Uncertainty-Penalized Bayesian Information Criterion for Parametric Partial Differential Equation Discovery

Pongpisit Thanasutives
Graduate School of Information Science and Technology
Osaka University
Suita, Osaka, 565-0871, Japan
thanasutives@ai.sanken.osaka-u.ac.jp

Ken-ichi Fukui
SANKEN (The Institute of Scientific and Industrial Research)
Osaka University
Ibaraki, Osaka, 567-0047, Japan
fukui@ai.sanken.osaka-u.ac.jp

Abstract

Data-driven discovery of partial differential equations (PDEs) has emerged as a promising approach for deriving underlying physics when domain knowledge about observed data is limited. Despite recent progress, the identification of governing equations and their parametric dependencies using conventional information criteria remains challenging in noisy situations, as the criteria tend to select overly complex PDEs. We introduce an extension of the uncertainty-penalized Bayesian information criterion (UBIC), which is adapted to solve parametric PDE discovery problems efficiently without computationally expensive PDE simulations. This extended UBIC uses quantified PDE uncertainty, accumulated across temporal or spatial points, to prevent overfitting in model selection. Numerical experiments on canonical PDEs show that our extended UBIC can identify the number of true terms and their varying coefficients accurately, even in the presence of noise.

1 Introduction

The discovery of governing partial differential equations (PDEs) through data-driven methods has been advanced significantly, with the development of sparse identification of nonlinear dynamics (SINDy) [1, 2]. These methods [3, 4] offer greater flexibility and accuracy by leveraging machine learning techniques on observed data, rather than relying on first-principles derivations. However, challenges remain, particularly in tuning regularization hyperparameters for sparse regression, which, if not properly done, can lead to the selection of overfitted or underfitted models. Recently, the uncertainty-penalized information criterion (UBIC) [5] has been developed to address the challenge by balancing model accuracy and model complexity while accounting for the quantified uncertainty of PDE coefficients to overcome the issue of overfitting. UBIC consistently outperforms traditional model selection criteria like AIC and BIC [6, 7], which tend to favor too complex models.

The difficulty increases when identifying parametric PDEs with spatially or temporally varying coefficients, as most state-of-the-art PDE discovery methods (e.g., [8, 9]) are designed for uncovering PDEs with constant coefficients and are thus not readily applicable for these tasks. Although existing approaches such as sequential grouped threshold ridge regression (SGTR) [10] and adaptive DLGA-PDE [11] offer possible solutions, the models they select are hyperparameter-sensitive. The adaptive DLGA-PDE also requires expensive computations, e.g., PDE solving/simulations or the learning

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

process to approximate a state variable. These limitations highlight the need for parsimony-enhanced approaches to achieve the computationally efficient, data-driven discovery of parametric PDEs.

Contributions. This paper introduces an extension of UBIC, adapted to solve parametric PDE discovery problems. The extended UBIC uses quantified PDE uncertainty as a complexity penalty to address the overfitting issue in model selection. It inherits the benefits of UBIC, including no computationally expensive PDE simulation required (different from SINDy-AIC [12]) and minimal dependence on hyperparameter tuning. Unlike any previous work, we provide confidence intervals for all coefficients, each evaluated at a particular time step or spatial grid point, as a byproduct of computing the UBIC.

2 Methodology

2.1 Problem Formulation

We consider the following parametric form of governing PDEs:

$$u_t = \mathcal{F}(u, u_x, u_{xx}, \ldots; \psi(x,t)) = \sum_{j=1} \mathcal{F}_j(u, u_x, u_{xx}, \ldots)\psi_j(x,t). \tag{1}$$

We aim to identify the nonlinear operator $\mathcal{F}$, which involves spatial derivatives of the state variable u, whose discretization $\mathbf{U} \in \mathbb{R}^{N_x \times N_t}$ on a spatio-temporal grid is given. $\mathcal{F}$ is parameterized by $\psi(x,t)$, which we assume reducible to either $\psi(x)$ or $\psi(t)$—spatially or temporally varying functions.

