



System identification for complex dynamical systems: a survey

Xiaoyu Zhang¹, Ziqi Li¹, Ruizhe Shi¹, Shiyu Wang¹, Feng Wang^{2,3}, Duxin Chen^{1,4}

Keywords:

System identification, complex dynamical systems, sparse Bayesian learning, recursive identification, multi-system identification, equation discovery, large language models

Citation: Zhang, X.; Li, Z.; Shi, R.; Wang, S.; Wang, F.; Chen, D. System identification for complex dynamical systems: a survey. *Complex Eng. Syst.* 2026, 6, 13. <http://dx.doi.org/10.20517/ces.2026.12>

Received: 11 Mar 2026

First Decision: 30 Apr 2026

Revised: 10 May 2026

Accepted: 25 May 2026

Published: 16 Jul 2026

Academic Editor:

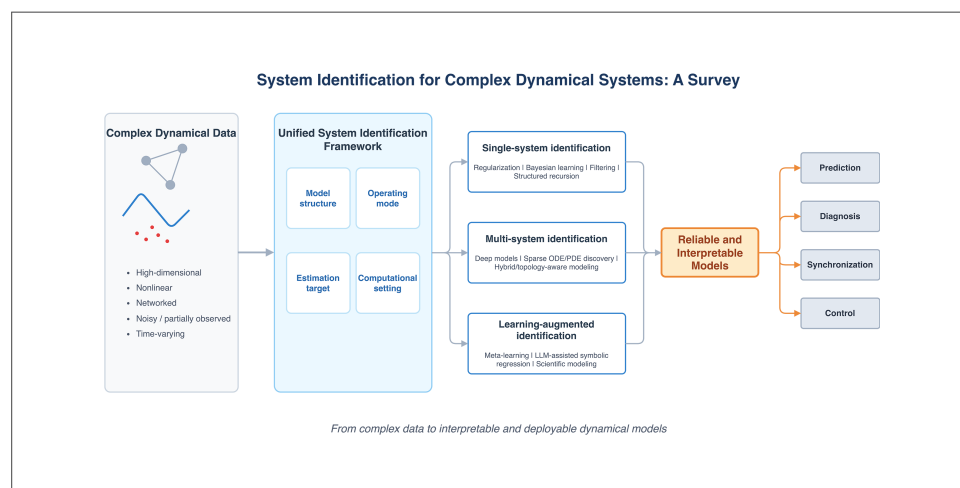
Feng Ding

Copy Editor:

Fangling Lan

Production Editor:

Fangling Lan



Abstract

System identification provides the data-to-model link for analysis, prediction, diagnosis, and control of dynamical systems. As engineering and scientific systems become increasingly nonlinear, high-dimensional, networked, time-varying, and partially observed, identification methods must handle noisy data, incomplete prior knowledge, safety constraints, and downstream control requirements. This review surveys system identification for complex dynamical systems through a unified framework that connects model structure, operating mode, estimation target, and computational setting. We first revisit single-system identification, including regularized regression, sparse Bayesian learning, Kalman-type filtering, and recent structured recursive methods such as auxiliary-model, multi-innovation, hierarchical, filtering-based, and coupled identification. We then review multi-system identification for families of interacting or related subsystems, covering deep neural architectures, sparse equation discovery for ordinary and partial differential equations, hybrid and topology-aware modeling, and meta-learning-based fast adaptation. Finally, we discuss large language model-assisted workflows for symbolic regression, equation discovery, feature generation, and sci-

¹School of Mathematics, Southeast University, Nanjing 211189, Jiangsu, China.

²Aliyun School of Big Data, Changzhou University, Changzhou 213164, Jiangsu, China.

³School of Computer Engineering, Suzhou Polytechnic University, Suzhou 215131, Jiangsu, China.

⁴Jiangsu Province Center for Applied Mathematical Sciences, Nanjing 211189, Jiangsu, China.

Correspondence to: Prof. Feng Wang, Aliyun School of Big Data, Changzhou University, Changzhou 213164, Jiangsu, China; School of Computer Engineering, Suzhou Polytechnic University, 106 Zhineng Ave., Suzhou 215131, Jiangsu, China. E-mail: wfeng@cczu.edu.cn

entific modeling. By comparing these methodological streams, this review highlights their strengths, limitations, and applicability to prediction, synchronization, diagnosis, and control. We identify key open challenges, including robust identification under biased or limited data, uncertainty-aware modeling, interpretable learning, reproducible benchmarking, and tighter integration between identification, experiment design, and control synthesis.

1. INTRODUCTION

In modern dynamical systems and control engineering, reliable mathematical models are indispensable for analysis, prediction, optimization, and feedback synthesis. Models underpin tasks ranging from stability assessment and performance verification to fault diagnosis and supervisory decision-making. Yet for many complex plants, first-principles descriptions are incomplete, uncertain, or only partially known, and system parameters may drift during nominal operation. These limitations are compounded by noise, unmodeled dynamics, and constraints in actuation and sensing. Consequently, there is a growing need for disciplined data-to-model pipelines that can extract faithful representations of system behavior from experimental or operational data across laboratory micro-experiments, industrial cyber-physical systems, and natural-science applications. System identification is precisely the discipline that addresses this need: it provides the methodological bridge from raw data to usable models, and thereby underlies the subsequent design of estimators and controllers.

Within this broader data-to-model-and-control pipeline, system identification, state estimation, and control form an interdependent triad. Control design requires predictive and certifiable models; identification extracts such models from data; and estimation provides latent states and uncertainty quantification for analysis and feedback. Among these three components, system identification plays a foundational role, as it determines the model structures and parameterizations on which estimation and control must operate. At the same time, the components are tightly coupled: identifiability and excitation conditions constrain what models can be learned; estimator structure and performance depend on model fidelity; and control architectures, in turn, shape the data available for identification. Understanding these interdependencies is essential for designing end-to-end workflows that remain robust under realistic experimental and operational constraints.

Against this backdrop, recent surveys delineate several important directions in modern system identification. A deep-learning-focused review by^[1] connects neural architectures to over-parameterization, regularization, stability, and multi-step prediction, clarifying how learning-based models reshape classical bias-variance considerations. A recent network-era survey^[2] emphasizes the challenges posed by missing data, outliers, adversarial disturbances, and distributed/edge data acquisition in large-scale cyber-physical systems. Earlier overview articles, such as^[3], cataloged a broad range of traditional parametric and intelligent techniques, while classical foundational work established the basic links among model classes, excitation conditions, identification criteria, and control design^[4]. Taken together, these surveys describe the broad historical landscape, but they do not yet fully capture several rapidly developing structured identification lines within the core recursive system identification literature.

In particular, recent system identification research has seen rapid advances in structured recursive and iterative methods for multivariable, nonlinear, and colored-noise systems. Representative advances include auxiliary-model identification for handling unmeasurable intermediate variables or internal signals, multi-innovation identification for improving data utilization and estimation accuracy, hierarchical identification for reducing computational complexity via model decomposition, filtering-based identification for mitigating bias induced by colored noise, and coupled or partially-coupled identification for exploiting shared information structures in multivariable systems. Importantly, these ideas are increasingly combined rather than used in isolation. Recent studies have reported auxiliary model-based maximum-likelihood multi-innovation recursive least-squares

methods for multiple-input multiple-output systems^[5], filtered hierarchical identification algorithms for multiple-input multiple-output (MIMO) systems with colored noises^[6], distributed partially-coupled recursive generalized extended least-squares methods for multivariate output-error systems^[7], hierarchical maximum-likelihood multi-innovation methods for Hammerstein-type nonlinear systems^[8], dynamic-factor and multi-innovation-based Elman-network modeling from measurements^[9], and two-stage iterative algorithms for Hammerstein nonlinear systems with non-uniform sampling^[10]. These developments indicate that a contemporary survey should not only revisit classical foundations and learning-based approaches, but also reflect the rapid evolution of structured recursive identification methods within the mainstream system identification literature.

The motivation for this review arises from both methodological and practical changes in contemporary system identification. Methodologically, the field has expanded from classical prediction-error and state-space estimation toward a diverse set of approaches, including regularized and Bayesian sparse regression, structured recursive identification, sparse equation discovery, deep neural modeling, meta-learning, and large language model (LLM)-assisted symbolic modeling. Practically, the target systems have also changed: complex engineering systems are increasingly nonlinear, high-dimensional, networked, heterogeneous, partially observed, and operated under safety and actuation constraints. As a result, identification methods must not only fit data, but also support downstream objectives such as prediction, synchronization, diagnosis, optimization, and control. Although existing surveys have made important contributions by reviewing classical methods, deep-learning-based identification, and network-era identification challenges^[1,2,3], they do not yet provide an integrated framework that simultaneously covers structured recursive methods, multi-system identification, physics–data equation discovery, fast adaptation across related systems, and emerging LLM-assisted workflows. This fragmentation motivates the present review, which aims to clarify how different identification paradigms are connected, what assumptions and limitations they involve, and how they can be selected or combined for complex dynamical systems.

More specifically, existing surveys still leave several practice-critical gaps. First, there is still a need for an end-to-end, practice-oriented organizing framework that explicitly ties model structure, operating mode, estimation target, and computational setting to downstream control objectives and constraints. Second, structured recursive identification methods for multivariable, nonlinear, and colored-noise systems are often discussed in specialized algorithmic papers, but have not been sufficiently integrated into broader survey-level accounts. Third, existing reviews treat system families only tangentially, with limited systematic coverage of multi-system identification, in which related subsystems share structure but differ in parameters or regimes. Fourth, scalable physics–data bridging remains fragmented: sparse/weak/integral equation discovery, probabilistic/Bayesian uncertainty quantification, and experiment design under safety/actuation limits are usually discussed in isolation rather than as parts of a unified pipeline. Fifth, rapid cross-task adaptation via meta-learning has only recently begun to appear in the identification literature and is not yet integrated into survey-level treatments. Finally, emerging LLM-assisted workflows for symbolic regression and modeling lack consolidated guidance on validation, reproducibility, and safe deployment.

To make the distinctive scope of this review more explicit, Table 1 summarizes the main coverage of representative survey perspectives and highlights the broader scope of the present review.

This review aims to address these gaps. We adopt a concise organizing framework that views identification as a set of interconnected design choices—model structure (black/grey/white; parametric *vs.* nonparametric), operating mode (open *vs.* closed loop), estimation target (one-step-ahead prediction *vs.* free-run simulation), and computational setting (offline *vs.* online). This perspective helps articulate how bias–variance–robustness trade-offs, informative experiment design, and uncertainty reporting interact in practice. Within this framework, we first develop single-system identification as the foundation, covering prediction-error criteria, regularization and sparsity for parsimony and interpretability, recursive and augmented-state filtering for joint state–parameter

Table 1. Comparison of representative survey perspectives and the present review

Study or perspective	Main focus and coverage
Classical system-identification overviews ^[3,4]	Classical model classes; excitation conditions; identification criteria; traditional parametric and intelligent methods; links with control design
Deep-learning-focused system-identification surveys ^[1]	Neural architectures; over-parameterization; regularization; stability; multi-step prediction; learning-based model structures
Network-era system-identification studies ^[2]	Missing data; outliers; adversarial disturbances; distributed data acquisition; edge data acquisition; large-scale cyber–physical systems
Present review	Unified data-to-model framework; single-system identification; multi-system identification; structured recursive identification; sparse and Bayesian learning; PDE-oriented equation discovery; meta-learning adaptation; LLM-assisted system identification; links to prediction, diagnosis, synchronization, optimization, and control

tracking, and recent structured recursive identification methods such as auxiliary-model, multi-innovation, hierarchical, filtering-based, and coupled identification. We then extend the discussion to multi-system identification, where families of related subsystems share structure but differ in parameters or regimes, with an emphasis on deep learning, topology-aware regression, sparse equation discovery, and hybrid dynamical modeling in data-rich settings. Building on these two layers, we further synthesize learning-augmented developments that bridge physics and data—including sparse libraries and weak/integral equation-discovery variants, probabilistic/Bayesian formulations for principled uncertainty, meta-learning for rapid cross-task adaptation, and LLM-assisted symbolic regression—and highlight when these tools improve sample efficiency, noise robustness, and downstream controllability. Throughout, we foreground reproducible practice: input and experiment design under safety/actuation limits, loss selection aligned with end-use objectives, diagnostic validation via residual analysis and information criteria or Pareto-front analyses, and transparent reporting of parameter and state uncertainty for risk-aware control.

To improve readability and help readers locate representative studies, Table 2 summarizes the main research directions, representative researchers or works, and their roles in this review.

The symbols used in this paper are summarized in Table 3.

Figure 1 provides a roadmap of the method landscape and illustrates how the three main parts of this survey relate to the organizing framework.

Building on this motivation, the remainder of the paper is organized as follows. Section 2 (Single-System Identification) introduces foundational formulations and criteria, and reviews representative algorithms including sparsity-driven regression (e.g., Orthogonal matching pursuit (OMP)/The least absolute shrinkage and se-

Table 2. Representative literature map for system identification of complex dynamical systems

Research direction	Representative researchers or works	Representative contribution
Classical foundations	Åström; Kalman; Ljung; sub-space identification	Prediction-error principles; state estimation; excitation conditions; state-space realization
Sparse and regularized identification	Tibshirani; Tipping; Wipf; Brunton	LASSO; RVM; sparse Bayesian learning; sparse nonlinear dynamics
Structured recursive identification	Ding; Wang; Xing; Liu; Qin and Ji	Auxiliary-model, multi-innovation, hierarchical, filtering-based, and coupled identification
ODE and PDE equation discovery	Brunton; Rudy; Schaeffer; Fasel; Both; Chen	SINDy; PDE-FIND; weak formulations; ensemble methods; PINN-assisted discovery
Multi-system and hybrid identification	Han; Yuan; topology-aware and hybrid-system identification studies	Network reconstruction; regime-dependent dynamics; switching logic
Meta-learning-based identification	Finn; Lin; Iwata	Fast adaptation across related systems; task-level transfer; Koopman-based meta-learning
LLM-assisted identification	Shojaee; Du; Grayeli; Wang; Taskin; Yang	LLM-assisted symbolic regression; equation discovery; feature generation; scientific modeling

