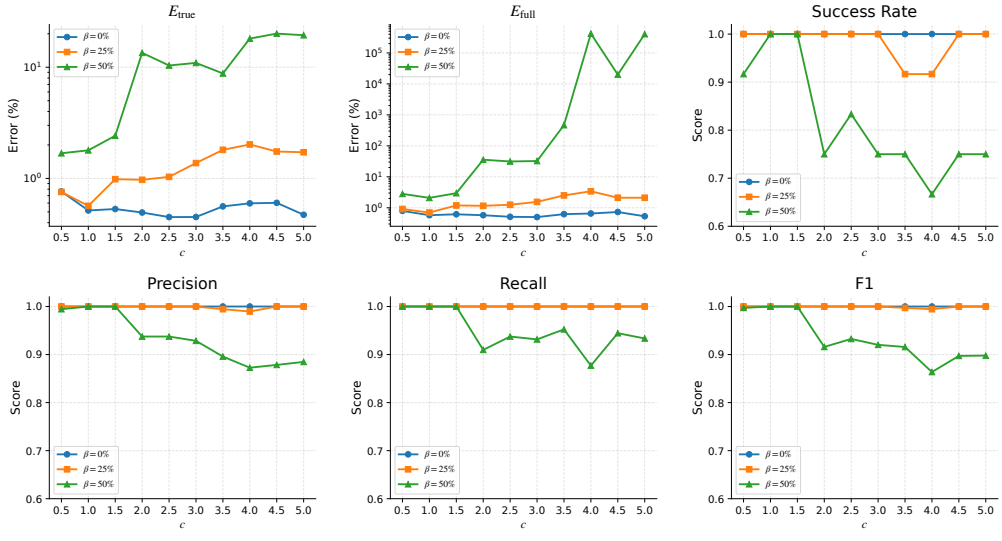

Figure 5: Effect of the Cauchy parameter c on PDE discovery performance. The curves show the average results across 12 PDE systems under different NDEC scenarios.

Table 16: Sensitivity analysis of the complexity weight α in the equation scoring function. Detailed results are reported for individual PDE systems.

Dataset	Metric	α										
		0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
$\beta = 0\%$												
	E_{true}	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037
	E_{full}	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037
Advection	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	1.0202	0.0331	0.0331	0.0331	0.0331	0.0331	0.0331	0.0331	0.0331	0.0331	0.0331
	E_{full}	283.75	0.0157	0.0157	0.0157	0.0157	0.0157	0.0157	0.0157	0.0157	0.0157	0.0157
Conv.–diff.	P	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.2213	0.0514	0.0514	0.0514	0.0514	0.0514	0.0514	0.0514	0.0514	0.0514	0.0514
	E_{full}	0.1129	0.0121	0.0121	0.0121	0.0121	0.0121	0.0121	0.0121	0.0121	0.0121	0.0121
Burgers	P	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.1387	0.0872	0.0872	0.0872	0.0872	0.0872	0.0872	0.0872	0.0872	0.0872	0.0872
	E_{full}	0.7393	0.0977	0.0977	0.0977	0.0977	0.0977	0.0977	0.0977	0.0977	0.0977	0.0977
KdV	P	0.17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.29	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	1.9185	2.3265	2.3265	2.3265	2.3265	2.3265	2.3265	2.3265	2.3265	2.3265	2.3265
	E_{full}	6.9674	2.4772	2.4772	2.4772	2.4772	2.4772	2.4772	2.4772	2.4772	2.4772	2.4772
KS	P	0.12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.21	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.1135	0.0311	0.0311	0.0311	0.0311	0.0311	0.0311	0.0311	0.0311	0.0311	0.0311
	E_{full}	0.1907	0.0402	0.0402	0.0402	0.0402	0.0402	0.0402	0.0402	0.0402	0.0402	0.0402
Allen–Cahn	P	0.43	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	9.9532	0.0347	0.0347	0.0347	0.0347	42.444	42.444	42.444	42.444	83.511	83.511
	E_{full}	15.536	0.0408	0.0408	0.0408	0.0408	58.951	58.951	58.951	58.951	86.706	86.706
Chaffee–Infante	P	0.14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	0.67	0.67	0.67	0.67	0.33	0.33
	F1	0.24	1.00	1.00	1.00	1.00	0.80	0.80	0.80	0.80	0.50	0.50
	E_{true}	7.9040	0.6492	0.6492	0.6492	0.6492	0.6492	0.6492	0.6492	0.6492	0.6492	0.6492
	E_{full}	580.79	0.7259	0.7259	0.7259	0.7259	0.7259	0.7259	0.7259	0.7259	0.7259	0.7259
Compound	P	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	2.3317	0.5693	0.5693	0.5693	0.5693	0.5693	0.5693	0.5693	0.5693	0.5693	0.5693
	E_{full}	DIV.	0.6960	0.6960	0.6960	0.6960	0.6960	0.6960	0.6960	0.6960	0.6960	0.6960
Burgers2D	P	0.13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.24	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Continued on next page

