Identification of Hybrid Systems: A Tutorial

Methods, Challenges, and Recent Advances

Based on Paoletti et al. 2007
with Recent Developments (2020-2024)

Workshop on Automata Learning

September 10, 2025

Outline

What are Hybrid Systems?

Definition

Hybrid systems are heterogeneous dynamical systems whose behavior is determined by interacting continuous and discrete dynamics Paoletti et al. 2007.

Continuous dynamics:

- Variables from continuous sets
- Differential/difference equations
- Traditional control theory

Discrete dynamics:

- Variables from finite sets
- Logic conditions
- Event-driven behavior

Critical CPS Applications

- **Autonomous vehicles**: Safety verification
- **Industrial robotics**: Multi-modal control
- **Medical devices**: Critical monitoring
- **Smart grids**: Energy management

Why Hybrid System Identification?

Challenges:

- First-principles modeling often impossible
- Complex interactions between continuous and discrete dynamics
- Discontinuous behaviors
- Multiple operating modes

Solutions:

- Data-driven approaches
- Automatic model structure detection
- Parameter estimation from observations
- Universal approximation properties

Key Insight - The "Learn-First" Necessity

The "model-first" imperative in modern Cyber-Physical Systems (CPS) creates a "learn-first" necessity when models are absent. Hybrid system identification provides automated means to learn necessary models from data for formal verification, control design, and digital twin creation.

Piecewise Affine Systems - State Space Form

Discrete-time Switched Affine Model

$$x_{k+1} = A_{\sigma(k)} x_k + B_{\sigma(k)} u_k + f_{\sigma(k)} + w_k \quad (1)$$

$$y_k = C_{\sigma(k)} x_k + D_{\sigma(k)} u_k + g_{\sigma(k)} + v_k \quad (2)$$

Where:

- $x_k \in \mathbb{R}^n$: continuous state
- $u_k \in \mathbb{R}^p$: input
- $y_k \in \mathbb{R}^q$: output
- $\sigma(k) \in \{1, \ldots, s\}$: discrete state
- w_k, v_k: noise terms

PWA Switching Rule

$$\sigma(k) = i \quad \text{iff} \quad \begin{bmatrix} x_k \\ u_k \end{bmatrix} \in \Omega_i$$

PWA Systems - Input-Output Form

PWARX Model

$$y_k = \theta_{\sigma(k)}^T \begin{bmatrix} r_k \\ 1 \end{bmatrix} + e_k$$

Where:

- $r_k = [y_{k-1}^T \dots y_{k-n_a}^T \; u_k^T \; u_{k-1}^T \dots u_{k-n_b}^T]^T$: regression vector
- $\sigma(k) = i$ iff $r_k \in \mathcal{R}_i$
- $\{\mathcal{R}_i\}_{i=1}^s$: polyhedral partition of regressor domain

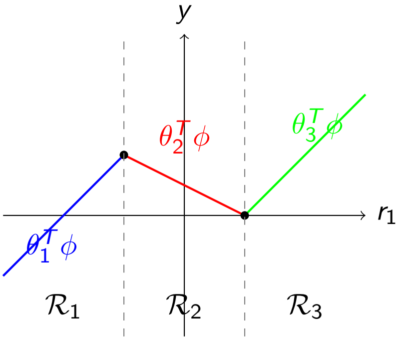
Polyhedral Regions

$$\mathcal{R}_i = \left\{ r \in \mathbb{R}^d : H_i \begin{bmatrix} r \\ 1 \end{bmatrix} \leq 0 \right\}$$

Key Property

PWA models have universal approximation properties and can represent various hybrid system

PWA Map Visualization



Key Features:

- Discontinuous PWA map with $s = 3$ regions
- Each region has its own affine submodel
- Switching occurs at hyperplane boundaries

General Identification Problem

Problem Statement

Given input-output data $\{(u_k, y_k)\}_{k=1}^{N}$, estimate:

1. Model orders n_a, n_b
2. Number of submodels s
3. Parameter vectors $\{\theta_i\}_{i=1}^{s}$
4. Polyhedral regions $\{\mathcal{R}_i\}_{i=1}^{s}$
5. Discrete state sequence $\{\sigma(k)\}_{k=1}^{N}$

Key Challenges

1. **Classification problem**: Which data point belongs to which submodel?
2. **Parameter estimation**: How to estimate submodel parameters?
3. **Region estimation**: How to determine polyhedral boundaries?
4. **Model selection**: How to choose the number of submodels s?