Table 3. Mathematical symbols used in this review

Symbol	Meaning
$\mathcal{D} = \{u(t), y(t)\}_{t=1}^N$	Input–output data record of length N
$u(t), y(t)$	Input and output at time t
$\hat{y}(t t-1; \theta)$	One-step-ahead prediction under model parameters θ
$\mathcal{M}(\theta)$	Model class/structure parameterized by θ
θ	Model parameters to be estimated
$\mathcal{E}$	Hyperparameters (e.g., regularization, kernel, model order)
$J(\theta; \mathcal{D})$	Identification objective (loss) on data $\mathcal{D}$
$\ell(\cdot, \cdot)$	Data-fit loss (e.g., squared error / negative log-likelihood)
λ	Regularization weight
$\mathcal{R}(\theta)$	Regularization term controlling model complexity
$\varphi(k)$	Regressor vector at sample index k
$\Theta(\mathbf{X})$	Library/dictionary matrix constructed from data $\mathbf{X}$
Ξ	Sparse coefficient matrix (e.g., in SINDy/PDE-FIND)
Q, R	Process and measurement noise covariance matrices (Kalman filtering)

lection operator (LASSO), Bayesian Relevance vector machine (RVM)/Sparse Bayesian learning (SBL)), regularization and information criteria for model selection, recursive/augmented-state filtering (Kalman Filter (KF)/Extended Kalman Filter (EKF)/Unscented Kalman Filter (UKF)) for joint state–parameter estimation, and recent structured recursive identification methods for multivariable, nonlinear, and colored-noise systems, including auxiliary-model, multi-innovation, hierarchical, filtering-based, and coupled identification. Section 3 (Multi-System Identification) extends to families of related subsystems, covering deep-learning-based models for spatio–temporal dynamics, sparse equation discovery (sparse identification of nonlinear dynamics (SINDy)

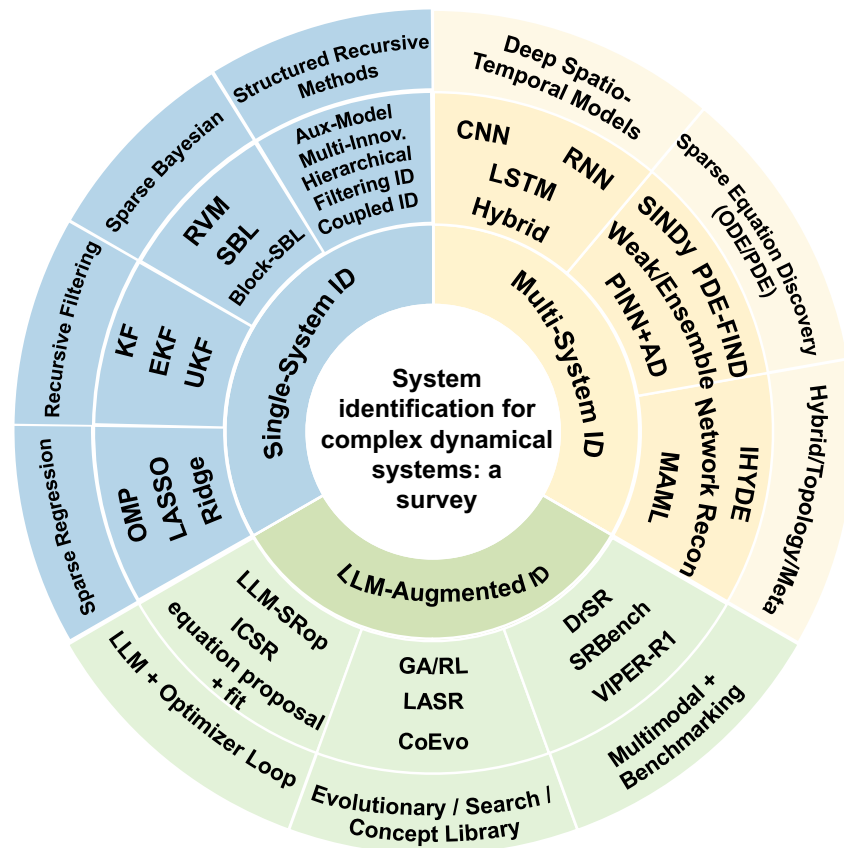


Figure 1. Overview of system identification for complex dynamical systems: single-system, multi-system, and LLM-augmented paradigms.

and weak/integral variants, PDE-FIND), hybrid/grey-box modeling (Koopman operators and piecewise dynamics such as IHYDE), networked reconstruction, and meta-learning for rapid cross-task adaptation. Section 4 (LLM for Identification) surveys emerging LLM-assisted workflows for symbolic regression and modeling pipelines, including evolutionary search, multimodal integration, and benchmarking with explicit validation protocols. Section 5 (Summary and Discussion) synthesizes the main methodological lessons and discusses limitations and open challenges—such as certifiable stability and uncertainty, robust experiment design under safety/actuation limits, distributed and federated identification, and integration with foundation models and causal evaluation. Section 6 (Conclusion) then summarizes the major findings of this review and highlights future directions for reliable, interpretable, and control-oriented system identification.

2. SINGLE-SYSTEM IDENTIFICATION

Single-system identification addresses the modeling and analysis of an individual subsystem from measured input–output data. A reliable mathematical model is foundational for analysis, optimization, and control design: it provides insight into underlying mechanisms, supports performance improvement, and enables effective controller synthesis. As a classical branch of system identification, single-system identification has matured over more than five decades and has been widely applied in industrial control and related areas. Nevertheless, advances in sensing, micro-experiments, and digital measurement have yielded high-volume, high-fidelity observations from increasingly complex—and often networked—systems. These developments pose new challenges for single-system identification and motivate methods that remain effective under modern data regimes^[11].

Formally, consider an input–output record $\mathcal{D} = \{u(t), y(t)\}_{t=1}^N$ (and possibly additional measured signals). Given a model class $\mathcal{M}(\theta)$ (e.g., input–output models, state-space models, or nonlinear parametric/nonparametric models) with parameters θ and hyperparameters $\mathcal{E}$, identification estimates θ by optimizing a chosen criterion, most commonly a prediction-error or likelihood-based objective:

$$\theta^* = \arg \min_{\theta} J(\theta; \mathcal{D}) = \frac{1}{N} \sum_{t=1}^N \ell(\hat{y}(t|t-1; \theta), y(t)) + \lambda \mathcal{R}(\theta), \quad (1)$$

where θ^* denotes the estimated parameter vector, $J(\theta; \mathcal{D})$ is the identification objective evaluated on the data record $\mathcal{D}$, $\hat{y}(t|t-1; \theta)$ denotes the one-step-ahead prediction implied by $\mathcal{M}(\theta)$, $\ell(\cdot, \cdot)$ is a data-fit loss such as squared error or negative log-likelihood, $\lambda \geq 0$ is the regularization weight, and $\mathcal{R}(\theta)$ is an optional regularization term controlling model complexity.

Operationally, single-system identification ingests observational data, is driven by a well-defined criterion, and returns an estimated model, typically within a closed-loop workflow in which models are validated, refined, and redeployed. The central difficulty lies in faithfully capturing the system’s essential dynamics under practical constraints. Typical challenges include growth in system scale and order; multivariate coupling in MIMO settings; pronounced nonlinearities and nonstationarities; measurement and process noise; and engineering constraints such as actuation limits, safety requirements, and sparsity of informative excitations. Taken together, these factors define the core technical barriers for contemporary single-system identification.

2.1 Fundamental knowledge

Before introducing classical algorithms for single-system identification, it is necessary to clarify several fundamental concepts, including the main classification schemes, essential analytical tools, and the historical evolution of the field. These elements provide the theoretical foundation and “global picture” on which the more detailed methods in the next subsection are built.

Classification of single-system identification. Classical single-system identification is most often formulated through prediction-error or likelihood-based estimation for parametric input–output and state-space models. In many widely used settings (e.g., linear regression models, ARX/ARMAX with fixed regressors, or locally approximated nonlinear models), the resulting estimators reduce to least-squares-type problems, while more general nonlinear and closed-loop cases require iterative optimization and dedicated treatments. With modern sensing and high-dimensional measurements, the number of candidate regressors (or basis functions) can be large, leading to redundant representations and ill-conditioned estimation. This motivates sparsity-promoting identification, which seeks parsimonious model structures by enforcing sparsity on parameters (or on a chosen function library), retaining only a small set of active terms that explain the data within an acceptable error level.

According to their methodological principles, the single-system identification methods reviewed in Section 2.2 can be grouped into three families that organize the methods emphasized in this review:

1. Regularized regression methods: These operate in a static regression setting and control model complexity via sparsity-promoting or shrinkage penalties. Greedy algorithms such as OMP^[12] seek sparse supports, whereas penalized estimators such as LASSO and ridge/elastic-net regression^[13,14,15] yield efficient solutions with explicit bias–variance control.
2. Sparse Bayesian methods: These introduce hierarchical probabilistic priors to induce sparsity and quantify uncertainty. Representative examples include the RVM^[16,17], classical SBL^[18,19], and block-structured extensions^[20,21,22].

3. Recursive state-space and filtering methods: These work in a dynamical state-space framework and perform online state and parameter estimation via Bayesian filtering. The KF and its nonlinear extensions (EKF, UKF) [23,24,25,26] are canonical examples, and they form the basis of augmented-state joint state-parameter identification reviewed in Section 2.2 [27,28].

These three families provide an organizing map for the single-system methods emphasized in this review; the specific algorithms in the next subsection can be seen as concrete instances or refinements within this organizing framework. Beyond these three families, recent literature has rapidly expanded a fourth line of development, namely structured recursive identification methods, which are especially important for multivariable, nonlinear, and colored-noise systems. This line includes auxiliary-model identification, multi-innovation identification, hierarchical identification, filtering-based identification, and coupled identification, and will be discussed separately at the end of this section.

Classical backbone and scope of this review. Single-system identification also has a well-established classical backbone, including PEM/ML estimation for parametric input–output and state-space models [29], subspace identification for state-space realization from input–output data [30], and closed-loop identification techniques that explicitly account for feedback-induced bias and reduced excitation [31,32]. These paradigms provide rigorous foundations for consistency, identifiability, and statistical efficiency, and remain central in industrial practice. In this review, we treat them as essential background and focus our technical exposition on sparsity-driven regression, SBL, and recursive filtering.

Analytical foundations. Three analytical tools underpin most identification algorithms: regression analysis, gradient-based optimization, and Bayesian inference.

For a linear model

$$y(k) = \varphi^T(k)\theta, \quad (2)$$

where $y(k)$ denotes the measured output at sample index k , $\varphi(k)$ is the regressor vector constructed from input, output, or state-related signals, and θ is the unknown parameter vector to be estimated. Parameter estimation is typically achieved by minimizing a loss function, such as the sum of squared residuals or the negative log-likelihood. Nonlinear cases often adopt iterative optimization schemes such as gradient descent:

$$\theta_{k+1} = \theta_k - \alpha \nabla_{\theta} J(\theta_k), \quad (3)$$

where θ_k denotes the parameter estimate at iteration k , θ_{k+1} is the updated estimate, α is the learning rate controlling the update step, $J(\theta_k)$ is the objective function evaluated at θ_k , and $\nabla_{\theta} J(\theta_k)$ denotes the gradient of the objective with respect to the parameters.

From a probabilistic perspective, sparse modeling can be expressed in Bayesian form:

$$\hat{s} = \arg \min_s \|\mathbf{y} - \mathbf{A}s\|_2^2 + \lambda \|s\|_0, \quad (4)$$

where $\hat{s}$ is the estimated sparse coefficient vector, $\mathbf{y}$ denotes the observation vector, $\mathbf{A}$ is the sensing or dictionary matrix, $\|\mathbf{y} - \mathbf{A}s\|_2^2$ measures the data-fitting error, $\|s\|_0$ counts the number of nonzero entries in s , and

$\lambda \geq 0$ controls the strength of sparsity regularization. This formulation corresponds to a maximum a posteriori (MAP) estimation problem under suitable priors. Different priors yield different regularization effects—for instance, Laplace priors correspond to LASSO-type sparsity, while Gaussian–Gamma priors lead to Bayesian RVM formulations^[33,17,34].

Research progress. The development of single-system identification is often described along three intertwined threads. The classical parametric era (1960s–1990s) established PEM/ML formulations for input–output and state-space models, together with frequency-domain methods and subspace identification, forming the backbone of industrial practice. The regularization and Bayesian era (late 1990s–2010s) emphasized robustness and uncertainty quantification, including regularized least-squares, kernel and Gaussian process viewpoints, and Bayesian identification. The sparsity-aware era (mid-2000s onward) introduced compressed-sensing-inspired methods and SBL (e.g., LASSO/OMP and RVM/SBL) to enable parsimonious structure selection in high-dimensional settings, and to connect naturally to equation discovery and modern learning pipelines.

Today, single-system identification remains a cornerstone for adaptive control, signal processing, and intelligent cyber–physical systems, providing data-driven models for prediction, diagnosis, and decision-making under noise, constraints, and limited excitation.

2.2 Classic algorithms

This subsection reviews representative classical algorithmic families for single-system identification, structured according to the three families outlined above. We first consider sparse regression methods (greedy and convex) for static regression, then sparse Bayesian formulations that provide probabilistic sparsity and uncertainty quantification, and finally recursive state-space and filtering approaches for dynamic, online estimation. These methods form the classical backbone of single-system identification. More recent structured recursive identification methods for multivariable, nonlinear, and colored-noise systems—including auxiliary-model, multi-innovation, hierarchical, filtering-based, and coupled identification schemes—will be discussed separately in the next subsection.

Orthogonal matching pursuit identification. OMP is a prototypical greedy method for sparse approximation. It was initially proposed as a recursive function approximation scheme with applications to wavelet decomposition^[12] and later became a cornerstone algorithm in sparse representation and compressed sensing. In each iteration, OMP selects the dictionary atom most correlated with the current residual and orthogonally projects the data onto the span of the selected atoms. In this way, a sparse representation is constructed through a sequence of locally optimal decisions. Although global optimality is not guaranteed, the method is computationally efficient and has proven effective in high-dimensional regimes.

The modern understanding of OMP is closely tied to the theory of compressed sensing. Sparse recovery problems are typically formulated as underdetermined linear systems with a sparsity prior, $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, where $m < n$ and infinitely many solutions satisfy $Ax = b$. Imposing that most entries of x are zero defines a well-posed recovery objective—one seeks the sparsest coefficient vector consistent with the data^[35]. Within this setting, a series of theoretical results established that greedy schemes such as OMP can recover the true sparse solution under suitable conditions on the sensing matrix. Tropp^[36] analyzed OMP under mutual-coherence assumptions, while subsequent work refined recovery guarantees in terms of the Restricted Isometry Property (RIP) and related conditions^[20,37]. These developments positioned OMP as a practical alternative to convex ℓ_1 minimization, offering similar recovery performance with lower per-iteration cost.

Beyond classical signal reconstruction, OMP has been widely adopted for variable selection in linear regression and for model structure discovery in identification tasks. For instance, in high-dimensional settings with $p \gg m$,

OMP iteratively builds a parsimonious subset of predictors by adding the most informative columns of the design matrix and re-estimating coefficients via least squares. In dynamical systems and system identification, the same mechanism is used to select a compact set of basis functions from a predefined library to explain observed dynamics. A representative example is the SINDy framework^[38], where OMP (or closely related sparse regressors) is employed to identify the active terms of governing equations from time-series data, thereby bridging sparse approximation and data-driven equation discovery.