Table 16 continued

Dataset	Metric	α										
		0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
	E_{true}	3.0185	0.0361	0.0361	0.0361	0.0361	0.0361	0.0361	0.0361	0.0361	0.0361	0.0361
	E_{full}	34.800	0.0379	0.0379	0.0379	0.0379	0.0379	0.0379	0.0379	0.0379	0.0379	0.0379
Adv.–diff. (2D)	P	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.18	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	1.5044	0.0356	0.0356	0.0356	0.0356	0.0356	0.0356	0.0356	71.286	71.286	71.286
	E_{full}	6.9287	0.0482	0.0482	0.0482	0.0482	0.0482	0.0482	0.0482	75.108	75.108	75.108
NLS	P	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.33	0.33
	F1	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.50	0.50
	E_{true}	5.1586	2.3197	2.3197	2.3197	2.3197	2.3197	2.3197	89.280	89.280	89.280	89.280
	E_{full}	30.767	2.7207	2.7207	2.7207	2.7207	2.7207	2.7207	90.178	90.178	90.178	90.178
λ–ω	P	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.14	0.14	0.14	0.14
	F1	0.07	1.00	1.00	1.00	1.00	1.00	1.00	0.25	0.25	0.25	0.25
$\beta = 25\%$												
	E_{true}	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038
	E_{full}	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038
Advection	P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	1.3305	0.1165	0.1165	0.1165	0.1165	0.1165	0.1165	0.1165	0.1165	0.1165	0.1165
	E_{full}	666.53	0.1084	0.1084	0.1084	0.1084	0.1084	0.1084	0.1084	0.1084	0.1084	0.1084
Conv.–diff.	P	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.0806	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647
	E_{full}	0.3089	0.0418	0.0418	0.0418	0.0418	0.0418	0.0418	0.0418	0.0418	0.0418	0.0418
Burgers	P	0.17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.29	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.1985	0.0897	0.0897	0.0897	0.0897	0.0897	0.0897	0.0897	0.0897	0.0897	0.0897
	E_{full}	4.3644	0.0761	0.0761	0.0761	0.0761	0.0761	0.0761	0.0761	0.0761	0.0761	0.0761
KdV	P	0.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.33	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	2.6009	1.0132	1.0132	1.0132	1.0132	1.0132	1.0132	1.0132	1.0132	1.0132	1.0132
	E_{full}	6.0463	1.0168	1.0168	1.0168	1.0168	1.0168	1.0168	1.0168	1.0168	1.0168	1.0168
KS	P	0.12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.21	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.1749	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185
	E_{full}	0.2407	0.0241	0.0241	0.0241	0.0241	0.0241	0.0241	0.0241	0.0241	0.0241	0.0241
Allen–Cahn	P	0.38	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.55	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	7.9875	0.0879	0.0879	0.0879	0.0931	42.373	42.373	42.373	42.373	83.440	83.440
	E_{full}	17.144	0.1167	0.1167	0.1167	0.1099	58.933	58.933	58.933	58.933	86.664	86.664