2.2 Best-subset Regression

Suppose, without loss of generality to spatially varying cases, Equation (1) is formulated as systems of linear equations, with temporal dependency. Given there are N_t time steps and N_x spatial points, the linear system evaluated at a time $t = t_i$ is expressed by

$$\mathbf{U}_{\boldsymbol{t}}^{\boldsymbol{i}} = \mathbf{Q}^{\boldsymbol{i}}\boldsymbol{\xi}^{\boldsymbol{i}} = \sum_{j=1}^{N_q} \xi_j^i \boldsymbol{q}_{\boldsymbol{j}}^{\boldsymbol{i}};\ \mathbf{Q}^{\boldsymbol{i}} = \begin{pmatrix} | & | & | & | \\ \boldsymbol{q}_{\mathbf{1}}^{\boldsymbol{i}} & \cdots & \boldsymbol{q}_{\boldsymbol{j}}^{\boldsymbol{i}} & \cdots \\ | & | & | & | \end{pmatrix} \in \mathbb{R}^{N_x \times N_q}. \tag{2}$$

$\mathbf{U}_{\boldsymbol{t}}$ is the first-order time derivative numerically computed with Kalman smoothing. Every $\mathbf{Q}^{\boldsymbol{i}}$ comprises overcomplete N_q candidate terms, each term potentially serving as a true $\mathcal{F}_j$. We define the candidate library $\mathbf{Q}$ as a block-diagonal matrix constructed by all $\mathbf{Q}^{\boldsymbol{i}}$ matrices, building a single system for the parametric PDE discovery problem: $\mathbf{U}_{\boldsymbol{t}} = \mathbf{Q}\boldsymbol{\Xi}$. We solve for the solution with s_k support size (the number of nonzero terms), satisfying

$$\hat{\boldsymbol{\Xi}} = \underset{\boldsymbol{\Xi}}{\arg\min} \sum_{i=1}^{N_t} \left\| \underline{\mathbf{U}}_{\boldsymbol{t}}^{\boldsymbol{i}} - \underline{\mathbf{Q}}^{\boldsymbol{i}}\boldsymbol{\xi}^{\boldsymbol{i}} \right\|_2^2 + \lambda \left\|\boldsymbol{\xi}^{\boldsymbol{i}}\right\|_2^2 \text{such that} \left\|\boldsymbol{\xi}^{\boldsymbol{i}}\right\|_0 = s_k, \forall k \leq N_s; \tag{3}$$

where $\underline{\mathbf{U}}$ and $\underline{\mathbf{Q}}$ are the validation data on which $\boldsymbol{\Xi} \in \mathbb{R}^{N_q N_t}$ (a tall column vector collecting every $\boldsymbol{\xi}^{\boldsymbol{i}}$) is evaluated. We set $\lambda = \frac{1}{N_t}\sum_{i=1}^{N_t} \left\| \underline{\mathbf{U}}_{\boldsymbol{t}}^{\boldsymbol{i}} - \underline{\mathbf{Q}}^{\boldsymbol{i}}\boldsymbol{\xi}_{\mathbf{LS}}^{\boldsymbol{i}} \right\|_2^2 / \left\|\boldsymbol{\xi}_{\mathbf{LS}}^{\boldsymbol{i}}\right\|_2^2$; where $\boldsymbol{\xi}_{\mathbf{LS}}^{\boldsymbol{i}}$ is the least-squares solution—leveraging all of the candidate terms, to balance between the residual sum of squares (RSS) loss and the L2-norm penalty. For each time step, the best-subset solver based on mixed-integer optimization (MIOSR) [13] is used to impose sparsity, gathering $\boldsymbol{\xi}^{\boldsymbol{i}}$ of consecutive support sizes with zero L2-norm penalty (not a sensitive hyperparameter). We prefer MIOSR over SGTR to ensure that potential PDEs with some support sizes are not overlooked. We achieve the group sparsity by controlling that the support set $\{j \mid \left|\xi_j^i\right| > 0\}$, is the same, say s_k, for every time step. Since we cannot infer the optimal number of nonzero terms solely from Equation (3), the model selection step is performed next.