Formally, treat the columns of $A = [\varphi_1, \dots, \varphi_n]$ as atoms. At each iteration, OMP selects the atom with maximal absolute correlation with the current residual, augments the active set, recomputes the least-squares coefficients over the active set, and updates the residual by orthogonal projection. The process terminates when the residual norm falls below a prescribed tolerance or a sparsity budget is reached. A standard version of OMP is summarized in Algorithm 1 (we use b for the observation vector, A for the dictionary, and x for the coefficient vector; in spatio-temporal identification, A corresponds to the library $\Theta(\cdot)$ and x to the sparse coefficient vector).

Algorithm 1 Orthogonal matching pursuit (OMP)

Require: dictionary $A = [\varphi_1, \dots, \varphi_n] \in \mathbb{R}^{m \times n}$, observation $b \in \mathbb{R}^m$, tolerance $\varepsilon > 0$, (optional) sparsity cap κ

Ensure: sparse coefficient vector $x \in \mathbb{R}^n$

```

1:  $r^{(0)} \leftarrow b$ ;  $S^{(0)} \leftarrow \emptyset$ ;  $x^{(0)} \leftarrow \mathbf{0}$ 
2: for  $t = 1, 2, \dots$  do
3:   Atom selection:  $j^* \leftarrow \arg \max_{j \notin S^{(t-1)}} |\langle \varphi_j, r^{(t-1)} \rangle|$ 
4:   Support update:  $S^{(t)} \leftarrow S^{(t-1)} \cup \{j^*\}$ 
5:   Least squares on active set:  $x_{S^{(t)}}^{(t)} \leftarrow \arg \min_z \|A_{S^{(t)}} z - b\|_2$ 
6:   Set  $x^{(t)}$  to zero outside  $S^{(t)}$  and to  $x_{S^{(t)}}^{(t)}$  on  $S^{(t)}$ 
7:   Residual update:  $r^{(t)} \leftarrow b - Ax^{(t)}$ 
8:   if  $\|r^{(t)}\|_2 \leq \varepsilon$  or  $|S^{(t)}| = K$  then
9:     break
10:  end if
11: end for
12: return  $x^{(t)}$ 

```

LASSO identification. LASSO, introduced by Tibshirani^[13], is a foundational method for sparse regression and has become a standard tool for system identification. Compared with ordinary least squares (OLS), which may suffer from high variance and poor interpretability in the presence of multicollinearity, LASSO introduces an l_1 penalty that simultaneously shrinks coefficients and performs variable selection. In its Lagrangian form, the objective reads

$$\min_{\beta} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1, \quad (5)$$

where Y denotes the response vector, X the design matrix, β the regression coefficients, and $\lambda \geq 0$ is a regularization parameter. Equivalently, LASSO can be posed as a constrained least-squares problem

$$(\hat{\alpha}, \hat{\beta}) = \underset{\alpha, \beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \alpha - \mathbf{x}_i^T \beta)^2 \right\} \quad \text{s.t.} \quad \sum_{j=1}^p |\beta_j| \leq t, \quad (6)$$

where α is an intercept term, $\hat{\beta}$ is the estimated coefficient vector, N is the number of samples, p is the number of predictors, y_i denotes the i -th response, $\mathbf{x}_i$ is the corresponding predictor vector, and $\lambda \geq 0$ is a tuning parameter controlling the degree of shrinkage. This formulation makes explicit the “budget” on the l_1 -norm of the coefficients.

Conceptually, LASSO may be viewed as a convex relaxation of l_0 -penalized subset selection. Whereas ridge regression^[14] employs an l_2 penalty to stabilize estimates without enforcing exact sparsity, the l_1 penalty of LASSO yields sparse solutions while remaining computationally tractable. Subsequent work developed efficient algorithms and regularization paths, most notably least-angle regression (LARS)^[39] and coordinate-descent implementations such as `glmnet`^[40], which enable large-scale applications and extensions to generalized linear models.

On the theoretical side, a rich literature has established the statistical properties of LASSO in high-dimensional settings. Elastic Net^[15] combines l_1 and l_2 penalties to handle strongly correlated predictors, while the adaptive LASSO^[41] achieves oracle properties under suitable conditions. High-dimensional analyses, such as Bickel *et al.*^[42], provide recovery guarantees and error bounds, clarifying how sparsity, sample size, and noise level jointly determine estimation accuracy.

These developments have made LASSO-based sparse regression a natural building block for system identification. In many physical systems, only a few terms dominate the underlying dynamics, so that the governing equations are sparse with respect to a high-dimensional function library. This principle is operationalized in SINDy^[38], where the response y corresponds to estimated temporal derivatives ($\dot{\mathbf{X}}$), the design matrix $\mathbf{x}$ is instantiated as a dictionary of candidate functions ($\Theta(\mathbf{X})$), and the coefficient vector β encodes a sparse set of active terms defining the governing equations. The resulting optimization problem takes the form

$$\hat{\Xi} = \underset{\Xi}{\operatorname{argmin}} \left\{ \|\dot{\mathbf{X}} - \Theta(\mathbf{X})\Xi\|_2^2 + \lambda \|\Xi\|_1 \right\}, \quad (7)$$

where $\dot{\mathbf{X}}$ denotes the matrix of estimated temporal derivatives, $\Theta(\mathbf{X})$ is the candidate-function library constructed from the state data $\mathbf{X}$, Ξ is the unknown coefficient matrix, $\hat{\Xi}$ is the identified sparse coefficient matrix, $\|\dot{\mathbf{X}} - \Theta(\mathbf{X})\Xi\|_2^2$ measures the regression error, $\|\Xi\|_1$ promotes sparsity, and $\lambda \geq 0$ controls the sparsity level. In the ODE setting, this formulation recovers the original SINDy model. For systems governed by partial differential equations, the same sparse-regression idea underlies PDE-FIND and its extensions^[43], which augment the candidate library with spatial derivatives and employ tailored regularization strategies. These PDE-oriented variants, together with robustness enhancements such as weak and ensemble formulations, are discussed in more detail in the PDE identification discussion. In system identification practice, LASSO thus plays two complementary roles. First, it offers an interpretable mechanism for model structure selection and complexity control, bridging classical regression and modern compressed sensing. Second, through frameworks such as SINDy, PDE-FIND, and their variants, LASSO-based sparse regression enables the discovery of parsimonious, physically meaningful dynamical models from noisy measurements, which can subsequently support prediction, control design, and model-based diagnostics.

Ridge regression identification. Ridge regression arises from the broader idea of Tikhonov regularization for ill-posed inverse problems, where a quadratic penalty is added to stabilize solutions of noisy or ill-conditioned linear systems^[44]. In the statistical literature, Hoerl and Kennard^[14] introduced ridge regression as a practical remedy for multicollinearity in multiple regression, showing that shrinking coefficients towards zero can substantially reduce estimation variance and improve prediction accuracy when predictors are strongly correlated.

Formally, ridge regression augments the least-squares objective with an L_2 -norm penalty on the coefficients,

$$\min_{\beta} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2, \quad (8)$$

where Y denotes the response vector, X the design matrix, β the regression coefficients, and $\lambda \geq 0$ the regularization parameter controlling the bias–variance trade-off. A larger λ enforces stronger shrinkage and improves numerical stability when $X^T X$ is ill-conditioned or nearly singular, whereas a smaller λ recovers the ordinary least-squares solution at the expense of increased variance. Because the problem remains strictly convex, ridge admits a unique closed-form solution and can be efficiently implemented via standard linear algebra routines.

In system identification, ridge-type regularization provides a natural baseline for regularized parameter estimation. Early work adopted it mainly as a pragmatic tool to handle noisy, collinear regressors and limited data records. More recently, ridge has been interpreted through the lens of reproducing-kernel Hilbert spaces and Gaussian priors, where the L_2 penalty corresponds to enforcing smoothness or energy constraints on impulse responses or nonlinear kernels^[45,46]. Within this framework, regularized least-squares estimators become central components of kernel-based and Gaussian-process identification schemes for both linear and nonlinear dynamic systems, with provable robustness and uncertainty quantification.

An important extension of ridge regression is the Elastic Net (EN), proposed by Zou and Hastie^[15]. The Elastic Net combines the sparsity-inducing L_1 penalty of LASSO with the grouping and stabilization properties of the L_2 penalty:

$$\min_{\beta} \|Y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2. \quad (9)$$

where $\lambda_1 \geq 0$ controls the L_1 sparsity-inducing penalty, $\lambda_2 \geq 0$ controls the L_2 shrinkage penalty, and the other symbols follow the ridge and LASSO formulations above. This hybrid formulation addresses limitations of pure L_1 or L_2 regularization in high-dimensional settings, encouraging groups of correlated predictors to enter or leave the model together while retaining exact sparsity. For system identification problems with many candidate regressors (e.g., large function libraries, long memory structures, or multiple input channels), Elastic Net–type penalties offer a flexible compromise between variable selection and coefficient shrinkage.

Taken together, ridge regression and its Elastic Net extension form a key branch of regularization-based identification methods. They provide numerically stable and conceptually simple building blocks and integrate seamlessly with more advanced frameworks, such as kernel-based regularization and Bayesian system identification. In practice, the choice between LASSO, ridge, and Elastic Net is driven by data characteristics (e.g., degree of collinearity, signal-to-noise ratio, and dimensionality) and modeling goals (sparsity, grouping, or smoothness), and these methods collectively underpin many modern sparse and regularized system identification pipelines.

Relevance vector machine identification. SBL provides a probabilistic framework for sparse modeling and variable selection, and offers a principled alternative to convex regularization methods such as LASSO and ridge regression. A central milestone in this line of work is the RVM proposed by Tipping^[16], which recasts sparse regression and classification within a Bayesian kernel–machine formulation. In contrast to support vector machines (SVMs), which rely on margin maximization and typically yield dense sets of support vectors, the RVM enforces sparsity through hierarchical priors while retaining fully probabilistic predictions.

In the standard RVM formulation, the system output is modeled as

$$\mathbf{t} = \Phi \mathbf{w} + \mathbf{z}, \quad (10)$$

where Φ is a design matrix composed of basis or kernel functions, $\mathbf{w}$ is the weight vector, and $\mathbf{z}$ denotes Gaussian noise. SBL imposes independent Gaussian priors with individual precision parameters on each weight, which are themselves governed by hyperpriors. Maximization of the marginal likelihood then drives many precisions to infinity, effectively pruning the corresponding weights and leaving only a small subset of “relevance vectors” active in the final model^[16]. Subsequent work developed efficient training algorithms for this framework, notably fast marginal-likelihood maximization schemes for large-scale problems^[17] and a unifying analysis of SBL/automatic relevance determination (ARD) that clarified its relationship to ℓ_1 -type regularization and basis selection^[18].

From a system identification perspective, the RVM/SBL family occupies an intermediate position between purely deterministic regularization and fully Bayesian state-space modeling. On the one hand, it inherits many advantages of convex sparse regression—automatic model-order selection, mitigation of multicollinearity, and robustness in high-dimensional settings. On the other hand, its Bayesian formulation provides uncertainty quantification on both weights and predictions, which is crucial for risk-aware control and experiment design. Empirically, RVM models tend to be substantially sparser than their SVM counterparts for comparable predictive accuracy, reducing evaluation cost and facilitating online adaptation.

These properties have motivated a growing body of applications in identification and diagnostics of dynamical systems. In fault diagnosis and condition monitoring, RVM-based classifiers have been employed for bearing fault detection and fault classification in rotating machinery, often outperforming SVMs under small-sample and noisy conditions^[47]. In environmental and hydrological modeling, multivariate RVMs have been used for spatio-temporal prediction of root-zone soil moisture, demonstrating the ability of SBL to capture complex spatial dependencies with compact models^[48]. More recently, RVM variants have been embedded in equation-discovery and multi-task frameworks, where Bayesian sparsity is exploited to identify governing terms shared across multiple operating regimes while accounting for task-specific noise^[49]. These developments highlight the suitability of RVM/SBL as a bridge between interpretable sparse models and probabilistic system identification.

In this review, we therefore regard RVM-based sparse Bayesian identification as a representative of the broader SBL paradigm. It complements greedy and convex-regularization approaches by providing (i) automatic relevance determination without manual tuning of sparsity levels, (ii) probabilistic predictions with quantified uncertainty, and (iii) typically very compact model representations. These characteristics make RVM/SBL particularly attractive for single-system identification in data-limited regimes and for multi-system scenarios where uncertainty and model parsimony are central concerns. Figure 2 illustrates the conceptual structure of RVM-based sparse identification.

Classic sparse Bayesian identification. SBL extends sparse modeling into a fully probabilistic framework by introducing hierarchical priors over model parameters. Building on the automatic relevance determination (ARD) idea underlying the RVM^[16], SBL generalizes Bayesian sparse regression to broad classes of linear inverse problems and basis-selection tasks^[18,19]. In contrast to explicit L_1 -penalized formulations such as LASSO, sparsity here emerges from inference over latent hyperparameters that control the variance of individual coefficients.

A generic observation model can be written as

$$\mathbf{T}_{ij} = \Phi(\mathbf{t}) \mathbf{W}_{ij} + \boldsymbol{\varepsilon}, \quad (11)$$

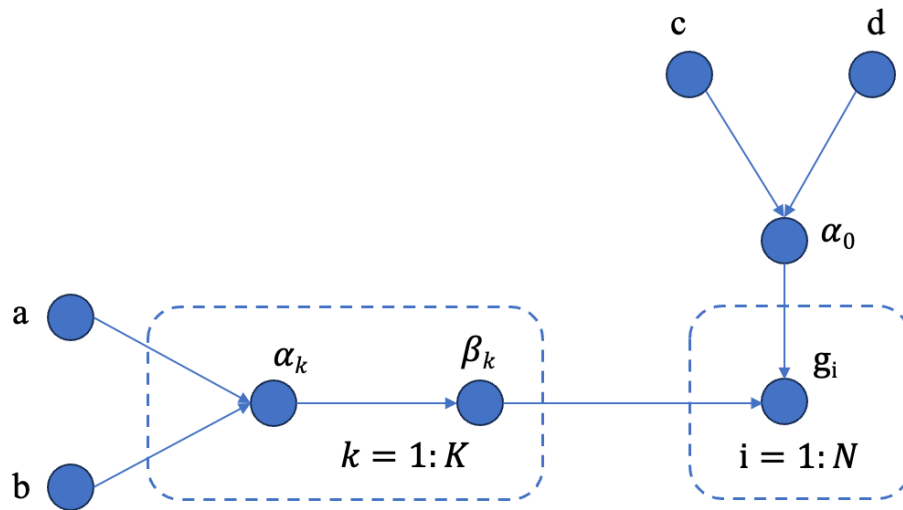


Figure 2. Conceptual structure of relevance vector machine (RVM) identification

Table 4. Equivalent penalty structures for common sparse Bayesian models

Model	Effective Penalty Term
Normal–Jeffreys	$N \log(\mu_k)$
Normal–Gamma	$\left(\frac{N}{2} - \frac{\alpha}{K}\right) N \log(\mu_k) - \log K^{\frac{N}{2}-1}(\gamma \mu_k)$
Normal–Inverse Gaussian	$(N - \frac{1}{2}) \log(q_k) - \log N(\gamma q_k)$

where $\Phi(i)$ denotes a basis (dictionary) matrix, $\mathbf{W}_{ij}$ is the unknown coefficient vector, and ε is Gaussian noise. SBL assumes Gaussian priors over each coefficient with individual precision (inverse-variance) hyperparameters, which themselves follow suitably chosen hyperpriors (e.g., Normal–Jeffreys, Normal–Gamma, or Normal–Inverse Gaussian). Under Gaussian likelihood and prior assumptions, the posterior over $\mathbf{W}_{ij}$ remains Gaussian with analytically tractable mean and covariance updates. Different hyperprior choices are equivalent to different implicit penalty structures on $\mathbf{W}_{ij}$, as summarized in Table 4. In this way, sparse Bayesian identification unifies classical sparse regression and probabilistic inference, achieving both automatic model selection and uncertainty quantification—two properties that are particularly valuable in noisy or data-limited identification settings.