Continued on next page

Table 16 continued

Dataset	Metric	α										
		0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Chaffee–Infante	P	0.12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	0.67	0.67	0.67	0.67	0.33	0.33
	F1	0.22	1.00	1.00	1.00	1.00	0.80	0.80	0.80	0.80	0.50	0.50
Compound	E_{true}	9.0240	3.3221	0.4219	0.4219	0.4219	0.4219	0.4219	0.4219	0.4219	0.4219	0.4219
	E_{full}	550.42	3.6500	0.5722	0.5722	0.5722	0.5722	0.5722	0.5722	0.5722	0.5722	0.5722
	P	0.10	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.18	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Burgers2D	E_{true}	2.9329	0.7345	0.7345	0.7345	0.7345	0.7345	0.7345	0.7345	0.7345	0.7345	0.7345
	E_{full}	DIV.	1.0170	1.0170	1.0170	1.0170	1.0170	1.0170	1.0170	1.0170	1.0170	1.0170
	P	0.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.21	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Adv.–diff. (2D)	E_{true}	4.3295	0.0580	0.0580	0.0580	0.0580	0.0580	0.0580	0.0580	0.0580	0.0580	0.0580
	E_{full}	39.497	0.0538	0.0538	0.0538	0.0538	0.0538	0.0538	0.0538	0.0538	0.0538	0.0538
	P	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
NLS	E_{true}	2.0717	0.0729	0.0729	0.0729	0.0729	0.0729	0.0729	0.0729	70.907	70.907	70.907
	E_{full}	9.9055	0.0853	0.0853	0.0853	0.0853	0.0853	0.0853	0.0853	75.020	75.020	75.020
	P	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.33	0.33
	F1	0.07	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.50	0.50
λ–ω	E_{true}	16.261	4.1128	4.1128	4.1128	4.1128	4.1128	4.1128	88.902	88.902	88.902	88.902
	E_{full}	62.362	5.2903	5.2903	5.2903	5.2903	5.2903	5.2903	90.049	90.049	90.049	90.049
	P	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.14	0.14	0.14	0.14
	F1	0.07	1.00	1.00	1.00	1.00	1.00	1.00	0.25	0.25	0.25	0.25
$\beta = 50\%$												
Advection	E_{true}	25.829	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058
	E_{full}	57.244	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058
	P	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Conv.–diff.	E_{true}	0.7251	0.1734	0.1734	0.1734	0.1734	0.1734	0.1734	0.1734	0.1734	0.1734	0.1734
	E_{full}	767.30	0.1163	0.1163	0.1163	0.1163	0.1163	0.1163	0.1163	0.1163	0.1163	0.1163
	P	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Burgers	E_{true}	0.5089	0.0304	0.0304	0.0304	0.0304	0.0304	0.0304	0.0304	0.0304	0.0304	0.0304
	E_{full}	1.1411	0.0235	0.0235	0.0235	0.0235	0.0235	0.0235	0.0235	0.0235	0.0235	0.0235
	P	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.18	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
KdV	E_{true}	0.2912	0.2176	0.2176	0.2176	0.2176	0.2176	0.2176	0.2176	0.2176	0.2176	0.2176
	E_{full}	5.9417	0.2048	0.2048	0.2048	0.2048	0.2048	0.2048	0.2048	0.2048	0.2048	0.2048
	P	0.14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Continued on next page