2.3 Model Selection

We minimize an information criterion to select the optimal support size s^* in the strictly increasing sequence of all available support sizes, $(s_k)_{k=1}^{N_s}$. An information criterion is expressed by $-2\log L(\hat{\boldsymbol{\Xi}}) + \mathcal{C}(a_N, \hat{\boldsymbol{\Xi}}, \mathcal{P})$; where L is the likelihood function, and $\mathcal{C}(a_N, \hat{\boldsymbol{\Xi}}, \mathcal{P})$ is the total complexity penalty defined with a_N, a sequence of positive numbers. For example, $2s_k$ and $\log(N)s_k$; where $N = N_x N_t$, is the complexity penalty for AIC and BIC, respectively. $\mathcal{P}$ is any other necessary information, e.g., the complexity measures of ICOMP (informational complexity criterion) [14] or the UBIC's quantified PDE uncertainty. Considering a particular support size of s_k, we propose an extension of the original UBIC (incorporating a fixed threshold $\zeta = 10^{-5}$ to prevent underflowing) for the parametric PDE discovery as follows:

$$
\begin{gathered}
\text{UBIC} = N\log\left(\frac{2\pi}{N}\left\|\mathbf{U}_{\boldsymbol{t}} - \mathbf{Q}\hat{\boldsymbol{\Xi}}_{\boldsymbol{\mu}}\right\|_2^2 + \zeta\right) + \log(N)(\mathfrak{U} + s_k); \\
\mathfrak{U} = 10^{\lambda^*}\mathcal{V},\ \mathcal{V} = \frac{\text{V}}{\text{V}_{\max}},\ \text{V} = \Sigma_{i=1}^{N_t} R_i,\ \text{and}\ R_i = \frac{\Sigma_{j=1}^{s_k}\sigma_j^i}{\left\|\hat{\boldsymbol{\xi}}_{\boldsymbol{\mu}}^{\boldsymbol{i}}\right\|_1}.
\end{gathered}
\tag{4}
$$

According to [5], we compute the uncertainty $\mathfrak{U}$ of the s_k-support-size PDE using the tuned data-dependent λ^* and the scaled coefficient of variation $\mathcal{V}$. At each time step, V accumulates an instability ratio R_i, defined as the total posterior standard deviation divided by the L1-norm of the posterior mean coefficient vector. Both the posterior covariance matrix ($\in \mathbb{R}^{s_k \times s_k}$) and mean coefficient vector ($\in \mathbb{R}^{s_k}$) are obtained using Bayesian automatic relevance determination (ARD) regression [15]. $\text{V}_{\max}$ is the maximum value of V over all available support sizes. With the temporal (or spatial) accumulation, we essentially derive the extended UBIC for the parametric PDE discovery. After the best PDE has been decided, a physics-informed neural network [16] or a Fourier neural operator [17] may be employed to simulate the state variable, on which UBIC is calculated to additionally verify the validity of the equation.