Two Approaches to Region Definition

Approach 1: Fixed Partition

- Regions defined a priori
- Simple data classification
- Standard linear identification
- Exponential growth with dimension

Advantages:

- Computationally simple
- Well-understood theory

Disadvantages:

- Curse of dimensionality
- Many regions may be empty

Approach 2: Estimated Partition

- Regions estimated from data
- Regions shaped to data clusters
- Coupled classification and estimation
- Computationally challenging

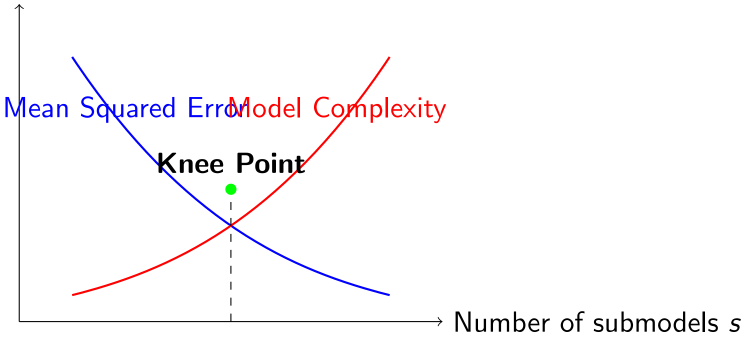
Advantages:

- Data-adaptive regions
- Fewer regions needed

Disadvantages:

- Nonconvex optimization
- Local minima issues

Trade-off Between Accuracy and Complexity



Key Insight: The optimal number of submodels is typically chosen at the "knee" of the complexity-accuracy trade-off curve.

MIN PFS Problem Amaldi and Mattavelli 2002

Find the minimum number s of parameter vectors $\{\theta_i\}$ such that:

Modern Identification Landscape

Method	Dynamics	Key Innovation	Paradigm
DAINARX	Nonlinear NARX	Model fittability principle	Threshold-free
FaMoS	Linear ARX	Decision tree interpretability	Fast baseline
LearnHA	Polynomial ODEs	Handles resets	Complete HA learnin
HySynth	Affine ODEs	ϵ-capture guarantees	Online synthesis
Classical (Paoletti)	PWA/SARX	Four procedures survey	Foundation methods

Canonical Identification Pipeline:

1. Trace Segmentation $\rightarrow$ partition data into mode segments
2. Segment Clustering $\rightarrow$ group segments by same dynamics
3. Mode Characterization $\rightarrow$ learn flow functions F_q
4. Guard/Reset Inference $\rightarrow$ learn discrete transitions

Key Distinction

- **Algebraic:** Focuses on SARX models (switching sequence estimation)

The Evolution: From Heuristics to Principles

Traditional Heuristics

Derivative-based:

- Detect "drastic variation"
- User-defined thresholds
- Noise sensitive

Similarity-based:

- Dynamic Time Warping
- Signal shape comparison
- Can misgroup dynamics

Modern Principles

Model Fittability:

- Mathematical consistency
- No arbitrary thresholds
- Data-driven robustness

Model Mergeability:

- Test shared dynamics
- Direct mathematical test
- Formal guarantees

Key Trend

From fragile heuristics to robust mathematical principles.

Classical Algebraic Procedure - Main Idea

Key Insight

View identification of multiple ARX models as identification of a single "lifted" ARX model that encodes all submodels simultaneously.

Hybrid Decoupling Constraint

If data are generated by model with s submodels, then:

$$\prod_{i=1}^{s}(b_i^T z_k) = 0, \quad \forall k$$

where $b_i = [1, \theta_i^T]^T$ and $z_k = [-y_k, \phi_k^T]^T$.