Table 16 continued

Dataset	Metric	α										
		0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
	E_{true}	4.6314	7.3953	7.3953	7.3953	7.3953	7.3953	7.3953	7.3953	7.3953	7.3953	7.3953
	E_{full}	6.1648	7.4444	7.4444	7.4444	7.4444	7.4444	7.4444	7.4444	7.4444	7.4444	7.4444
KS	P	0.12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.21	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	0.2984	0.0312	0.0312	0.0312	0.0312	0.0312	0.0312	0.0312	0.0312	0.0312	0.0312
	E_{full}	0.1717	0.0366	0.0366	0.0366	0.0366	0.0366	0.0366	0.0366	0.0366	0.0366	0.0366
Allen–Cahn	P	0.33	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	25.771	0.2750	0.2750	0.2750	0.2750	42.504	42.504	42.504	42.504	83.576	83.576
	E_{full}	42.351	0.3146	0.3146	0.3146	0.3146	58.970	58.970	58.970	58.970	86.743	86.743
Chaffee–Infante	P	0.16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	0.67	0.67	0.67	0.67	0.33	0.33
	F1	0.27	1.00	1.00	1.00	1.00	0.80	0.80	0.80	0.80	0.50	0.50
	E_{true}	6.4780	3.1934	0.8906	0.8906	0.8906	0.8906	0.8906	0.8906	0.8906	0.8906	0.8906
	E_{full}	611.29	3.4336	0.9441	0.9441	0.9441	0.9441	0.9441	0.9441	0.9441	0.9441	0.9441
Compound	P	0.09	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.17	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	9.5217	1.6862	1.6862	1.6862	1.6862	1.6862	59.857	59.857	59.857	59.857	59.857
	E_{full}	DIV.	2.7276	2.7276	2.7276	2.7276	2.7276	99.995	99.995	99.995	99.995	99.995
Burgers2D	P	0.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	0.50	0.50	0.50	0.50	0.50	0.50
	F1	0.20	1.00	1.00	1.00	1.00	1.00	0.67	0.67	0.67	0.67	0.67
	E_{true}	4.6096	0.3066	0.3066	0.3066	0.3066	0.3066	0.3066	0.3066	0.3066	0.3066	0.3066
	E_{full}	53.702	0.3552	0.3552	0.3552	0.3552	0.3552	0.3552	0.3552	0.3552	0.3552	0.3552
Adv.–diff. (2D)	P	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1	0.16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	E_{true}	5.2177	0.2153	0.2153	0.2153	0.2153	0.2153	0.2153	0.2153	34.967	69.744	69.744
	E_{full}	24.256	0.3012	0.3012	0.3012	0.3012	0.3012	0.3012	0.3012	52.883	74.789	74.789
NLS	P	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.67	0.33	0.33
	F1	0.07	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	0.50	0.50
	E_{true}	43.148	15.463	10.246	10.246	10.246	50.489	88.968	88.968	88.968	88.968	88.968
	E_{full}	143.86	34.478	12.614	12.614	12.614	64.583	90.063	90.063	90.063	90.063	90.063
λ–ω	P	0.04	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R	1.00	1.00	1.00	1.00	1.00	0.57	0.14	0.14	0.14	0.14	0.14
	F1	0.07	0.93	1.00	1.00	1.00	0.73	0.25	0.25	0.25	0.25	0.25

F.7 Effect of Local Window Scale

To investigate the impact of the local window size on CARE-PDE, we conduct a sensitivity analysis by varying the window sizes in both spatial and temporal dimensions. Specifically, we introduce a scaling factor s to adjust the initial and maximum window sizes simultaneously, and evaluate different settings with $s \in \{0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5\}$, where $s = 1$ corresponds to the default window configuration, as shown in Fig. 7. From the experimental results, excessively small windows contain limited spatiotemporal observations, which often makes it difficult to con-