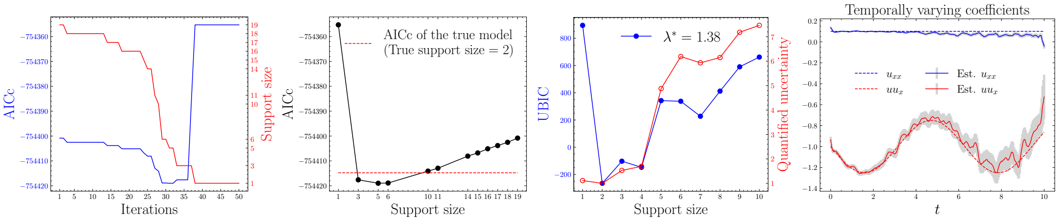
Spectral density based transformation. The validation data $\underline{\mathbf{Q}}$ in frequency space is obtained by applying discrete Fourier transformation over the temporal axis to every $\mathbf{Q}_{\boldsymbol{j}} = \begin{pmatrix} | & | & | & | \\ \boldsymbol{q}_{\boldsymbol{j}}^{\mathbf{1}} & \cdots & \boldsymbol{q}_{\boldsymbol{j}}^{\boldsymbol{i}} & \cdots \\ | & | & | & | \end{pmatrix} \in \mathbb{R}^{N_x \times N_t}$, and removing entries corresponding to low-power frequencies—less than the ninety percentile. The transformation is beneficial not only when deciding the optimal coefficient vector with s_k support size, but also when selecting the optimal s^*. We generalize the RSS to $\left\|T(\mathbf{U}_{\boldsymbol{t}}) - T(\mathbf{Q}\hat{\boldsymbol{\Xi}}_{\boldsymbol{\mu}})\right\|_2^2$; where T transforms $\mathbf{U}_{\boldsymbol{t}}$ and $\mathbf{Q}\hat{\boldsymbol{\Xi}}_{\boldsymbol{\mu}}$ to new representations, i.e., mapping $T(\mathbf{U}_{\boldsymbol{t}}) = \tilde{\mathbf{U}}_{\boldsymbol{t}}$. Every $\tilde{\mathbf{U}}_{\boldsymbol{t}}^{\boldsymbol{i}}$ is a numerical result from a trapezoidal integration applied along the spatial axis (the frequency/temporal axis for a spatially-dependent PDE) of estimated power spectral density (PSD) using a periodogram. The integration limits the sample number and therefore facilitates the model selection step, as [18, 5] have shown that conventional information criteria tend to select overfitted PDEs when the number of samples is large. The PSD representation is noise-tolerant with its clear characteristics, exhibiting larger values for true data-generating frequencies (see Appendix A.3). The integration is ablated if the estimated PSD is a one-dimensional vector.

3 Results and Discussion

As detailed in Table 1, we experiment with three canonical parametric PDEs: the time-dependent Burgers' equation, the spatially-dependent advection-diffusion (AD) PDE, and the spatially-dependent chaotic Kuramoto-Sivashinsky (KS) PDE. Each entry of the noise-free simulated solution $\mathbf{U}$ is perturbed with ϵ%-sd (standard deviation) Gaussian noise sampled from $\frac{\epsilon}{100} \times \text{sd}(\mathbf{U}) \times \mathcal{N}(0, 1)$. The noise levels are listed in Table 1. We apply a Savitzky-Golay filter to smooth the resulting distorted data before computing partial derivatives. We refer readers to [5] for a discussion on the positive effects of the data denoising. The candidate terms include powers of u up to the cubic degree, which are multiplied by spatial derivatives of u up to the fourth order. All experiments were run

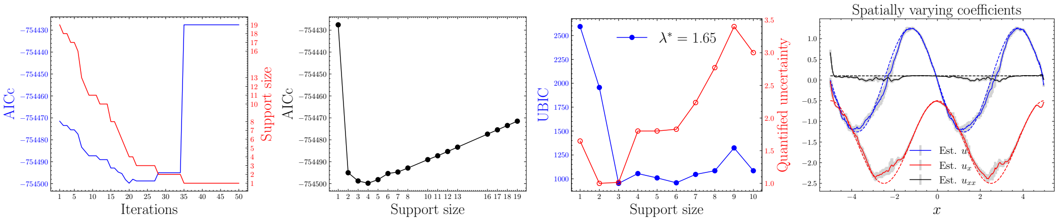
Table 1: Parametric PDE datasets from [10]. These datasets are visualized in Appendix A.1.