In Table 4, μ_k denotes a coefficient- or basis-specific variance-related hyperparameter, q_k denotes the corresponding scale variable in the Normal–Inverse Gaussian formulation, N is the number of observations, K is the number of candidate basis terms or coefficient groups, and α and γ are hyperparameters controlling the shape and scale of the corresponding sparse Bayesian priors.

Early SBL work focused on establishing these hierarchical formulations and clarifying their relationship to convex and nonconvex sparsity penalties [18,19]. Subsequent studies exploited SBL’s robustness and automatic relevance determination to tackle increasingly challenging inverse problems, including compressive sensing, deconvolution, and structured sparsity. This progression naturally connects SBL to system identification, where ill-conditioned regressors, correlated excitations, and limited data are pervasive.

On the application side, SBL has been deployed across a wide range of domains relevant to identification and modeling. In block-sparse recovery, Sant *et al.* [22] proposed a total variation (TV)–regularized SBL framework to handle unknown and nonuniform block structures, combining linear TV and region-aware logarithmic TV penalties to enhance sensitivity to block patterns and improve robustness under mixed sparsity. In high-mobility wireless communications, Wei *et al.* [50] developed an off-grid SBL-based channel estimation scheme for orthogonal time–frequency space (OTFS) systems, mitigating grid mismatch via first-order linear approximations and achieving high-precision estimation under low-resolution sampling.

SBL has also emerged as a powerful engine for large-scale data-driven modeling. In radio environment mapping, Wang *et al.* [51] combined SBL with environment-aware dictionaries and information-theoretic sensor placement, enabling efficient three-dimensional spectrum reconstruction. For biomedical signal decoding, Wang *et al.* [52] proposed the SBLEST framework, which integrates spatial–temporal filtering and classification within a single SBL objective, automatically learning neurophysiologically interpretable EEG features and outperforming deep networks in small-sample regimes. In structural health monitoring, Wang *et al.* [53] introduced an improved SBL scheme (iSBL) with Bayesian regularization to suppress overfitting and enhance generalization, delivering accurate modeling of nonstationary vibration signals for infrastructure safety assessment.

Despite these advances, classic SBL still faces challenges in high-dimensional and real-time identification scenarios. Inference can be computationally demanding, and hyperparameter learning often relies on hand-tuned update rules or empirical heuristics [22,50]. Current research therefore explores faster variational or approximate-inference schemes, more expressive yet tractable hyperprior families, and self-adaptive hyperparameter learning, as well as extensions to multimodal fusion and online intelligent systems. Overall, SBL provides a flexible probabilistic backbone for sparse identification, sitting at the interface between traditional regularization, Bayesian learning, and modern large-scale inverse problems.

Block Bayesian identification. Within the SBL family, block sparse Bayesian learning (BSBL) generalizes element-wise sparsity to group-structured signals. Instead of treating coefficients as independent, BSBL assumes that coefficients within the same block are jointly relevant while sparsity is enforced across blocks. This idea was systematized by Zhang and Rao, who showed that many practical signals (e.g., multichannel or temporally correlated sources) are naturally block-sparse and that explicitly modeling intra-block correlation can substantially improve recovery performance [20]. Their subsequent extension of SBL algorithms for block-sparse signals with intra-block correlation [21] formalized the BSBL framework: each block is associated with its own covariance structure and hyperparameters, so that both the block-level sparsity pattern and the within-block correlation are inferred from data.

Following these foundational works, a series of studies focused on algorithmic efficiency and on relaxing assumptions about the block structure. Liu *et al.* proposed a fast marginalized BSBL algorithm that marginalizes out the coefficients and operates directly on hyperparameters, yielding significant computational savings for large-scale problems while preserving the advantages of BSBL in correlated settings [54]. Fang *et al.* [55] further introduced pattern-coupled SBL for recovery of block-sparse signals with unknown block patterns. By coupling the hyperparameters of neighboring coefficients, their hierarchical prior encourages clustered sparsity without requiring explicit knowledge of block boundaries, thus broadening the applicability of block-type models to scenarios where only loose prior information on structure is available.

More recent developments integrate richer structural priors into the BSBL framework. For example, total-variation–regularized SBL formulations incorporate linear and region-aware TV penalties to better capture piecewise smooth or spatially contiguous blocks, improving robustness to mixed sparsity patterns and nonuniform block sizes [22]. Collectively, these advances indicate a clear research trajectory: from early BSBL models

assuming known block partitions and simple intra-block correlations, toward flexible, data-driven formulations that can learn complex, overlapping, or partially specified group structures while retaining the probabilistic interpretability of SBL.

For system identification, block Bayesian approaches are particularly attractive when the underlying dynamics exhibit modular or subsystem-level sparsity. Basis functions corresponding to the same physical mechanism, spatial neighborhood, or input–output channel can be grouped into blocks, so that BSBL simultaneously identifies which modules are active and how their parameters co-vary. The resulting models provide not only sparse representations but also block-level uncertainty quantification, which is valuable for robust control design, fault diagnosis, and adaptive decision-making under noisy or incomplete observations. Overall, block Bayesian identification represents an important evolution of sparse Bayesian methods toward structured, interpretable, and scalable modeling of high-dimensional dynamical systems.

KF and its extensions. Beyond static sparse regression, Kalman-type filters provide a complementary, model-based route to dynamic system identification and belong to the third family in our organizing framework. Since Kalman's seminal work on linear filtering and prediction^[23], the KF has become the standard recursive estimator for linear Gaussian state–space models, with numerous expositions and engineering-oriented tutorials consolidating its theory and implementation^[24]. For system identification, this line of work is particularly attractive because it naturally accommodates process and measurement noise, supports online updating, and yields uncertainty estimates together with state and parameter trajectories.

In the classical linear setting, the KF addresses the discrete-time state–space model

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1}, \quad z_k = Hx_k + v_k,$$

where x_k denotes the latent system state at time step k , u_{k-1} is the control input, z_k is the measurement, A is the state-transition matrix, B is the input matrix, H is the measurement matrix, and w_{k-1} and v_k denote zero-mean process and measurement noises with covariance matrices Q and R , respectively. The algorithm alternates between prediction and correction steps to update the posterior mean and covariance of x_k , thereby minimizing the mean-square estimation error. Within system identification, this framework is routinely used in an *augmented-state* formulation, where unknown (possibly time-varying) parameters are appended to the physical state, $x'_k = [x_k^\top, \theta_k^\top]^\top$, and modeled as a slow random walk. This strategy, systematized in Ljung's monograph on system identification^[29], turns parameter estimation into a recursive filtering problem and lays the groundwork for nonlinear extensions.

To handle nonlinear dynamics and/or measurement relations, the EKF linearizes the system around the current estimate via first-order Taylor expansion. Early analyses by Ljung showed that, under mild regularity and excitation conditions, the EKF can exhibit favorable asymptotic behavior when used as a parameter estimator for (mildly) nonlinear systems^[25]. On the application side, EKF-based identification has been widely adopted in structural health monitoring and vibration analysis, where slowly varying stiffness or damping parameters must be tracked in real time from noisy sensor data^[27]. These studies collectively highlight both the appeal and the limitations of EKF: it is computationally efficient and easy to integrate with existing state–space models, but its performance can degrade when linearization errors and Jacobian sensitivities become pronounced.

Motivated by these shortcomings, Julier and Uhlmann introduced the UKF, which replaces local linearization with the Unscented Transform to propagate a deterministic set of sigma points through the nonlinear dynamics^[26]. The UKF captures posterior means and covariances up to second order without requiring analytical

Jacobians, offering improved accuracy and numerical robustness for strongly nonlinear or high-dimensional systems. In the identification literature, UKF-type schemes have therefore become a standard choice for joint state–parameter estimation in control, navigation, and robotics, especially when the underlying process models are only moderately well-known and must be refined from data.

More recently, research has begun to fuse Kalman filtering with deep learning in order to handle complex, non-stationary environments. A representative example is the “deep–Kalman” architecture, where recurrent neural networks—often long short-term memory (LSTM) models—are used to learn nonlinear process dynamics or noise structures, while the Kalman update equations enforce probabilistic consistency and provide uncertainty estimates^[28]. Such hybrid schemes illustrate a broader trend: combining physically motivated state–space models with data-driven components, so that the filter can inherit interpretability and stability from classical control theory while leveraging neural networks to model residual dynamics more expressively. This line of work points toward next-generation identification methods that unify model-based and learning-based paradigms, and will be revisited in the multi-system and learning-augmented sections of this review.

At the same time, another important recent development within single-system identification is the rapid emergence of structured recursive parameter-estimation methods for multivariable, nonlinear, and colored-noise systems. In contrast to the classical sparse-regression and Kalman-filtering paradigms reviewed above, these methods emphasize decomposition, auxiliary modeling, innovation expansion, filtering transformations, and explicit exploitation of coupling structures. Since these directions now constitute an important branch of the contemporary system identification literature, we review them separately below.

2.3 Modern structured recursive identification methods

Beyond regularization-oriented regression, SBL, and Kalman-type state-space filtering, recent single-system identification has developed a rich family of structured recursive and iterative methods. A common goal of these methods is to improve estimation accuracy, computational efficiency, and robustness in multivariable, nonlinear, and colored-noise settings. Auxiliary-model identification introduces surrogate models to handle unmeasurable intermediate variables or noise-corrupted quantities; multi-innovation identification extends the innovation from a single instant to a stacked innovation vector so as to improve data utilization; hierarchical identification decomposes a high-dimensional estimation problem into lower-dimensional subproblems; filtering-based identification transforms the data or model to mitigate the bias induced by colored noise; and coupled identification explicitly exploits shared or partially-coupled information structures in multivariable systems. Recent work increasingly combines these ideas rather than treating them as isolated techniques.

Auxiliary-model identification. Auxiliary-model identification is particularly useful when the identification model contains unmeasurable intermediate variables, internal signals, or nuisance quantities that cannot be handled directly by standard least-squares-type estimators. The basic idea is to introduce an auxiliary model and use its generated signals to facilitate parameter estimation in the original nonlinear or stochastic system. Recent studies show that this idea has become an important component of modern recursive identification. For example, Wang *et al.*^[5] proposed an auxiliary model-based maximum-likelihood recursive least-squares framework for MIMO systems and further extended it to a multi-innovation version with improved estimation accuracy. For Wiener nonlinear output-error systems, Tian *et al.*^[56] developed an auxiliary-model gradient-based iterative identification method with a moving data window, illustrating how auxiliary models can be combined with iterative optimization in nonlinear settings. For Hammerstein nonlinear systems with non-uniform sampling, Qin and Ji^[10] proposed two-stage gradient-based iterative algorithms based on key-term separation, further demonstrating the continuing development of decomposition- and auxiliary-model-oriented ideas in nonlinear identification. Overall, auxiliary-model identification provides an effective way to transform difficult estimation

problems into more tractable recursive or iterative forms, especially when internal variables are unavailable, or the original estimation equations are difficult to solve directly.

Multi-innovation identification. The central idea of multi-innovation identification is to replace the conventional one-step innovation with a stacked innovation vector over multiple sampling instants, thereby improving information reuse and parameter-estimation accuracy. This strategy is particularly effective when the data are noisy, the model is multivariable, or the recursive estimator suffers from slow convergence under single-innovation updates. In recent literature, multi-innovation methods are frequently integrated with other structured approaches rather than used alone. In a neural-network modeling context, Zhou *et al.*^[9] developed a dynamic-factor and multi-innovation-based output-input feedback Elman-network modeling approach from measurements, indicating that the multi-innovation principle also extends beyond linear parametric models to recurrent nonlinear structures. For multivariable Hammerstein-input-nonlinear systems, Wang and Liu^[8] proposed hierarchical maximum-likelihood multi-innovation identification methods, showing that multi-innovation updates can be combined with hierarchical decomposition to improve both efficiency and estimation quality. In addition, Ding *et al.*^[57] proposed hierarchical stochastic-gradient and hierarchical multi-innovation stochastic-gradient identification methods for multivariable ARX models, further confirming that multi-innovation has become an important enhancement mechanism in modern recursive identification.

Hierarchical identification. Hierarchical identification aims to reduce the computational burden of high-dimensional estimation by decomposing a complex identification problem into several lower-dimensional subproblems or virtual subsystems. This principle is especially relevant for multivariable systems, where direct estimation often leads to large parameter vectors, high matrix-inversion costs, and reduced numerical efficiency. Recent studies show that hierarchical ideas are now widely embedded into recursive least-squares, stochastic-gradient, and maximum-likelihood schemes. Xing *et al.*^[6] developed highly efficient filtered hierarchical identification algorithms for MIMO systems with colored noises by combining filtering transformations, auxiliary models, and subsystem decomposition. For multi-input ARX systems with partially-coupled information vectors, Ding *et al.*^[58] proposed hierarchical recursive-gradient identification methods that explicitly exploit model structure to reduce redundant computation. Wang and Liu^[8] further showed that hierarchical decomposition can also be integrated with maximum-likelihood and multi-innovation ideas in Hammerstein-input-nonlinear systems. Overall, hierarchical identification has evolved from a computational device for large-parameter problems into a broader structural principle for modern recursive system identification.

Filtering-based identification. Filtering-based identification is motivated primarily by the difficulty of identifying systems in the presence of colored disturbances. When the noise is correlated rather than white, standard recursive estimators may become biased or inefficient. The filtering identification idea addresses this issue by transforming the original model or data into a filtered identification form in which the disturbance structure is easier to handle. Xu *et al.*^[59] provided a recent formulation of this idea by converting systems with colored noise into identification models with white-noise terms and then deriving a corresponding parameter-estimation algorithm. Building on this line, Xing *et al.*^[6] proposed filtered auxiliary-model recursive least-squares and four-stage filtered hierarchical recursive least-squares algorithms for MIMO systems with colored noises. Ding *et al.*^[60] further extended the filtering idea to multivariable equation-error ARMA-like systems and developed filtered hierarchical generalized-extended stochastic-gradient, recursive-gradient, and least-squares methods, together with their multi-innovation counterparts. These studies show that filtering-based identification has become a key tool for combining colored-noise robustness with recursive efficiency, especially in multivariable and noise-correlated settings.