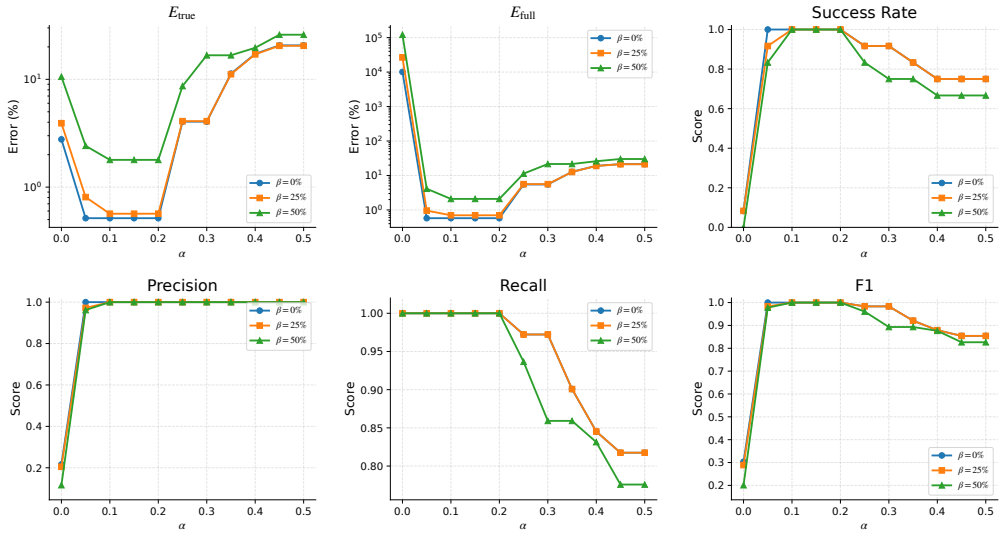

Figure 6: Effect of the complexity weight α in the equation scoring function on PDE discovery performance. The curves show the average results across 12 PDE systems under different NDEC scenarios.

struct sufficient valid windows that satisfy the reliability requirements. This issue becomes more pronounced as β increases. Even when some tasks can produce sufficient valid windows, limited local information may reduce the reliability of local projections and integral constraints, leading to inaccurate derivative features and unstable equation recovery. As the window size increases, local regions provide more effective samples for projection and constraint aggregation, thereby improving the stability of derivative estimation and reducing the influence of random perturbations. However, excessively large windows weaken the locality of the representation and may mix different dynamical behaviors across regions, resulting in degraded recovery performance. Overall, selecting an appropriate window size is essential for balancing local sample sufficiency and representation locality, and the default window configuration achieves a suitable trade-off.

F.8 Ablation Study

To examine the contribution of each component in CARE-PDE, we conduct a complete combinatorial ablation study over CRFR, AGRE, RGLP, LICA, and SSR. We conduct the ablation study by replacing modules. For field reconstruction, CRFR is replaced with the conventional ℓ_2 loss, while AGRE is disabled by assigning equal reliability to all observations. For derivative and constraint construction, RGLP is substituted with automatic differentiation (AD), and removing LICA reverts the formulation to pointwise strong-form features. For sparse identification, SSR is replaced by regression without stability selection. We evaluate all 2^5 module configurations under the same experimental protocol. Table 17 reports the results averaged over the 12 PDE systems, including coefficient errors, structural metrics, and the exact recovery rate.

The results show that the complete CARE-PDE consistently achieves the most stable overall performance, whereas replacing individual components results in varying degrees of degradation. The effects of replacing CRFR or removing AGRE become more evident as heterogeneous corruption increases, confirming the importance of robust reconstruction and reliability-aware weighting. Replacing RGLP with AD primarily affects derivative estimation, whereas removing LICA leads to a more pronounced decline in structural recovery, indicating that local projection and integral aggregation reduce sensitivity to local estimation errors.

Removing SSR has a relatively smaller effect under mild conditions but becomes more important as the perturbation level increases, reflecting its role in suppressing unstable candidate terms during structure selection. Performance further deteriorates when multiple components are replaced simul-