Dataset	PDE	Varying coefficient	N_x, N_t	ϵ
Burgers	$u_t = a(t)uu_x + 0.1u_{xx}$ $x \in [-8, 8]$ and $t \in [0, 1]$	$a(t) = -(1 + \frac{\sin(t)}{4})$	$256, 256$	4
AD	$u_t = a'(x)u + a(x)u_x + 0.1u_{xx}$ $x \in [-5, 5]$ and $t \in [0, 5]$	$a(x) = -1.5 + \cos(\frac{2\pi x}{5})$	$256, 256$	4
KS	$u_t = a(x)uu_x + b(x)u_{xx}$ $+c(x)u_{xxxx}$ $x \in [-20, 20]$ and $t \in [0, 100]$	$a(x) = 1 + 0.25\sin(\frac{2\pi x}{20})$ $b(x) = -1 + 0.25e^{-\frac{(x-2)^2}{5}}$ $c(x) = -1 - 0.25e^{-\frac{(x+2)^2}{5}}$	$512, 512$	10^{-2}



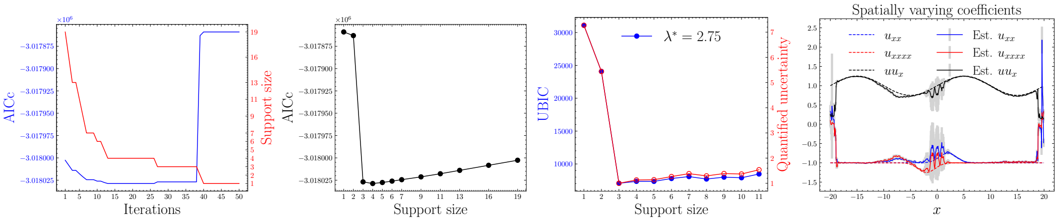
(a) SGTR (Baseline) (b) Ours (c) Identified coefficients

Figure 1: **Burgers' PDE.**



(a) SGTR (Baseline) (b) Ours (c) Identified coefficients

Figure 2: **Advection-diffusion PDE.** Dashed lines in (c) denote the true spatially varying coefficients.



(a) SGTR (Baseline) (b) Ours (c) Identified coefficients

Figure 3: **Kuramoto-Sivashinsky PDE.**

on an Intel i7 CPU with 32 GB of RAM. The code is publicly available at `https://github.com/Pongpisit-Thanasutives/parametric-discovery`.

We compare our method with the widely adopted SGTR baseline, which evaluates models using corrected AIC (AICc) [19] for finite sample sizes, under noisy situations. In Figures 1(a), 2(a), and 3(a), although SGTR converges, it fails to explore certain support sizes, including the true one of the Burgers' PDE, raising concerns about how SGTR imposes sparsity through hard thresholding. The AICc losses have led to the selection of too complex or overfitted models. In contrast, our UBIC, calculated with the PSD-based transformation, utilizes the quantified uncertainty to penalize overfitted models and identifies the correct governing equations despite the high noise levels, consistently outperforming the SGTR baseline, as shown in Figures 1(b), 2(b), and 3(b). From the three experiments, the wall-clock time for our model selection step, which is performed after applying the best-subset regression, is approximately 10 to 20 seconds. Following the model selection results by our UBIC, Figures 1(c), 2(c), and 3(C) present the posterior coefficients with twice their standard deviation, representing about the 95% confidence intervals. These intervals demonstrate regions (in space or time) where the instability in estimating the posterior coefficients is relatively high, offering insights that can further improve the PDE discovery method by circumventing these unstable regions. In Appendix A.2, we uncover symbolic expressions for the varying coefficients.

4 Conclusion

Our main contribution is the extended UBIC for identifying governing parametric PDEs. The extended UBIC, computed with the PSD-based transformation, leverages accumulated PDE uncertainty to overcome the overfitting problem in the model selection step, disambiguating the true governing parametric PDE from overfitted PDEs with unnecessary candidate terms. The ability to compute confidence intervals for varying coefficients enhances the interpretability of potential models, providing comprehensive insights into their stability. Since a failure to completely include true terms, which dictate the dynamics, in the candidate library can degrade PDE discovery results, as discussed in Appendix A.4, we plan to use the proposed UBIC as the fitness function in a genetic-algorithm-based PDE discovery framework [11], eliminating the limitation imposed by the overcompleteness assumption and thus allowing us to better tackle real-world ODE/PDE discovery problems.