Coupled identification. Coupled identification exploits the fact that, in many multivariable systems, different subsystem identification equations are not fully independent but share common or partially shared informa-

tion vectors. Instead of estimating each subsystem separately, coupled identification explicitly uses this shared structure to reduce redundancy and improve computational efficiency. Liu et al. [7] proposed a distributed identification method based on a partially-coupled recursive generalized extended least-squares algorithm for multivariate input-output-error systems with colored noises, showing that partially shared information structures can be embedded directly into recursive estimation. Ding et al. [58] further developed hierarchical recursive-gradient identification methods for multi-input ARX systems with partially-coupled information vectors. More recently, Ding et al. [57] derived coupled identification models for multivariable ARX systems and combined them with hierarchical stochastic-gradient and hierarchical multi-innovation stochastic-gradient algorithms. Overall, coupled identification should be viewed as a structural extension of recursive system identification: it is not merely concerned with “multiple systems,” but with how shared information can be explicitly encoded within the estimator for a single multivariable plant.

Taken together, auxiliary-model, multi-innovation, hierarchical, filtering-based, and coupled identification methods represent an important modern branch of single-system identification. They are especially relevant for multivariable, nonlinear, and colored-noise systems, where classical one-shot least-squares formulations are often insufficient. More broadly, these methods illustrate a notable recent trend: modern identification algorithms increasingly arise from structured combinations of modeling, filtering, decomposition, and recursive optimization ideas rather than from a single estimation principle alone.

3. MULTI-SYSTEM IDENTIFICATION

With the rapid evolution of modern industry and the increasing complexity of real-world systems, identification problems have shifted from isolated single units to families of interacting subsystems. Large-scale processes in power grids, transportation networks, and cyber–physical infrastructures are typically high-dimensional, nonlinear, and stochastic, with strong spatial and temporal coupling. As a result, the gap between traditional single-system models and actual system behaviors has widened. In the context of large-scale digitalization and Industry 4.0, massive amounts of heterogeneous data—often collected over networks of sensors and controllers—have become available, stimulating a paradigm shift toward data-driven, multi-system modeling. Modern deep neural networks (DNNs), with their strong representational capacity, and sparse physics-informed models now form two complementary pillars of contemporary multi-system identification at the intersection of system theory and artificial intelligence.

3.1 Fundamental knowledge

Multi-system identification typically involves spatio–temporal data that characterize coupled dynamic processes across multiple subsystems. Before introducing representative algorithms, it is useful to outline a generic formulation and a taxonomy that jointly cover the methods in Section 3.2.

Preliminary formulation. A common abstraction is to view each spatial location (i, j) as a local subsystem with an associated time series $T_{ij} \in \mathbb{R}^M$. A shared dictionary of temporal basis functions $\phi(t) \in \mathbb{R}^{M \times P}$ is constructed, and local dynamics are encoded by coefficient vectors $\mathbf{W}_{ij} \in \mathbb{R}^P$:

$$T_{ij} = \phi(t)\mathbf{W}_{ij}. \quad (12)$$

where T_{ij} denotes the time series associated with spatial location or subsystem (i, j) , $\phi(t)$ is the shared temporal basis or dictionary evaluated at time t , and $\mathbf{W}_{ij}$ is the local coefficient vector encoding the dynamics of subsystem (i, j) . Sparse identification then seeks the most compact representation of $\mathbf{W}_{ij}$ that accurately reconstructs T_{ij} , thereby capturing the dominant structural dependencies of the multi-system.

To account for model mismatch and measurement noise, the identification problem is often relaxed with an error tolerance:

$$\min_{\mathbf{W}_{ij}} \|\mathbf{W}_{ij}\|_0 \quad \text{s.t.} \quad \|T_{ij} - \phi(t)\mathbf{W}_{ij}\|_p^p < \delta, \quad (13)$$

where $\|\mathbf{W}_{ij}\|_0$ counts the number of nonzero coefficients, $\|T_{ij} - \phi(t)\mathbf{W}_{ij}\|_p^p$ measures the reconstruction error under the p -norm, and $\delta > 0$ is the admissible error tolerance. or equivalently expressed as a stochastic model with Gaussian perturbations:

$$T_{ij} = \phi(t)\mathbf{W}_{ij} + \xi, \quad \xi \sim \mathcal{N}(0, \sigma^2 I). \quad (14)$$

where ξ denotes the stochastic perturbation or measurement noise, $\mathcal{N}(0, \sigma^2 I)$ is a zero-mean Gaussian distribution, σ^2 is the noise variance, and I is the identity matrix with compatible dimension. The objective is to identify the sparsest $\mathbf{W}_{ij}$ that best explains the observed spatio-temporal evolution across all (i, j) , thereby uncovering the underlying mechanisms that govern the multi-system dynamics.

Research progress and methodological taxonomy. From the perspective of this review, the development of multi-system identification can be organized into three complementary methodological lines that jointly cover all algorithms discussed in Section 3.2:

1. Deep-learning-based multi-system modeling: The entire multi-system is treated as a high-dimensional input-output map, and deep architectures automatically extract spatial, temporal, and nonlinear coupling structures. Representative models include CNNs, RNNs/LSTMs, and their hybrids^[28,61,62,63,64,65,66,67,68,69]. These methods emphasize expressive power and scalability, and are particularly suitable for complex non-linear multivariate systems.
2. Sparse equation discovery and PDE-based frameworks: Here, multi-system or multi-field data are used to identify “shared structure + local parameters” in governing laws. SINDy^[38] and its AIC/weak-form/ensemble extensions^[70,71,72] are prototypical for ODE-level dynamics, while PDE-FIND and its deep-learning-augmented variants^[43,73,74,75,76,77,78,79,80] generalize the same sparse-regression philosophy to spatio-temporal fields. This family focuses on representing multi-system behavior through compact, physically interpretable equations within a unified sparse-regression framework.
3. Hybrid and topology-aware multi-system modeling, including fast adaptation: These methods explicitly characterize network structure, piecewise dynamics, and switching logic. Examples include network reconstruction and two-step topological sampling based on sparse regression^[81,82], as well as the IHYDE framework for hybrid dynamical systems^[83]. Later parts of this review extend this line with meta-learning-based fast identification and LLM-assisted workflows, which operate at the level of system families while retaining structural interpretability and compatibility with control and diagnosis.

This taxonomy provides the “overall” organizing logic for multi-system identification: Section 3.2 revisits concrete algorithms from each of these three methodological lines and analyzes their respective strengths and limitations.

3.2 Classic algorithms

In line with the taxonomy above, this subsection reviews classic algorithms for multi-system identification by grouping them into three methodological families: (i) deep-learning-based models for high-dimensional non-linear dynamics; (ii) sparse equation discovery for ODE and PDE models; and (iii) hybrid, topology-aware, and

fast-adaptation methods including meta-learning. For each family, we highlight representative formulations and application domains rather than exhaustively cataloging variants.

Deep-learning-based multi-system identification. The use of deep neural networks (DNNs) for dynamic system identification has gradually evolved from proof-of-concept demonstrations to more structured frameworks for large-scale multi-system settings. Early work showed that deep architectures can exploit hierarchical feature representations to capture highly nonlinear mappings between inputs and outputs, outperforming shallow models such as SVMs or single-layer perceptrons in prediction accuracy and noise robustness.

Within this broader trend, Forgione and Piga^[68] proposed a unified neural modeling framework that couples network structure design with identification criteria, targeting nonlinear dynamical systems measured under realistic, noisy conditions. Their results highlight how carefully designed DNN architectures can approximate continuous-time dynamics while respecting stability and regularization requirements. To mitigate the computational burden of large models, Sahu *et al.*^[69] introduced a compression-oriented scheme in which learned nonlinear mappings are approximated by piecewise linear surrogates, significantly reducing inference time and memory footprint with only modest loss in identification accuracy. These contributions illustrate a first line of research: making deep models practically deployable for system identification through architecture and complexity control.

In parallel, domain-specific studies have demonstrated that DNN-based identification can be effectively adapted to heterogeneous multi-system environments. Alshathri *et al.*^[67] applied deep networks to cellular communication systems (GSM, UMTS, LTE), showing that data-driven models can learn complex signal patterns without explicit feature engineering and generalize across operating conditions. In power systems, Cui *et al.*^[28] proposed a multi-layer LSTM (M-LSTM) framework for time-varying parameter identification in composite load models comprising ZIP components and induction motors, validating robustness on IEEE 23-bus and 68-bus benchmarks. These works collectively demonstrate how deep architectures can be tailored to distinct engineering domains while sharing a common identification viewpoint.

More recent studies have moved toward hybrid deep learning structures that explicitly address challenges such as parameter sensitivity, chaotic behavior, and strong stochasticity. For instance, Wang *et al.*^[66] developed an evolving interval type-2 fuzzy LSTM (eIT2F-LSTM) model that synchronously identifies chaotic systems under noise and random parameter variations, illustrating how uncertainty-aware modules can be embedded into recurrent architectures. Taken together, these contributions trace a clear research trajectory: from generic DNN predictors to domain-tailored and uncertainty-aware deep models that support large-scale, nonlinear multi-system identification.

Despite these advances, several limitations remain. Deep models often exhibit limited interpretability, making it challenging to translate identified structures into physically meaningful models or to certify stability and robustness. Moreover, training and tuning large architectures can be data- and computation-intensive, particularly when labeled data are scarce or heterogeneous across subsystems. These open issues motivate the hybrid and physics-informed approaches reviewed below, which seek to combine the representational power of DNNs with the structural transparency of classical identification.

Sparse equation discovery and PDE-based multi-system identification. Building on the sparse regression ideas outlined in Section 2.2, Brunton *et al.* introduced the SINDy framework^[38], which recasts model discovery as a sparse regression problem on a library of candidate nonlinear functions. Instead of assuming a priori knowledge of the governing equations, SINDy starts from time-series measurements of the state $\mathbf{x}(t)$ and seeks a parsimonious representation of the vector field

$$\frac{d}{dt}\mathbf{x}(t) = f(\mathbf{x}(t)), \quad (15)$$

where $\mathbf{x}(t)$ denotes the system state at time t , $\frac{d}{dt}\mathbf{x}(t)$ is its time derivative, and $f(\cdot)$ is the unknown vector field to be identified. Under the assumption that only a small number of terms in a high-dimensional function library are active. Concretely, the dynamics are approximated in the regression form

$$\dot{\mathbf{X}} = \Theta(\mathbf{X})\Xi, \quad (16)$$

where $\mathbf{X}$ denotes the sampled state-data matrix, $\dot{\mathbf{X}}$ collects numerical estimates of the corresponding time derivatives, $\Theta(\mathbf{X})$ is a dictionary of candidate functions such as polynomial or trigonometric terms evaluated on $\mathbf{X}$, and Ξ is a sparse coefficient matrix whose nonzero entries encode the active terms in the governing equations. Sparse regression—typically via sequential thresholded least squares—then identifies a small set of nonzero entries in Ξ , yielding an interpretable set of governing equations.

Since its introduction, SINDy has evolved along several interconnected directions. One line of work integrates SINDy with control and optimization. Kaiser *et al.* [84] proposed SINDy-MPC, which couples SINDy-based model discovery with model predictive control, enabling data-driven control design for systems whose dynamics are initially unknown. By learning low-order, interpretable models that are compatible with MPC constraints, this framework highlights the value of sparse models in bridging identification and control.

A second line of research focuses on model selection and robustness. In practice, SINDy often produces a family of candidate models along a sparsity–accuracy Pareto front, and numerical differentiation can amplify measurement noise. To address the first issue, Mangan *et al.* [70] proposed SINDy-AIC, which uses SINDy to generate candidate models and then applies the Akaike Information Criterion for principled, automatic model selection. To mitigate noise sensitivity, Schaeffer [71] introduced integral and weak-form variants that replace pointwise derivatives with integral constraints against smooth test functions, effectively averaging out high-frequency noise. More recently, Fasel *et al.* [72] developed Ensemble-SINDy (E-SINDy), which leverages bootstrap ensembles to quantify term inclusion probabilities and improve robustness in low-data, high-noise regimes.

A third, complementary thread extends SINDy to more complex dynamical settings. PDE-FIND and related methods generalize the sparse regression framework to partial differential equations by augmenting the library with spatial derivatives and enforcing appropriate regularization [43]. These PDE-oriented variants have been combined with Galerkin projections and reduced-order modeling for high-dimensional fluid flows [85], with physics-informed constraints in plasma dynamics [86], and with tailored function libraries in nonlinear optics [87] and computational chemistry [88]. Across these applications, the core SINDy principle remains the same: exploit the expected sparsity of governing laws in a rich function library to obtain compact, interpretable models from data.

The SINDy framework was extended from ordinary differential equations (ODEs) to partial differential equations (PDEs) in the seminal PDE-FIND method of Rudy *et al.* [43]. PDE-FIND (PDE Functional Identification of Nonlinear Dynamics) adapts the core SINDy idea of sparse regression to spatiotemporal fields. The key change lies in the construction of the candidate library Θ . Whereas the original SINDy for ODEs builds a library $\Theta(\mathbf{x})$ from functions of the state variables, PDE-FIND constructs a richer library $\Theta(U)$ from the field variable $U(x, t)$ and its partial spatial derivatives. Correspondingly, the target equation shifts from the ODE form $\frac{d}{dt}\mathbf{x}(t) = f(\mathbf{x}(t))$ to the PDE form

$$u_t = N(u, u_x, u_{xx}, u_{xxx}, \dots), \quad (17)$$

where $u(x, t)$ denotes the spatio-temporal field variable, u_t is its time derivative, $u_x, u_{xx}, u_{xxx}, \dots$ denote spatial derivatives of different orders, and $N(\cdot)$ is the unknown nonlinear function that defines the governing PDE.

In practice, u_t and U are sampled at multiple space-time points. The time derivative U_t is reshaped into a single column vector, while the library Θ is assembled by evaluating candidate terms (e.g., $U, U^2, U_x, U_{xx}, UU_x$) at the same sampling locations. This yields a large, typically overdetermined, linear system

$$U_t = \Theta(U, U_x, U_{xx}, \dots) \xi, \quad (18)$$

Here, U_t is the vectorized time derivative evaluated at sampled space-time points, $\Theta(U, U_x, U_{xx}, \dots)$ is the candidate library containing field variables and spatial derivatives, and ξ is the sparse coefficient vector associated with the candidate PDE terms. Rudy et al. [43] advocate Sequential Threshold Ridge Regression (STRidge), which is empirically more robust than standard LASSO when the columns of Θ are highly correlated.