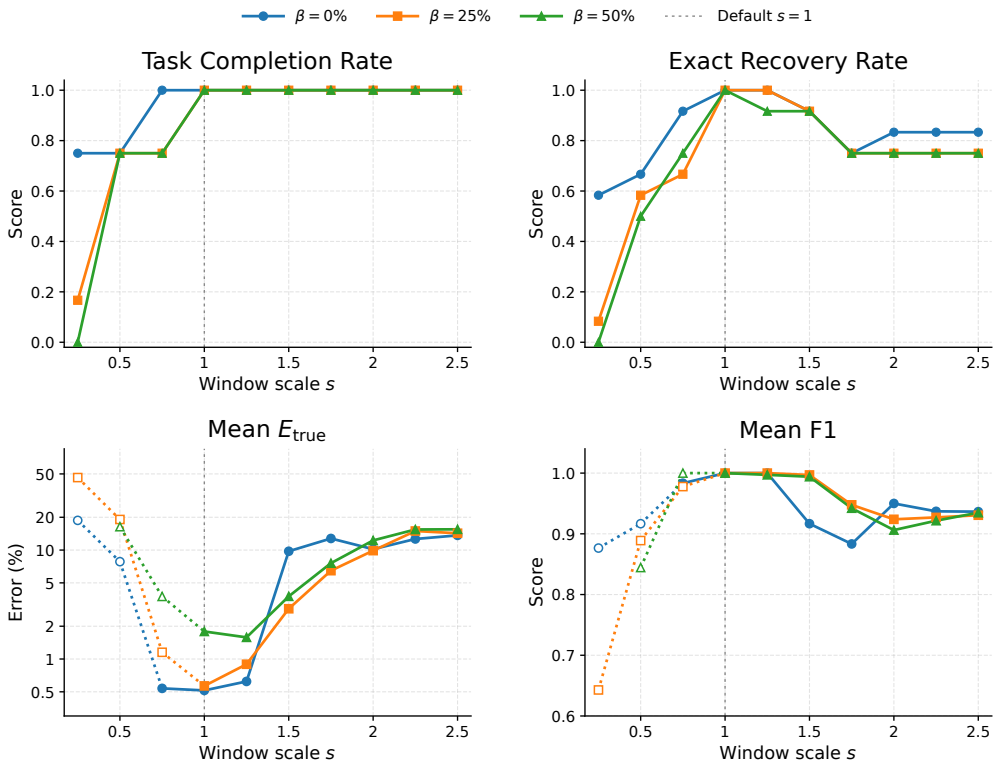

Figure 7: Effect of the local window scale on PDE discovery performance. Solid curves denote the averaged results over 12 PDE systems under different NDEC scenarios, while dotted curves represent the averaged results computed only on successfully completed systems.

taneously, indicating that the robustness of CARE-PDE arises from the synergistic contributions of its components, with each module playing an indispensable role in the overall framework.

Table 17: Complete ablation results of CARE-PDE over the 12 PDE systems under the NDEC settings considered in the main text. Checkmarks and crosses indicate whether each component is enabled or removed, respectively. SR denotes the exact recovery rate over the 12 systems. The full CARE-PDE configuration is shown in bold.

Components					Mean performance over 12 PDE systems					
CRFR	AGRE	RGLP	LICA	SSR	$E_{\text{true}}\downarrow$	$E_{\text{full}}\downarrow$	P↑	R↑	F1↑	SR↑
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 0\%)$										
✓	✓	✓	✓	✓	**0.5148**	**0.5763**	**1.00**	**1.00**	**1.00**	**100.00%**
×	✓	✓	✓	✓	0.6436	0.7488	1.00	1.00	1.00	100.00%
✓	×	✓	✓	✓	0.5148	0.5763	1.00	1.00	1.00	100.00%
✓	✓	×	✓	✓	8.6496	8.7578	0.92	0.92	0.92	91.67%
✓	✓	✓	×	✓	5.3545	5.5817	0.94	1.00	0.96	83.33%
✓	✓	✓	✓	×	0.5094	0.5712	1.00	1.00	1.00	100.00%
×	×	✓	✓	✓	0.6440	0.7467	1.00	1.00	1.00	100.00%
×	✓	×	✓	✓	8.7662	8.9203	0.92	0.92	0.92	91.67%
×	✓	✓	×	✓	5.9000	6.1440	0.92	1.00	0.95	75.00%
×	✓	✓	✓	×	0.6254	0.7347	1.00	1.00	1.00	100.00%
✓	×	×	✓	✓	8.6496	8.7578	0.92	0.92	0.92	91.67%
✓	×	✓	×	✓	5.3545	5.5817	0.94	1.00	0.96	83.33%