Hybrid Decoupling Polynomial

$$p_s(z) = \prod_{i=1}^{s}(b_i^T z) = h^T \nu_s(z)$$

Algebraic Procedure - Advantages & Limitations

Advantages

- Closed-form solution
- No initialization required
- Can estimate model orders
- Can estimate number of submodels
- Provably correct (noiseless case)

Limitations

- Sensitive to noise
- Only for SARX models
- May not work with nonlinear disturbances
- Classification may be suboptimal for PWARX

Number of Submodels

$$s = \arg\min\{i : \operatorname{rank}(L_i(K)) = M_i(K) - 1\}$$

When to Use

- System truly switched affine
- Moderate noise levels
- Unknown model structure
- Need automatic model selection

DAINARX - Revolutionary Threshold-Free Approach

Key Innovation: Model Fittability Principle

Core Idea: Replace subjective thresholds with objective mathematical consistency. A changepoint occurs when a growing data segment can no longer be explained by a single NARX model.

DAINARX Pipeline:

1. **Model Fittability Segmentation**: Detect changepoints when NARX model achieves near-zero error on growing window
2. **Model Mergeability Clustering**: Group segments that can be fit by same NARX instance
3. **Nonlinear Mode Learning**: Identify high-order NARX dynamics for each cluster
4. **Guard/Reset Inference**: Learn state-dependent switching conditions from transition data

Revolutionary Features

- **Derivative-agnostic**: Works directly with sampled data

Clustering-based Procedure - Key Properties

Advantages

- No prior knowledge needed
- Single tuning parameter (c)
- Can distinguish submodels with same parameters in different regions
- Handles discontinuous dynamics

Parameter c Selection

- Small c: Less noise filtering, fewer mixed datasets
- Large c: More noise filtering, more mixed datasets
- Trade-off between noise and purity

Limitations

- Requires fixed s, n_a, n_b
- Performance depends on ratio of mixed/pure local datasets
- May fail with overestimated model orders

When to Use

- No prior system knowledge
- Prescribed model structure
- Need to distinguish submodels in different regions

FaMoS - Fast Model Learning for CPS

Key Innovation: Decision Tree Representation

Goal: Fast identification of linear ARX dynamics with human-interpretable switching logic represented as decision trees.

FaMoS Key Features:

- High-order linear ARX dynamics
- Decision tree switching logic
- Fast baseline for hybrid CPS
- Human-interpretable structure
- Scalable to industrial systems

Classification Rule

$$\sigma(k) = i^* \quad \text{where } i^* = \arg\max_{i=1,\dots,s} p((y_k, r_k)|\sigma(k) = i)$$

Bayes Update

Bayesian Procedure - Advanced Features

Likelihood Function

$$p((y_k, r_k)|\theta) = p_e(y_k - \theta^T \phi_k)$$

where $p_e(\cdot)$ is the pdf of the noise term.

Misclassification Weights for Region Estimation

For data point (y_k, r_k) attributed to mode i, the price for misclassification into mode j is:

$$\nu_{i,j}(r_k) = \log \frac{p((y_k, r_k)|\sigma(k) = i)}{p((y_k, r_k)|\sigma(k) = j)}$$

Implementation:

- Particle filtering algorithms for numerical implementation
- Modified Multicategory RLP (MRLP) for region estimation
- Automatic weight computation reduces misclassification penalties

Bayesian Procedure - Evaluation

Advantages

- Incorporates prior knowledge naturally
- Automatic misclassification weights
- Probabilistic framework
- Various parameter estimates available (MAP, expectation)

Limitations

- Requires fixed s, n_a, n_b
- Sensitive to initialization
- Requires specification of prior pdfs
- Computational complexity with particle filters

Applications

Successfully applied to:

- Pick-and-place machines Juloski et al. 2004
- Industrial process control

When to Use

- Prior knowledge available
- Physical insight into system modes
- Need probabilistic uncertainty quantification
- Can tolerate computational cost

LearnHA - Complete Hybrid Automata Learning

Key Innovation: Full HA Synthesis

Capability: Learn complete hybrid automata with polynomial ODEs, including external inputs and state resets - features often missing from other frameworks.

LearnHA Capabilities:

- First-order polynomial ODEs
- External input handling
- Linear state resets
- Complete hybrid automaton synthesis
- DTW-based segment clustering

Parameter Estimation

$$\theta_i = \arg\min_{\theta} \max_{(y_k, r_k) \in D_i} |y_k - \phi_k^T \theta|$$

(ℓ_∞ projection estimator)

Bounded-error Procedure - Model Reduction

Submodel Merging

Merge submodels i and j if:

$$\alpha_{i,j} = \frac{\|\theta_i - \theta_j\|}{\min\{\|\theta_i\|, \|\theta_j\|\}} < \alpha$$

Submodel Removal

Remove submodel i if cardinality of data set D_i is less than βN.