A major limitation of this baseline PDE-FIND pipeline is its sensitivity to noise. Numerical differentiation of spatial data—especially for higher-order derivatives such as u_{xx} or u_{xxxx} —is notoriously ill-conditioned and strongly amplifies measurement noise, which can easily corrupt the identified equation. This challenge has motivated a series of more robust SINDy-based PDE identification methods. A central idea is the Weak SINDy formulation [73], which replaces pointwise differentiation with an integral (weak) form: the PDE is multiplied by smooth test functions and integrated over the domain. This operation effectively averages out high-frequency perturbations, substantially improving robustness for noisy spatiotemporal data and enabling the discovery of PDEs in weak form, in line with related ensemble-weak formulations [72].

A second limitation of sparse-regression approaches like PDE-FIND is their reliance on a pre-defined, “closed” candidate library: if a true term is absent, it cannot be recovered. To relax this constraint, genetic-algorithm (GA)-based symbolic discovery has emerged as a complementary direction. These methods explore the space of equation structures via evolutionary operations such as crossover and mutation, allowing discovery even from incomplete or biased libraries. Xu et al. [74] proposed DLGA-PDE, a hybrid framework that first trains a deep neural network surrogate to generate denoised meta-data and compute stable derivatives via automatic differentiation, and then applies a GA to identify the PDE terms, achieving robustness under noisy and data-scarce regimes. To further enlarge the search space, Chen et al. [75] introduced SGA-PDE, a symbolic GA that encodes PDE terms as binary trees composed of basic operators (e.g., $+$, ∂_x) and operands (e.g., u, x), enabling gradient-free discovery of complex “open-form” equations, including fractional or composite structures beyond the reach of traditional sparse regression. More recently, Zhang et al. [76] proposed MORL4PDEs, which combines multi-objective optimization with reinforcement learning to tackle challenging PDE discovery problems involving higher-order derivatives, such as those targeted by SGA-PDE.

A third line of work addresses noise and differentiation issues by integrating deep learning, particularly physics-informed neural networks (PINNs), into the PDE identification pipeline. The central idea is to first train a neural network to approximate the spatiotemporal field, learning a smooth surrogate $\hat{u}(x, t)$ that effectively denoises the raw measurements; automatic differentiation (AD) is then applied to $\hat{u}$ to compute all required temporal and spatial derivatives with machine precision, providing a clean candidate library for subsequent sparse regression. Two seminal contributions established this direction. Both et al. [77] proposed DeepMoD, which embeds sparse regression directly into the neural network loss so that the network parameters (approximating $\hat{u}$) and the PDE

coefficients ξ are optimized jointly, yielding strong robustness to noise and small data. Similarly, Chen *et al.* [78] introduced the PINN-SR framework, which also uses a PINN and AD for derivative evaluation but couples them with an alternating direction optimization (ADO) strategy that sequentially updates the network weights and sparse PDE coefficients to handle the nontrivial joint optimization. Building on these ideas, subsequent studies have proposed refined neural architectures and training strategies for PDE discovery [79,80], further tightening the integration between data-driven surrogates, sparse regression, and physical constraints.

Hybrid, topology-aware, and meta-learning-based multi-system identification. Complementary to purely data-driven DNN approaches and sparse equation discovery, a parallel family of methods exploits system topology, piecewise dynamics, and task-relatedness when identifying families of coupled subsystems.

A first line of work focuses on topology and network structure. From a broader perspective, network-from-dynamics methods aim to reconstruct interaction graphs from observed trajectories by casting the inverse problem as a sparse regression or optimization task [81]. Within this line of research, “two-step” frameworks first perform a coarse global sparse regression to screen candidate interactions, then refine models locally using the inferred network structure. A prototypical example is Han *et al.* [82], who study networked public-goods games and seek to reconstruct both direct and indirect interactions from time-series observations. In the first, global stage, they apply LASSO-based regression to infer an effective interaction matrix from payoff and strategy trajectories, thereby exploiting sparsity to prune implausible links and obtain a coarse-grained description of the network. This step emphasizes scalability and structural discovery: it reveals which nodes appear to influence which others, but the recovered couplings still mix direct and mediated effects. In the subsequent refinement stage, Han *et al.* introduce a matrix-transformation and optimization procedure to separate direct from indirect influences and to eliminate spurious edges [82]. By enforcing additional structural constraints, they sharpen the initial coarse network into a compact, interpretable topology that more faithfully reflects the underlying game interactions. Numerical experiments on simulated networks demonstrate that the two-step pipeline can accurately recover the true adjacency structure from noisy, finite-length time series.

A second line of work focuses on hybrid and piecewise dynamics. Yuan *et al.* (2019) proposed the IHYDE framework (Identification of HYbrid Dynamical systems) [83], which aims to infer both the local dynamical laws and the associated switching logic of hybrid systems directly from data. Rather than assuming a single global model, IHYDE explicitly acknowledges that different operating regimes may follow distinct dynamics, and formulates the identification task as an optimization problem in which hybrid models are selected to minimize prediction error while maintaining parsimony. At a high level, IHYDE proceeds in two stages. First, the observed time series are partitioned into segments that are internally coherent in terms of their dynamical behavior. This segmentation can be viewed as assigning each data point (or short time window) to one of several latent subsystems, effectively uncovering a piecewise structure in the state space. Second, for each subsystem, sparse regression is applied to identify a low-complexity dynamical model from a candidate library, yielding distinct local equations for different regimes. In this way, IHYDE generalizes the sparse-identification philosophy (as in SINDy-type methods) from single-system settings to hybrid dynamical systems, jointly learning regime-dependent models and switching structure from data [83]. Yuan *et al.* demonstrate that this piecewise identification paradigm is applicable across a wide range of benchmark and engineering systems [83], including delayed relays, hysteresis loops, phototactic robots, nonlinear hybrid circuits, autonomous vehicles, industrial process monitoring, power-grid fault detection, real-time smart-grid modeling, and electrophysiological models of human atrial action potentials, as well as non-hybrid dynamical systems.

A third line of work targets fast adaptation across families of related systems via meta-learning. Traditional system identification techniques are typically designed for a single system and aim to construct an accurate model from scratch. This often requires substantial amounts of representative data and extensive parameter tuning.

In many real-world scenarios, however, engineers encounter not one isolated system but a family of related systems—for example, batches of robots with slightly different dynamics, power converters with manufacturing tolerances, or vehicles exhibiting distinct driving styles. Repeating full identification for each member of such a family is both data-intensive and computationally inefficient, hindering rapid adaptation and large-scale deployment.

Meta-learning-based identification offers a principled way to address this limitation. Rather than learning a single model, meta-learning seeks to acquire a transferable learning procedure or prior over models from a distribution of related identification tasks $p(\mathcal{T})$ [89]. By training across multiple historical tasks, the algorithm captures structural regularities shared within a system family and distills them into meta-knowledge θ . For a new task $\mathcal{T}$ drawn from $p(\mathcal{T})$, and given a small support dataset $D_{\mathcal{T}}^{\text{sup}}$, an adaptation operator $\text{Adapt}(\cdot)$ produces task-specific parameters that are evaluated on a query set $D_{\mathcal{T}}^{\text{qry}}$:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{\mathcal{T} \sim p(\mathcal{T})} [\mathcal{L}(\text{Adapt}(\theta, D_{\mathcal{T}}^{\text{sup}}), D_{\mathcal{T}}^{\text{qry}})]. \quad (19)$$

where θ denotes the meta-parameters shared across tasks, θ^* is the optimized meta-parameter vector, $\mathcal{T}$ is an identification task sampled from the task distribution $p(\mathcal{T})$, $D_{\mathcal{T}}^{\text{sup}}$ and $D_{\mathcal{T}}^{\text{qry}}$ are the support and query datasets for task $\mathcal{T}$, $\text{Adapt}(\cdot)$ denotes the task-specific adaptation operator, and $\mathcal{L}(\cdot)$ is the task loss evaluated on the query set. Once the optimal meta-parameters θ^* are obtained, the model can be rapidly adapted to a new system using only a small amount of data, enabling efficient few-shot identification [90].

A canonical optimization-based approach is Model-Agnostic Meta-Learning (MAML) [89]. MAML aims to learn an initialization θ such that a few gradient steps on task-specific data yield a performant model. For each identification task $\mathcal{T}_i$, the inner update adapts the parameters using the support set D_i^{sup} :

$$\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(\theta), \quad (20)$$

where θ is the shared initialization, θ'_i is the task-adapted parameter vector for task $\mathcal{T}_i$, α denotes the inner-loop learning rate, and $\mathcal{L}_{\mathcal{T}_i}(\theta)$ is the loss of task $\mathcal{T}_i$ evaluated on its support data. The outer update then refines the shared initialization by aggregating query-set losses across tasks:

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(\theta'_i), \quad (21)$$

where β is the meta-learning rate, the summation aggregates losses over tasks sampled from $p(\mathcal{T})$, and ∇_{θ} denotes differentiation with respect to the shared initialization θ . Because θ'_i depends on θ , the exact gradient involves second-order terms; in practice, first-order variants (FOMAML) are often used to reduce computational cost. Viewed from a system identification perspective, this procedure learns an initialization that is “close” to the optimal parameters for all systems in the family, so that only minor task-specific adaptation is needed.

Meta-learning has been combined with several identification paradigms, ranging from black-box sequence models to grey-box physical models and Koopman operator formulations. In black-box identification, neural sequence models such as RNNs, LSTMs, or GRUs directly approximate input–output dynamics,

$$\mathbf{h}_t = f_{\theta}(\mathbf{h}_{t-1}, y_{t-1}, u_t), \quad \hat{y}_t = g_{\theta}(\mathbf{h}_t), \quad (22)$$

where $\mathbf{h}_t$ is the hidden state at time t , $\mathbf{h}_{t-1}$ is the previous hidden state, y_{t-1} is the previous output, u_t is the current input, $f_\theta(\cdot)$ is the recurrent state-transition function, $g_\theta(\cdot)$ is the output mapping, and $\hat{y}_t$ is the predicted output. with task loss typically defined as the mean squared error between predictions and measurements^[90]. In grey-box and physics-informed identification, meta-learning can be wrapped around parametric physical models via hypernetworks g_θ that map limited support data to estimates of physical parameters, $\hat{\phi}_i = g_\theta(D_i^{\text{sup}})$, which are then inserted into mechanistic simulators. In Koopman operator-based identification, meta-learning helps discover shared nonlinear embeddings $\psi_\theta(\cdot)$ such that

$$\psi_\theta(x_{t+1}) \approx \mathbf{K}_t \psi_\theta(x_t), \quad (23)$$

where x_t and x_{t+1} denote consecutive system states, $\psi_\theta(\cdot)$ is the learned nonlinear lifting or embedding map parameterized by θ , and $\mathbf{K}_t$ denotes the task-specific linear dynamics matrix in the lifted space^[91]. Bayesian meta-learning further interprets θ as a hierarchical prior over model parameters^[92], providing uncertainty quantification crucial for risk-aware control, adaptive experiment design, and safe deployment in changing environments.

Taken together, hybrid, topology-aware, and meta-learning-based methods provide a third pillar for multi-system identification. They explicitly exploit structural priors and task-relatedness, complementing deep-learning-based and sparse equation discovery approaches and enabling scalable identification of large networks and families of systems.

4. LLM FOR IDENTIFICATION

In addition to classical single- and multi-system methods, recent work explores LLMs as flexible assistants for model discovery, symbolic regression, and feature engineering. This section reviews LLM-based identification workflows in four stages, emphasizing how they interact with numerical optimization and scientific constraints.

Initial symbolic regression and iterative optimization stage. A first wave of work uses LLMs as flexible generators of candidate analytical models and couples them with classical numerical fitting. In this line, LLM-SR^[93] formulates symbolic regression as a three-step loop of hypothesis generation, data-driven evaluation, and experience management: the LLM proposes candidate expressions from structured task descriptions, numerical optimizers fit free parameters and score each expression on observational data, and high-performing histories are fed back into subsequent prompts. In-Context Symbolic Regression (ICSR)^[94] follows a similar philosophy but emphasizes in-context learning: it first samples random seed functions, then iteratively asks the LLM to propose modified expressions, with coefficients estimated by nonlinear least squares and performance feedback as textual cues. Automated model discovery frameworks^[95] push this idea further by letting LLMs generate, evaluate, and refine mechanistic models directly from raw time-series (e.g., Lotka–Volterra dynamics), without pre-specifying a closed-form structure. LLM4ED^[96] explicitly positions LLMs as equation-discovery engines, combining prior scientific knowledge embedded in the model with self-optimization and evolutionary operators to produce, score, and iteratively improve candidate equations.

Evolutionary search and concept-learning enhancement stage. To improve both efficiency and scientific plausibility of equation search, a second research thread tightens the coupling between LLMs and evolutionary optimization. LASR^[97] augments LLM-based symbolic regression with a concept library: candidate expressions are iteratively evolved via mutation and crossover, while the LLM abstracts reusable “concepts” that are stored and recombined in later generations. CoEvo^[98] further casts equation discovery as an open-ended co-evolutionary process, in which LLM reasoning, population-based evolutionary strategies, and a dynamically growing knowledge base interact to explore a richer hypothesis space spanning natural-language descriptions

and executable code. Domain-specific work such as AutoTurb^[99] applies this evolutionary LLM paradigm to turbulence modeling, where algebraic closure relations are generated by the LLM and repeatedly filtered by CFD-based performance evaluation. A complementary study^[100] systematically analyzes when and why evolutionary search is necessary on top of LLM generation, highlighting that, for many scientific design problems, LLMs alone tend to under-explore the hypothesis space without explicit search operators.

Multi-reasoning mechanisms and multimodal integration stage. More recent contributions enrich the reasoning modes and input modalities available to LLM-based identification systems. DrSR^[101] introduces a dual-reasoning framework: a data-aware module extracts structural relations among variables directly from the dataset, while an inductive module distills successful and failed patterns from the generation history into an “idea library” that guides future equation proposals. VIPER-R1^[102] extends symbolic regression to a vision-symbolic setting, using Motion Structure Induction to infer equations from visual trajectories and then refining the generation policy via reinforcement learning (GRPO). SGA^[103] adopts a bilevel optimization perspective: the LLM explores discrete expression structures and initial parameter guesses at the outer level, while simulation and gradient-based search refine continuous parameters at the inner level, with a Top- κ selection mechanism maintaining a diverse pool of promising models. Collectively, these works move from single-pass prompt-response generation toward iterative, multi-signal reasoning pipelines where data, prior knowledge, and search dynamics are tightly intertwined.

Benchmarking and domain-specialized application stage. As algorithmic patterns stabilize, a fourth line of work focuses on systematic evaluation and domain adaptation for LLM-assisted identification. LLM-SRBench^[104] provides a benchmark for scientific equation discovery, emphasizing not only numerical accuracy but also physical plausibility and interpretability of the discovered laws. PhysGym^[105] probes the capabilities and limitations of LLMs in interactive physics discovery under controlled prior-knowledge settings, revealing an imbalance between memorization and genuine exploration. Beyond equation discovery per se, LLM-FE^[106] transfers the same ideas to feature engineering, where LLMs generate interpretable transformation programs whose utility is iteratively judged by downstream predictive performance. Domain-specific systems such as D3^[107] in pharmacology and AIME^[108] in ecosystem modeling demonstrate how LLMs can assist in constructing white-box or hybrid dynamical models, integrating parameter estimation and mechanistic validation loops. Taken together, these efforts suggest a trajectory in which LLMs evolve from generic code or text generators into task-aware identification assistants, embedded within principled pipelines of optimization, uncertainty assessment, and domain-specific evaluation.