References

[1] 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.

[2] 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.

[3] Saso Dzeroski and Ljupco Todorovski. Discovering dynamics: From inductive logic programming to machine discovery. *Journal of Intelligent Information Systems*, 4(1):89–108, 1995.

[4] Michael Schmidt and Hod Lipson. Distilling Free-Form Natural Laws from Experimental Data. *Science*, 324(5923):81–85, 2009.

[5] Pongpisit Thanasutives, Takashi Morita, Masayuki Numao, and Ken-ichi Fukui. Adaptive Uncertainty-Penalized Model Selection for Data-Driven PDE Discovery. *IEEE Access*, 12:13165–13182, 2024.

[6] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[7] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

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

[9] Jens Berg and Kaj Nyström. Data-driven discovery of PDEs in complex datasets. *Journal of Computational Physics*, 384:239–252, 2019.

[10] Samuel Rudy, Alessandro Alla, Steven L. Brunton, and J. Nathan Kutz. Data-Driven Identification of Parametric Partial Differential Equations. *SIAM Journal on Applied Dynamical Systems*, 18(2):643–660, 2019.

[11] Hao Xu, Dongxiao Zhang, and Junsheng Zeng. Deep-learning of parametric partial differential equations from sparse and noisy data. *Physics of Fluids*, 33(3):037132, 03 2021.

[12] N. M. Mangan, J. N. Kutz, S. L. Brunton, and J. L. Proctor. Model selection for dynamical systems via sparse regression and information criteria. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2204):20170009, 2017.

[13] Dimitris Bertsimas and Wes Gurnee. Learning sparse nonlinear dynamics via mixed-integer optimization. *Nonlinear Dynamics*, 111(7):6585–6604, 2023.

[14] Hamparsum Bozdogan and Dominique M.A. Haughton. Informational complexity criteria for regression models. *Computational Statistics & Data Analysis*, 28(1):51–76, 1998.

[15] David JC MacKay et al. Bayesian nonlinear modeling for the prediction competition. *ASHRAE transactions*, 100(2):1053–1062, 1994.

[16] 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.

[17] Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier Neural Operator for Parametric Partial Differential Equations. In *International Conference on Learning Representations*, 2021.

[18] Pongpisit Thanasutives, Takashi Morita, Masayuki Numao, and Ken-ichi Fukui. Noise-aware physics-informed machine learning for robust PDE discovery. *Machine Learning: Science and Technology*, 4(1):015009, Feb 2023.

[19] Clifford M. Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 06 1989.

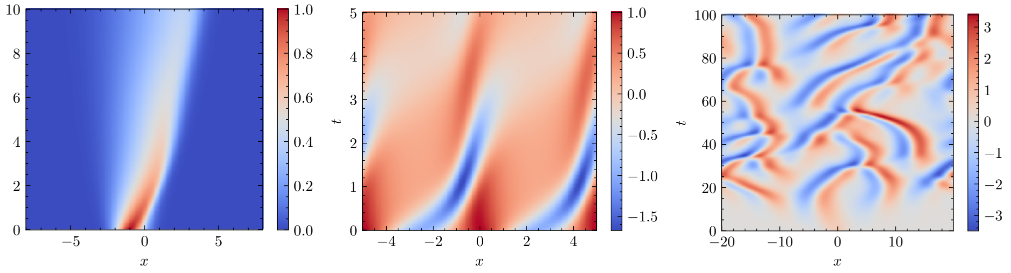
[20] Miles Cranmer. Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl, 2023.