Continued on next page

Table 17 continued

Components					Mean performance over 12 PDE systems					
CRFR	AGRE	RGLP	LICA	SSR	$E_{\text{true}} \downarrow$	$E_{\text{full}} \downarrow$	P↑	R↑	F1↑	SR↑
✓	×	✓	✓	×	0.5094	0.5712	1.00	1.00	1.00	100.00%
✓	✓	×	×	✓	9.5491	10.588	0.88	0.92	0.89	83.33%
✓	✓	×	✓	×	8.6433	8.7519	0.92	0.92	0.92	91.67%
✓	✓	✓	×	×	4.4956	5.1097	0.91	1.00	0.94	83.33%
×	×	×	✓	✓	8.7668	8.9185	0.92	0.92	0.92	91.67%
×	×	✓	×	✓	5.8996	6.1423	0.92	1.00	0.95	75.00%
×	×	✓	✓	×	0.6253	0.7325	1.00	1.00	1.00	100.00%
×	✓	×	×	✓	9.7905	10.246	0.90	0.92	0.91	83.33%
×	✓	×	✓	×	8.7678	8.9244	0.92	0.92	0.92	91.67%
×	✓	✓	×	×	4.7104	5.3119	0.90	1.00	0.93	75.00%
✓	×	×	×	✓	9.5491	10.588	0.88	0.92	0.89	83.33%
✓	×	×	✓	×	8.6433	8.7519	0.92	0.92	0.92	91.67%
✓	×	✓	×	×	4.4956	5.1097	0.91	1.00	0.94	83.33%
✓	✓	×	×	×	9.2206	10.228	0.86	0.92	0.88	75.00%
×	×	×	×	✓	9.7906	10.246	0.90	0.92	0.91	83.33%
×	×	×	✓	×	8.7679	8.9225	0.92	0.92	0.92	91.67%
×	×	✓	×	×	4.7095	5.3101	0.90	1.00	0.93	75.00%
×	✓	×	×	×	9.7979	10.257	0.90	0.92	0.91	83.33%
✓	×	×	×	×	9.2206	10.228	0.86	0.92	0.88	75.00%
×	×	×	×	×	9.7976	10.257	0.90	0.92	0.91	83.33%
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 25\%)$										
✓	✓	✓	✓	✓	**0.5662**	**0.7005**	**1.00**	**1.00**	**1.00**	**100.00%**
×	✓	✓	✓	✓	1.8700	2.3478	1.00	1.00	1.00	100.00%
✓	×	✓	✓	✓	0.8065	0.9469	1.00	1.00	1.00	100.00%
✓	✓	×	✓	✓	8.8089	8.9980	0.92	0.92	0.92	91.67%
✓	✓	✓	×	✓	10.433	10.400	0.88	0.92	0.89	75.00%
✓	✓	✓	✓	×	0.5561	0.7027	1.00	1.00	1.00	100.00%
×	×	✓	✓	✓	6.7009	10.128	0.97	0.97	0.97	91.67%
×	✓	×	✓	✓	10.109	10.610	0.92	0.92	0.92	91.67%
×	✓	✓	×	✓	13.243	16.552	0.84	0.91	0.87	75.00%
×	✓	✓	✓	×	1.9847	2.9620	0.99	1.00	0.99	91.67%
✓	×	×	✓	✓	9.0867	9.2472	0.92	0.92	0.92	91.67%
✓	×	✓	×	✓	10.605	12.286	0.86	0.92	0.88	75.00%
✓	×	✓	✓	×	0.8449	0.9965	1.00	1.00	1.00	100.00%
✓	✓	×	×	✓	10.046	11.319	0.86	0.92	0.88	75.00%
✓	✓	×	✓	×	8.8051	8.9997	0.92	0.92	0.92	91.67%
✓	✓	✓	×	×	7.7440	8.3116	0.91	1.00	0.94	75.00%
×	×	×	✓	✓	14.940	18.390	0.89	0.89	0.89	83.33%
×	×	✓	×	✓	32.132	37.018	0.61	0.74	0.64	50.00%
×	×	✓	✓	×	6.8120	10.753	0.96	0.97	0.97	83.33%
×	✓	×	×	✓	17.134	18.122	0.85	0.85	0.83	75.00%
×	✓	×	✓	×	10.216	11.209	0.91	0.92	0.91	83.33%
×	✓	✓	×	×	12.171	16.304	0.85	0.97	0.89	66.67%
✓	×	×	×	✓	14.968	19.653	0.83	0.89	0.85	66.67%
✓	×	×	✓	×	9.1199	9.2968	0.92	0.92	0.92	91.67%
✓	×	✓	×	×	8.9998	10.223	0.87	1.00	0.91	66.67%
✓	✓	×	×	×	10.183	11.630	0.85	0.92	0.88	75.00%
×	×	×	×	✓	27.405	30.661	0.71	0.77	0.70	58.33%
×	×	×	✓	×	16.830	1378.46	0.83	0.89	0.85	66.67%
×	×	✓	×	×	23.716	5270.32	0.66	0.94	0.74	41.67%