Recent 2026 studies further extend this line toward automated scientific discovery, physics-informed symbolic regression, and LLM-guided dynamical-system discovery^[109,110,111,112].

5. SUMMARY AND DISCUSSION

In this paper, we revisited system identification through a unified lens that couples model structure (from classical linear parametric models to sparse libraries and hybrid/grey-box formulations), operating mode (open versus closed loop), estimation target (one-step prediction versus free-run simulation), and computational setting (offline batch versus online recursive schemes). Within this organizing perspective, we traced a trajectory from single-system identification rooted in regression, Bayesian inference, and recursive estimation, to multi-system identification for networked and spatio-temporal settings, and finally to learning-augmented paradigms based on meta-learning and LLMs. The overarching theme is that modern system identification is no longer a single algorithmic choice, but a structured design space in which sparsity, uncertainty quantification, dynamical structure, recursive estimation, and data-driven representations must be balanced against constraints of robustness, scalability, and interpretability.

For single-system identification, we organized the discussion around several complementary methodological lines. The first includes regularization-oriented paradigms such as greedy pursuit (e.g., OMP), convex regularization with sparsity-promoting or shrinkage penalties (e.g., LASSO, ridge, and elastic net), and sparse Bayesian inference (e.g., RVM, SBL, and block-structured variants), which offer different routes to model structure selection, complexity control, and uncertainty-aware estimation. The second consists of recursive state-space methods (KF/EKF/UKF) for joint state-parameter estimation, which naturally connect identification, latent-state inference, and uncertainty propagation in noisy and potentially time-varying environments. Beyond these classical lines, we further highlighted a rapidly developing branch of structured recursive identification methods for multivariable, nonlinear, and colored-noise systems, including auxiliary-model identification, multi-innovation identification, hierarchical identification, filtering-based identification, and coupled identification. These methods enrich the classical toolbox by improving information reuse, reducing computational burden, handling correlated disturbances, and exploiting shared structural information within multivariable systems. Taken together, this part of the review shows that even in the “single-system” setting, effective identification increasingly depends on structured combinations of regularization, probabilistic modeling, recursive estimation, filtering transformations, and decomposition strategies.

Moving to multi-system identification, we highlighted how the focus shifts from isolated dynamics to families of coupled subsystems, spatio-temporal fields, and hybrid or piecewise models, as reflected in deep-learning-based identification, sparse equation discovery, PDE discovery, topology reconstruction, and hybrid dynamical-system identification [38,43,68,82,83]. Deep-learning-based approaches built on DNNs, CNNs, and RNNs/LSTMs (and their hybrids) have demonstrated strong approximation power and have been empirically effective under complex nonlinearities and noisy measurements, particularly when large volumes of data and appropriate regularization are available. At the same time, model-based frameworks such as SINDy and PDE-FIND generalize sparse regression to discover ODE and PDE models from time series and field data, while weak, ensemble, and physics-informed variants alleviate noise sensitivity and incorporate prior physical structure. Hybrid methods such as IHYDE extend this logic to switching and piecewise dynamical systems by jointly inferring regime partitions and local models, while topology-aware strategies show how network structure and local dynamics can be reconstructed from spatio-temporal observations. Meta-learning-based schemes further exploit task-relatedness across system families, enabling rapid adaptation and bridging black-box, grey-box, and Koopman-operator-based formulations. Collectively, these developments show that multi-system identification has evolved into a broad spectrum of approaches ranging from high-capacity function approximation to tightly physics-informed equation discovery and hybrid structural modeling.

On top of these foundations, LLM-assisted workflows point toward a new stage in which the object of learning is not only a model, but also an adaptive modeling strategy. LLM-based symbolic regression and equation-discovery pipelines use large models to generate candidate equations or feature transformations, which are then filtered, refined, and validated through numerical optimization, evolutionary search, and domain-specific simulation [93,96,97,101]. In this sense, LLMs are best understood not as replacements for existing identification methods, but as flexible hypothesis generators and interactive modeling assistants that can complement sparse regression, mechanistic modeling, and scientific validation pipelines. Recent benchmarking and domain-focused studies suggest both the promise of these tools and the need for stricter validation, reproducibility, and mechanism-level evaluation.

Taken together, these strands point to several cross-cutting lessons. First, there is no universally best identification method; effective practice increasingly arises from the compositional use of tools, such as combining sparse regression with Bayesian uncertainty, integrating auxiliary-model and filtering ideas into recursive estimators, or embedding deep and LLM-based modules within physics-informed and control-aware loops. Second, robustness and reliability remain central challenges: numerical differentiation, model misspecification, colored

noise, distribution shift, and limited or biased data can significantly degrade performance, especially in multi-variable, nonlinear, and networked systems^[1,2,6,7,71,72]. Third, the integration of identification with experiment design, control synthesis, and formal guarantees (e.g., stability, safety, and risk-aware performance bounds) is still incomplete, particularly for large-scale, hybrid, and partially observed settings. Fourth, as the field expands toward learning-augmented and foundation-model-based workflows, the need for standardized benchmarks, leakage-aware evaluation protocols, and reproducible reporting becomes even more pressing.

Addressing these challenges will require not only new algorithms, but also closer interaction among control theory, statistical inference, machine learning, scientific computing, and emerging foundation models. Within this evolving ecosystem, system identification continues to play a pivotal role as the bridge between raw data, interpretable dynamical structure, and deployable decision-making and control policies.

6. CONCLUSION

This review has examined system identification for complex dynamical systems from a unified perspective that connects model structure, operating mode, estimation target, and computational setting with downstream requirements in prediction, diagnosis, synchronization, and control. By organizing the literature into single-system, multi-system, and learning-augmented identification, this paper shows that modern system identification is no longer limited to parameter estimation for isolated systems. Instead, it has become a broader data-to-model design problem that must jointly consider sparsity, uncertainty, interpretability, scalability, robustness, and compatibility with control objectives.

The reviewed methods reveal a clear methodological evolution. Classical and regularized identification methods provide the theoretical and computational foundation for reliable model estimation; structured recursive methods improve robustness and efficiency for multivariable, nonlinear, and colored-noise systems; multi-system identification extends the focus to networked, spatio-temporal, hybrid, and family-structured dynamics; and learning-augmented approaches, including meta-learning and LLM-assisted workflows, open new possibilities for rapid adaptation and automated scientific modeling. Looking forward, important research directions include robust identification under noisy, biased, or limited data; uncertainty-aware and interpretable learning; scalable identification for large networked systems; stronger integration with experiment design and control synthesis; and standardized benchmarks for foundation-model-assisted identification. Progress in these directions will be essential for transforming complex dynamical data into reliable, transparent, and deployable models for engineering and scientific decision-making.

DECLARATIONS

Authors' contribution

Conceptualization and manuscript drafting: Zhang, X.
Literature collection and analysis: Li, Z.; Shi, R.; Wang, S.
Manuscript review, supervision, and revision: Wang, F.; Chen, D.
All authors read and approved the final manuscript.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool ChatGPT (GPT-4o, released 2024-05-13; accessed between March and July 2026) was used solely for language editing. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

This work was supported in part by the National Key Research and Development Program of China (No. 2025YFF0524100), the National Natural Science Foundation of China (NSFC) (Grant Nos. 62233004, 62273090, and T2541017), the Jiangsu Provincial Scientific Research Center of Applied Mathematics (No. BK20233002), and the Basic Research Program of Jiangsu Province (No. BK20253020).

Conflicts of interest

Chen, D. is an associate editor of the journal *Complex Engineering Systems*. Chen, D. was not involved in any steps of editorial processing, notably including reviewers' selection, manuscript handling and decision making, while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

- Pillonetto, G.; Aravkin, A.; Gedon, D.; Ljung, L.; Ribeiro, A. H.; Schon, T. B. Deep networks for system identification: a survey. *Automatica* **2025**, *171*, 111907. [DOI](#)
- Wang, Q. G.; Zhang, L. System identification in the network era: a survey of data issues and innovative approaches. *IEEE/CAA J. Autom. Sin.* **2025**, *12*, 1305-19. [DOI](#)
- Fu, L.; Li, P. The research survey of system identification method. In: 2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics; 2013, pp. 397-401. [DOI](#)
- Åström, K. J.; Eykhoff, P. System identification-a survey. *Automatica* **1971**, *7*, 123-62. [DOI](#)
- Wang, H.; Zhang, Q.; Liu, X. Auxiliary model-based maximum likelihood multi-innovation recursive least squares identification for multiple-input multiple-output systems. *J. Frankl. Inst.* **2024**, *361*, 107352. [DOI](#)
- Xing, H.; Ding, F.; Zhang, X.; Luan, X.; Yang, E. Highly-efficient filtered hierarchical identification algorithms for multiple-input multiple-output systems with colored noises. *Syst. Control Lett.* **2024**, *186*, 105762. [DOI](#)
- Liu, Q.; Chen, F.; Guo, Q.; Wang, X. Distributed identification based partially-coupled recursive generalized extended least squares algorithm for multivariate input-output-error systems with colored noises from observation data. *J. Comput. Appl. Math.* **2024**, *449*, 115976. [DOI](#)
- Wang, H.; Liu, X. Hierarchical maximum likelihood multi-innovation identification methods for a class of multivariable Hammerstein-input-nonlinear systems. *J. Frankl. Inst.* **2025**, *362*, 108143. [DOI](#)
- Zhou, Y.; Wu, Z.; Chen, Z.; Ma, H. Dynamic factor and multi-innovation-based output-input feedback Elman network modeling from measurements. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 1-13. [DOI](#)
- Qin, H.; Ji, Y. Two-stage gradient-based iterative algorithms for the Hammerstein nonlinear systems with non-uniform sampling using the key term separation. *J. Frankl. Inst.* **2026**, *363*, 108416. [DOI](#)
- Hjalmarsson, H. System identification of complex and structured systems. *Eur. J. Control* **2009**, *15*, 275-310. [DOI](#)
- Pati, Y. C.; Rezaiifar, R.; Krishnaprasad, P. S. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In Proceedings of 27th Asilomar Conference on Signals, Systems and Computers; 1993, pp. 40-4. [DOI](#)
- Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267-88. [DOI](#)
- Hoerl, A. E.; Kennard, R. W. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **2012**, *12*, 55-67. [DOI](#)
- Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301-20. [DOI](#)
- Tipping, M. E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211-44. [DOI](#)

17. Tipping, M. E.; Faul, A. C. Fast marginal likelihood maximisation for sparse Bayesian models. In Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics; 2003. pp. 276-83. Available from: <http://proceedings.mlr.press/r4/tipping03a.html> [Last accessed on 2 Jul 2026].
18. Wipf, D.; Rao, B. D. Sparse Bayesian learning for basis selection. *IEEE Trans. Signal Process.* **2004**, *52*, 2153-64. DOI
19. Ji, S.; Xue, Y.; Carin, L. Bayesian compressive sensing. *IEEE Trans. Signal Process.* **2008**, *56*, 2346-56. DOI
20. Zhang, Z.; Rao, B. D. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 912-26. DOI
21. Zhang, Z.; Rao, B. D. Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation. *IEEE Trans. Signal Process.* **2013**, *61*, 2009-15. DOI
22. Sant, A.; Leinonen, M.; Rao, B. D. Block-sparse signal recovery via general total variation regularized sparse Bayesian learning. *IEEE Trans. Signal Process.* **2022**, *70*, 1056-71. DOI
23. Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35-45. DOI
24. Welch, G.; Bishop, G. An introduction to the kalman filter. 2001. Available from: <https://techreports.cs.unc.edu/papers/95-041.pdf> [Last accessed on 2 Jul 2026].
25. Ljung, L. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Trans. Automat. Contr.* **1979**, *24*, 36-50. DOI
26. Julier, S. J.; Uhlmann, J. K. Unscented filtering and nonlinear estimation. *Proc. IEEE* **2004**, *92*, 401-22. DOI
27. Loh, C. H.; Tou, I. C. A system identification approach to the detection of changes in both linear and non-linear structural parameters. *Earthq. Eng. Struct. Dyn.* **1995**, *24*, 85-97. DOI
28. Cui, M.; Khodayar, M.; Chen, C.; Wang, X.; Zhang, Y.; Khodayar, M. E. Deep learning-based time-varying parameter identification for system-wide load modeling. *IEEE Trans. Smart Grid* **2019**, *10*, 6102-14. DOI
29. Simpkins, A. System identification: theory for the user, 2nd Edition (Ljung, L.; 1999) [On the Shelf]. *IEEE Robot. Automat. Mag.* **2012**, *19*, 95-6. DOI
30. Van Overschee, P.; De Moor, B. Subspace identification for linear systems: theory-implementation-applications. New York, NY: Springer; 2012. DOI
31. Van Den Hof, P. M.; Schrama, R. J. Identification and control-closed-loop issues. *Automatica* **1995**, *31*, 1751-70. DOI
32. Hjalmarsson, H.; Gevers, M.; De Bruyne, F. For model-based control design, closed-loop identification gives better performance. *Automatica* **1995**, *32*, 1659-73. DOI
33. Feng, J.; Jia, K.; Qin, C.; Zhu, S.; Yang, X.; Tian, J. Sparse Bayesian reconstruction method for multispectral bioluminescence tomography. *Chin. Opt. Lett.* **2010**, *8*, 1010-14. DOI
34. Bishop, C. M. Pattern recognition and machine learning. Springer; 2006. Available from: <https://link.springer.com/book/9780387310732> [Last accessed on 2 Jul 2026].
35. Donoho, D. L. Compressed sensing. *IEEE Trans. Inform. Theory* **2006**, *52*, 1289-306. DOI
36. Tropp, J. A. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory* **2004**, *50*, 2231-42. DOI
37. Tropp, J. A.; Gilbert, A. C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory* **2007**, *53*, 4655-66. DOI
38. Brunton, S. L.; Proctor, J. L.; Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3932-7. DOI
39. Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R. Least angle regression. *Ann. Statist.* **2004**, *32*, 407-99. DOI
40. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Soft.* **2010**, *33*, 1-22. DOI
41. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2012**, *101*, 1418-29. DOI
42. Bickel, P. J.; Ritov, Y.; Tsybakov, A. B. Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **2009**, *37*, 1705-32. DOI
43. Rudy, S. H.; Brunton, S. L.; Proctor, J. L.; Kutz, J. N. Data-driven discovery of partial differential equations. *Sci. Adv.* **2017**, *3*, e1602614. DOI
44. Tikhonov, A.; Tikhonov, A. N.; Tikhonov, A. Solution of incorrectly formulated problems and the regularization method. *Sov. Math Dok.* **1963**, *4*, 1035-8. Available from: <https://www.scienceopen.com/document?vid=aec7b578-395e-4e65-8a5a-e3b24d3f9095> [Last accessed on 2 Jul 2026].
45. Pillonetto, G.; Chen, T.; Chiuso, A.; De Nicolao, G.; Ljung, L. Regularized system identification: learning dynamic models from data. Springer; 2022. DOI
46. Care, A.; Carli, R.; Libera, A. D.; Romeres, D.; Pillonetto, G. Kernel methods and gaussian processes for system identification and control: a road map on regularized kernel-based learning for control. *IEEE Control Syst. Mag.* **2023**, *43*, 69-110. DOI
47. Widodo, A.; Kim, E. Y.; Son, J.; et al. Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine. *Expert Syst. Appl.* **2009**, *36*, 7252-61. DOI
48. Zaman, B.; Mckee, M. Spatio-temporal prediction of root zone soil moisture using multivariate relevance vector machines. *Open J. Mod. Hydrol.* **2014**, *4*, 80-90. DOI
49. Bee, S.; Dervilis, N.; Worden, K.; Bull, L. Multi-task equation discovery. *arXiv* **2025**, arXiv:250925400. DOI
50. Wei, Z.; Yuan, W.; Li, S.; Yuan, J.; Ng, D. W. K. Off-grid channel estimation with sparse Bayesian learning for OTFS systems. *IEEE Trans.*