[21] Kevin René Broløs, Meera Vieira Machado, Chris Cave, Jaan Kasak, Valdemar Stentoft-Hansen, Victor Galindo Batanero, Tom Jelen, and Casper Wilstrup. An approach to symbolic regression using feyn. *arXiv preprint arXiv:2104.05417*, 2021.

A Appendix

A.1 Dataset Visualization

An illustration of each dataset experimented in this paper is provided in Figure 4.



(a) **Burgers' PDE.** (b) **Advection-diffusion PDE.** (c) **Kuramoto-Sivashinsky PDE.**

Figure 4: Two-dimensional visualization of the noiseless datasets.

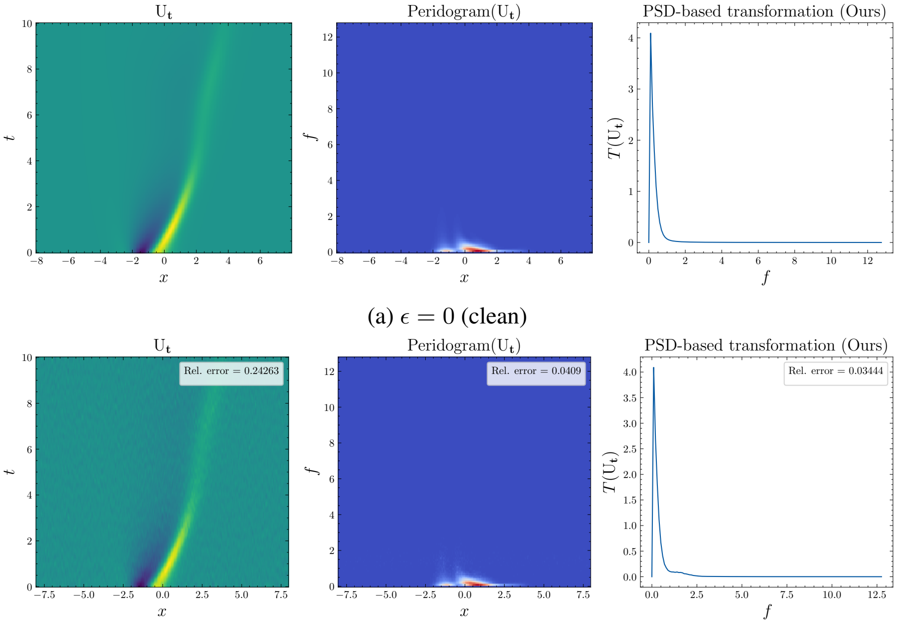
A.2 Symbolic Discovery of Varying Coefficients

Symbolic discovery of varying coefficients is achieved via the PySR package [20]. To prioritize parsimonious expressions of varying coefficients, we consider model rankings based on the PySR's score. We evaluate any selected interpretable expression $\hat{h}(x)$ against its ground truth $h(x)$ using the percentage relative coefficient error: $\mathrm{CE}(h(x), \hat{h}(x)) = 100 \times \frac{\left\|\hat{h}(x) - h(x)\right\|_1}{\left\|h(x)\right\|_1}$. We note that $\mathrm{CE}(h(t), \hat{h}(t))$ are calculated in the same manner. Table 2 lists the relative coefficient errors for every experiment in this paper. In the experiments of the parametric Burgers' and AD PDEs, we can uncover

Table 2: Symbolic expression of varying coefficients

Dataset	Symbolically discovered varying coefficient	%Coefficient error
Burgers	$a(t) = -0.25146\sin(t) - 0.95987, 0.08080$	$3.813, 19.20$
AD	$a'(x) = -\sin(1.2415x)$ $a(x) = \cos(1.244x) - 1.4216$ 0.06406	20.06 4.554 35.94
KS	$0.25113\sin(0.32253x) + 0.95955$ $-0.976754 + 0.353986e^{-0.493521(1-0.547288x)^2}$ $-0.966723 - 0.249388e^{-3.73627(0.429437x+1)^2}$	4.055 4.027 4.365