Continued on next page

Table 17 continued

Components					Mean performance over 12 PDE systems					
CRFR	AGRE	RGLP	LICA	SSR	$E_{\text{true}} \downarrow$	$E_{\text{full}} \downarrow$	P↑	R↑	F1↑	SR↑
×	✓	×	×	×	14.876	18.419	0.84	0.91	0.86	75.00%
✓	×	×	×	×	11.748	13.350	0.85	0.92	0.88	75.00%
×	×	×	×	×	26.186	564.12	0.60	0.88	0.68	33.33%
$(\gamma_{\text{L}}, \gamma_{\text{H}}, \beta) = (0\%, 25\%, 50\%)$										
✓	✓	✓	✓	✓	**1.7895**	**2.0907**	**1.00**	**1.00**	**1.00**	**100.00%**
×	✓	✓	✓	✓	16.704	555.27	0.85	0.89	0.86	66.67%
✓	×	✓	✓	✓	5.7925	9.1967	0.97	0.97	0.97	91.67%
✓	✓	×	✓	✓	9.5331	9.8585	0.92	0.92	0.92	91.67%
✓	✓	✓	×	✓	14.861	19.450	0.89	0.90	0.89	75.00%
✓	✓	✓	✓	×	2.2300	3.9206	0.99	1.00	0.99	91.67%
×	×	✓	✓	✓	33.235	18182.94	0.74	0.79	0.75	58.33%
×	✓	×	✓	✓	23.071	561.87	0.81	0.84	0.81	58.33%
×	✓	✓	×	✓	31.644	472123.62	0.77	0.80	0.73	58.33%
×	✓	✓	✓	×	16.263	97309.27	0.83	0.96	0.86	75.00%
✓	×	×	✓	✓	13.611	17.038	0.89	0.89	0.89	83.33%
✓	×	✓	×	✓	21.773	31.839	0.79	0.89	0.82	66.67%
✓	×	✓	✓	×	10.267	17.754	0.95	0.97	0.96	91.67%
✓	✓	×	×	✓	18.546	26.288	0.85	0.86	0.85	66.67%
✓	✓	×	✓	×	9.9014	10.864	0.91	0.92	0.91	83.33%
✓	✓	✓	×	×	13.587	20.132	0.81	0.92	0.84	75.00%
×	×	×	✓	✓	41.418	17436.17	0.69	0.72	0.70	50.00%
×	×	✓	×	✓	47.905	2168.90	0.59	0.66	0.57	33.33%
×	×	✓	✓	×	26.550	58695.31	0.70	0.88	0.74	58.33%
×	✓	×	×	✓	34.288	717524.07	0.74	0.75	0.70	50.00%
×	✓	×	✓	×	23.129	97418.19	0.80	0.90	0.82	66.67%
×	✓	✓	×	×	28.919	518425.86	0.62	0.86	0.68	41.67%
✓	×	×	×	✓	24.884	27.737	0.82	0.79	0.77	66.67%
✓	×	×	✓	×	18.072	25.569	0.87	0.89	0.88	83.33%
✓	×	✓	×	×	23.005	73021.06	0.71	0.94	0.77	50.00%
✓	✓	×	×	×	14.239	21.589	0.78	0.92	0.82	66.67%
×	×	×	×	✓	47.353	2137.99	0.56	0.64	0.54	25.00%
×	×	×	✓	×	32.970	116010.51	0.67	0.82	0.70	50.00%
×	×	✓	×	×	42.350	172724.47	0.44	0.77	0.52	16.67%
×	✓	×	×	×	30.277	467105.57	0.59	0.80	0.63	33.33%
✓	×	×	×	×	25.354	69252.48	0.64	0.89	0.71	41.67%
×	×	×	×	×	42.468	14182.32	0.45	0.71	0.50	16.67%