-
- Wireless Commun.* **2022**, *21*, 7407–26. DOI
51. Wang, J.; Zhu, Q.; Lin, Z.; et al. Sparse Bayesian learning-based 3-D radio environment map construction-Sampling optimization, scenario-dependent dictionary construction, and sparse recovery. *IEEE Trans. Cogn. Commun. Netw.* **2024**, *10*, 80–93. DOI
 52. Wang, W.; Qi, F.; Wipf, D. P.; et al. Sparse Bayesian learning for end-to-end EEG decoding. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 15632–49. DOI
 53. Wang, Q.; Dai, Y.; Ma, Z.; et al. Towards high-precision data modeling of SHM measurements using an improved sparse Bayesian learning scheme with strong generalization ability. *Struct. Health Monit.* **2023**, *23*, 588–604. DOI
 54. Liu, B.; Zhang, Z.; Fan, H.; Fu, Q. Fast marginalized block sparse bayesian learning algorithm. *arXiv* **2013**, arXiv:12114909 DOI
 55. Fang, J.; Shen, Y.; Li, H.; Wang, P. Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals. *IEEE Trans. Signal Proc.* **2014**, *63*, 360–72. DOI
 56. Tian, C.; Dong, Z.; Ji, Y.; Lin, X. Auxiliary model gradient-based iterative identification with moving data window for Wiener nonlinear output-error systems. *J. Frankl. Inst.* **2025**, *362*, 107992. DOI
 57. Ding, F.; Xiao, Y.; Xu, L.; Fang, Z. Hierarchical stochastic gradient and hierarchical multi-innovation stochastic gradient identification for multivariable ARX models. *Int. J. Adapt. Control Signal Process.* **2025**, *40*, 79–102. DOI
 58. Ding, F.; Luan, X.; Xu, L.; Zhang, X. Hierarchical recursive gradient parameter identification for multi-input ARX systems with partially-coupled information vectors. *Int. J. Adapt. Control Signal Process.* **2025**, *39*, 1978–95. DOI
 59. Xu, L.; Ding, F.; Zhang, X.; Zhu, Q. Novel parameter estimation method for the systems with colored noises by using the filtering identification idea. *Syst. Control Lett.* **2024**, *186*, 105774. DOI
 60. Ding, F.; Xu, L.; Zhang, X.; Xu, H.; Zhou, Y.; Luan, X. Hierarchical generalized extended parameter identification for multivariable equation-error ARMA-like systems by using the filtering identification idea. *Ann. Rev. Control* **2025**, *60*, 100993. DOI
 61. Bengio, Y. Learning deep architectures for AI. *Found. Trends Mach. Learn.* **2009**, *2*, 1–127. DOI
 62. Genc, S. Parametric system identification using deep convolutional neural networks. In 2017 International Joint Conference on Neural Networks (IJCNN); 2017. pp. 2112–9. DOI
 63. Wu, R.; Jahanshahi, M. R. Deep convolutional neural network for structural dynamic response estimation and system identification. *J. Eng. Mech.* **2019**, *145*, 04018125. DOI
 64. Wang, Y. A new concept using LSTM Neural Networks for dynamic system identification. In 2017 American control conference (ACC); 2017. pp. 5324–29. DOI
 65. Brusafferri, A.; Matteucci, M.; Portolani, P.; Spinelli, S. Nonlinear system identification using a recurrent network in a Bayesian framework. In 2019 IEEE 17th international conference on industrial informatics (INDIN); 2019. pp. 319–24. DOI
 66. Wang, H.; Luo, C.; Wang, X. Synchronization and identification of nonlinear systems by using a novel self-evolving interval type-2 fuzzy LSTM-neural network. *Eng. Appl. Artif. Intell.* **2019**, *81*, 79–93. DOI
 67. Alshathri, K.; Xia, H.; Lawrence, V.; Yao, Y. Cellular system identification using deep learning: GSM, UMTS and LTE. In 2019 28th Wireless and Optical Communications Conference (WOCC); 2019. pp. 1–4. DOI
 68. Forgione, M.; Piga, D. Model structures and fitting criteria for system identification with neural networks. In 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT); 2020. pp. 1–6. DOI
 69. Sahu, I.; Pal, A.; Ukil, A.; Majumdar, A. Compressing deep neural network: a black-box system identification approach. In 2021 International Joint Conference on Neural Networks (IJCNN); 2021. pp. 1–8. DOI
 70. Mangan, N. M.; Kutz, J. N.; Brunton, S. L.; Proctor, J. L. Model selection for dynamical systems via sparse regression and information criteria. *Proc. R. Soc. A. Math. Phys. Eng. Sci.* **2017**, *473*, 20170009. DOI
 71. Schaeffer, H.; McCalla, S. G. Sparse model selection via integral terms. *Phys. Rev. E.* **2017**, *96*, 023302. DOI
 72. Fasel, U.; Kutz, J. N.; Brunton, B. W.; Brunton, S. L. Ensemble-SINDy: robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proc. R. Soc. A.* **2022**, *478*, 20210904. DOI
 73. Messenger, D. A.; Bortz, D. M. Weak SINDy for partial differential equations. *J. Comput. Phys.* **2021**, *443*, 110525. DOI
 74. Xu, H.; Chang, H.; Zhang, D. DLGA-PDE: discovery of PDEs with incomplete candidate library via combination of deep learning and genetic algorithm. *J. Comput. Phys.* **2020**, *418*, 109584. DOI
 75. Chen, Y.; Luo, Y.; Liu, Q.; Xu, H.; Zhang, D. Symbolic genetic algorithm for discovering open-form partial differential equations (SGA-PDE). *Phys. Rev. Res.* **2022**, *4*, 023174. DOI
 76. Zhang, X.; Guan, J.; Liu, Y.; Wang, G. MORL4PDEs: data-driven discovery of PDEs based on multi-objective optimization and reinforcement learning. *Chaos Soliton Fract.* **2024**, *180*, 114536. DOI
 77. Both, G.; Choudhury, S.; Sens, P.; Kusters, R. DeepMoD: deep learning for model discovery in noisy data. *J. Comput. Phys.* **2021**, *428*, 109985. DOI
 78. Chen, Z.; Liu, Y.; Sun, H. Physics-informed learning of governing equations from scarce data. *Nat. Commun.* **2021**, *12*, 6136. DOI
 79. Stephany, R.; Earls, C. PDE-LEARN: using deep learning to discover partial differential equations from noisy, limited data. *Neural Netw.* **2024**, *174*, 106242. DOI
 80. Pervez, A.; Gavves, E.; Locatello, F. Mechanistic PDE networks for discovery of governing equations. In Forty-second International Conference on Machine Learning; 2025. Available from: <https://openreview.net/forum?id=7w1YGBPDMw> [Last accessed on 2 Jul 2026].
 81. Timme, M.; Casadiego, J. Revealing networks from dynamics: an introduction. *J. Phys. A Math. Theor.* **2014**, *47*, 343001. DOI
 82. Han, X.; Shen, Z.; Wang, W.; Lai, Y.; Grebogi, C. Reconstructing direct and indirect interactions in networked public goods game. *Sci.*

- Rep.* **2016**, *6*, 30241. DOI
83. Yuan, Y.; Tang, X.; Zhou, W.; et al. Data driven discovery of cyber physical systems. *Nat. Commun.* **2019**, *10*, 4894. DOI
 84. Kaiser, E.; Kutz, J. N.; Brunton, S. L. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proc. R. Soc. A* **2018**, *474*, 20180305. DOI
 85. Deng, N.; Noack, B. R.; Morzyński, M.; Pastur, L. R. Galerkin force model for transient and post-transient dynamics of the fluidic pinball. *J. Fluid Mech.* **2021**, *918*, A4. DOI
 86. Kaptanoglu, A. A.; Morgan, K. D.; Hansen, C. J.; Brunton, S. L. Physics-constrained, low-dimensional models for magnetohydrodynamics: first-principles and data-driven approaches. *Phys. Rev. E* **2021**, *104*, 015206. DOI
 87. Sorokina, M.; Sygletos, S.; Turitsyn, S. Sparse identification for nonlinear optical communication systems: SINO method. *Opt. Express* **2016**, *24*, 30433-43. DOI
 88. Boninsegni, L.; Nüske, F.; Clementi, C. Sparse learning of stochastic dynamical equations. *J. Chem. Phys.* **2018**, *148*, 241723. DOI
 89. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the 34th International Conference on Machine Learning; 2017, pp. 1126-35. Available from: <https://proceedings.mlr.press/v70/finn17a.html> [Last accessed on 2 Jul 2026].
 90. Lin, S.; Wang, H.; Zhang, J. System identification via meta-learning in linear time-varying environments. *arXiv* **2020**, arXiv:201014664 DOI
 91. Iwata, T.; Kawahara, Y. Meta-learning for koopman spectral analysis with short time-series. *arXiv* **2021**, arXiv:210204683. DOI
 92. Grant, E.; Finn, C.; Levine, S.; Darrell, T.; Griffiths, T. Recasting gradient-based meta-learning as hierarchical bayes. *arXiv* **2018**, arXiv:180108930. DOI
 93. Shojaei, P.; Meidani, K.; Gupta, S.; Farimani, A. B.; Reddy, C. K.; LLM-SR: scientific equation discovery via programming with large language models. In The Thirteenth International Conference on Learning Representations; 2025. Available from: <https://openreview.net/forum?id=m2nmp8P5in> [Last accessed on 2 Jul 2026].
 94. Merler, M.; Haitsiukevich, K.; Dainese, N.; Marttinen, P. In-context symbolic regression: Leveraging large language models for function discovery. *arXiv* **2024**, arXiv:240419094 DOI
 95. Li, M. Y.; Fox, E. B.; Goodman, N. D. Automated statistical model discovery with language models. *arXiv* **2024**, arXiv:240217879. DOI
 96. Du, M.; Chen, Y.; Wang, Z.; Nie, L.; Zhang, D. Llm4ed: large language models for automatic equation discovery. *arXiv* **2024**, arXiv:240507761. DOI
 97. Grayeli, A.; Sehgal, A.; Costilla-Reyes, O.; Cranmer, M.; Chaudhuri, S. Symbolic regression with a learned concept library. *Adv. Neural Inf. Proc. Syst.* **2024**, *37*, 44678-709. DOI
 98. Guo, P.; Zhang, Q.; Lin, X. Coevo: continual evolution of symbolic solutions using large language models. *arXiv* **2024**, arXiv:241218890. DOI
 99. Zhang, Y.; Zheng, K.; Liu, F.; Zhang, Q.; Wang, Z. AutoTurb: using large language models for automatic algebraic turbulence model discovery. *Phys. Fluids* **2025**, *37*, 015211. DOI
 100. Zhang, R.; Liu, F.; Lin, X.; Wang, Z.; Lu, Z.; Zhang, Q. Understanding the importance of evolutionary search in automated heuristic design with large language models. In International Conference on Parallel Problem Solving from Nature; 2024, pp. 185-202. DOI
 101. Wang, R.; Wang, B.; Li, K.; Zhang, Y.; Cheng, J. DrSR: LLM based scientific equation discovery with dual reasoning from data and experience. *arXiv* **2025**, arXiv:250604282. DOI
 102. Liu, J.; Lai, S.; Li, P.; et al. Mimicking the physicist's eye: a VLM-centric approach for physics formula discovery. *arXiv* **2025**, arXiv:250817380. DOI
 103. Ma, P.; Wang, T. H.; Guo, M.; et al. LLM and simulation as bilevel optimizers: a new paradigm to advance physical scientific discovery. *arXiv* **2024**, arXiv:240509783. DOI
 104. Shojaei, P.; Nguyen, N. H.; Meidani, K.; Barati Farimani, A.; Doan, K. D.; Reddy, C. K. LLM-SRBench: a new benchmark for scientific equation discovery with large language models. In Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada; 2025, pp. 55325-59. DOI
 105. Chen, Y.; Piekos, P.; Ostaszewski, M.; Laakom, F.; Schmidhuber, J. PhysGym: benchmarking LLMs in interactive physics discovery with controlled priors. *arXiv* **2025**, arXiv:250715550. DOI
 106. Abhyankar, N.; Shojaei, P.; Reddy, C. K. LLM-FE: automated feature engineering for tabular data with LLMs as evolutionary optimizers. *arXiv* **2025**, arXiv:250314434. DOI
 107. Holt, S.; Qian, Z.; Liu, T.; Weatherall, J.; van der Schaar, M. Data-driven discovery of dynamical systems in pharmacology using large language models. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 96325-66. DOI
 108. Spiliadis, S.; Rogers, J.; Boschetti, F.; et al. Data-driven discovery of mechanistic ecosystem models with LLMs. *bioRxiv* **2025**, 2025-7. DOI
 109. Kramer, S.; Cerrato, M.; Brugger, J.; Džeroski, S.; King, R. D. Automated scientific discovery: from equation discovery to autonomous discovery systems. *Mach. Learn.* **2026**, *115*, 109. DOI
 110. Taskin, B.; Xie, W.; Lazebnik, T. Knowledge integration for physics-informed symbolic regression using pre-trained large language models. *Sci. Rep.* **2026**, *16*, 1614. DOI
 111. Yang, J.; Venkatachalam, O.; Kianezhad, M.; Vadgama, S.; Yu, R. Think like a scientist: physics-guided LLM agent for equation discovery. *arXiv* **2026**, arXiv:2602.12259. DOI
 112. Bideh, A. Z.; Gryak, J. LLM-ODE: data-driven discovery of dynamical systems with large language models; 2026. DOI

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.