the correct mathematical expressions/structures of the varying coefficients with acceptable accuracy. For the parametric KS PDE case, we cannot initially retrieve the correct expression only for the varying coefficient of u_{xxxx}, as the suggested expression by PySR is $-0.9798+0.19909xe^{-0.26832x^2}$, which however hints that a common Gaussian function should be used instead due to its similar accuracy. We refine the initial expressions for u_{xx} and u_{xxxx} using Feyn's autorun functionality [21] with minimal complexity settings, ultimately discovering the Gaussian formulas with satisfactory coefficient errors of less than 5%. Since the SGTR baseline failed to identify the true governing equations in all of the three numerical experiments, quantifying its coefficient error is not applicable.



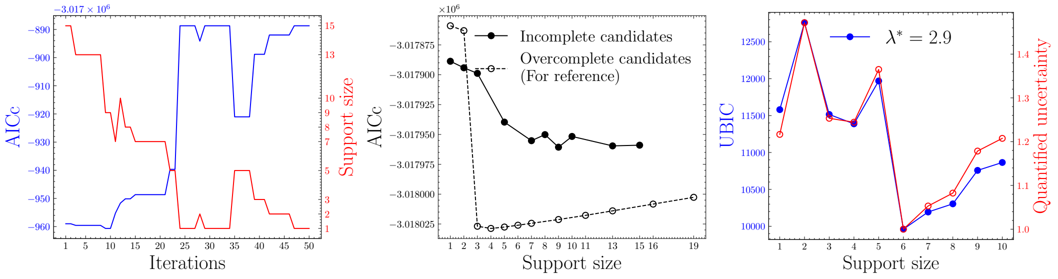
(a) $\epsilon = 0$ (clean)

(b) $\epsilon = 4$ (with Gaussian noise)

Figure 5: **Burgers' PDE.** We visualize different representations of $\mathbf{U}_t$ and measure their relative errors against the ground truth data.

A.3 Noise-robustness of PSD

We explore the noise-robustness of our PSD-based transformation using the parametric Burgers' PDE as an example. In Figure 5, different representations of $\mathbf{U}_t$ are presented. We evaluate the accuracy of each representation by comparing it to the ground truth data using the relative Frobenius-norm error. The fact that our PSD-based transformed representation matches its ground truth more closely than other representations under the noisy situation demonstrates its superior robustness. Therefore, we apply our extended UBIC with the PSD-based transformation.



(a) SGTR (Baseline)

(b) Ours

Figure 6: Parametric PDE discovery of the KS equation with incomplete candidate terms.

A.4 Parmetric PDE Discovery with Incomplete Candidate Terms

To better understand the negative impact of incomplete candidate terms on the parametric PDE discovery, we reduce the maximum derivative order considered in the KS experiment from 4 to 3. Subsequently, we reperform both the SGTR baseline and our proposed method. Figure 6(a) clearly shows that the potential PDEs or the best subsets are of inferior AICc when the candidate terms are incomplete. Although the UBIC values from the two cases——overcomplete and incomplete—cannot be directly compared because of the difference in the tuned λ^* values, our UBIC selects a PDE with increased complexity to compensate for the true derivative term omitted from the candidate library, as illustrated in Figure 6(b). Therefore, we stress the importance of extending our proposed UBIC to accommodate other PDE discovery algorithms that leverage more flexible structures, such as evolving genomes [11] or tree expressions, to encode PDEs.

B Broader Impacts

The uncertainty-penalized information criterion developed in this paper is developed to inspire the future advancement of parsimony-enhanced information criteria. Such criteria can be seamlessly integrated with any state-of-the-art PDE discovery methods to disregard incorrect governing PDEs that arise from the overfitting or underfitting in the model selection step. Our contribution holds promise for advancing interdisciplinary research and enhancing the efficiency of model discovery processes, leading to a deeper understanding of complex physical phenomena across a broad spectrum of scientific disciplines. We see no negative societal impacts of the work performed.