Outlier Detection

Data points that cannot satisfy the bound δ are automatically discarded during classification, enabling outlier detection.

Multi-output Extension:

$$\|y_k - f(r_k)\|_\infty \leq \delta, \quad \forall k$$

Bounded-error Procedure - Evaluation

Advantages

- Prescribed accuracy through δ
- Automatic outlier detection
- Model complexity control
- No prior knowledge needed
- Multi-output capability

Limitations

- Requires fixed n_a, n_b
- Suboptimal MIN PFS solution
- Multiple tuning parameters
- May need several attempts for good results

Trade-off Control

- Smaller $\delta \rightarrow$ more submodels, better fit
- Larger $\delta \rightarrow$ fewer submodels, worse fit

When to Use

- Need prescribed accuracy
- Presence of outliers
- Approximating nonlinear dynamics
- Multi-output systems

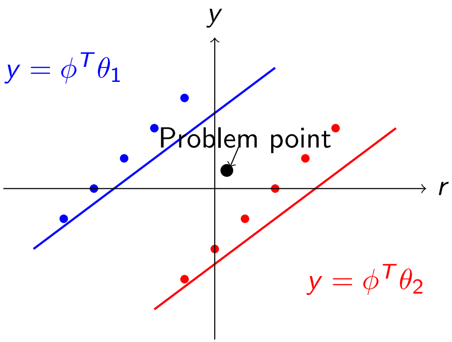
Comprehensive Method Comparison

Method	Dynamics	Scope	H.O.	Reset	Input	Key Feature
DAINARX	Nonlinear NARX	HA	Y	Y	Y	Threshold-free
FaMoS	Linear ARX	HA	Y	N	Y	Decision trees
LearnHA	Polynomial	HA	N	Y	Y	Complete HA
HySynth	Affine	HA	N	N	N	ϵ-capture
Classical	PWA/Linear	SA	$\pm$	$\pm$	$\pm$	Foundation

Trade-off Insight

Expressiveness vs. Tractability: Linear methods are fast but limited. Nonlinear methods are powerful but require more assumptions.

The Classification Challenge



Intersecting Submodels Problem

Data points near submodel intersections can be assigned to either submodel, leading to:

- Wrong data attribution
- Non-linearly separable clusters
- Poor region estimation

Linear Separation Techniques

After data classification, we need to estimate polyhedral regions. Two approaches:

Pairwise Separation

For each pair (A_i, A_j) of data clusters, find hyperplane:

$$w^T r_k + \gamma > 0, \quad \forall r_k \in A_i$$

$$w^T r_k + \gamma < 0, \quad \forall r_k \in A_j$$

Advantages:

- Computationally efficient
- Parallelizable

Disadvantages:

- May create holes in partition

Multi-class Separation

Find s classification functions simultaneously such that class i function is maximal for data in A_i.

Methods:

- Multi-category SVM (M-SVM) Cortes and Vapnik 1995
- Multi-category RLP (M-RLP) Bennett and Mangasarian 1992

Advantages:

- Guaranteed complete partition
- Global optimality

Handling Non-separable Data

Robust Linear Programming (RLP)

When data are not linearly separable:

$$\min_{w,\gamma,v_k} \quad \sum_k c_k v_k \tag{3}$$

$$\text{s.t.} \quad z_k[w^T r_k + \gamma] \geq 1 - v_k \tag{4}$$

$$v_k \geq 0, \quad \forall k \tag{5}$$

where $z_k = 1$ if $r_k \in A_i$, $z_k = -1$ if $r_k \in A_j$.

- v_k: slack variables (misclassification errors)
- c_k: misclassification weights (can be adaptive)
- Reduces to standard SVM with quadratic objective

Bayesian Extension

$p((y_k, r_k)|\sigma(k)=i)$

Recent Advances in Hybrid System Identification

Methodological Maturation

From Heuristics to Principles:

- **Traditional**: Derivative variation, signal similarity (DTW), user thresholds
- **Modern**: Model fittability principle, mathematical consistency
- **Benefits**: Robustness, formal guarantees, reduced manual tuning
- **Challenge**: Computational complexity, template requirements

Integration with Modern ML

Deep Learning Opportunities:

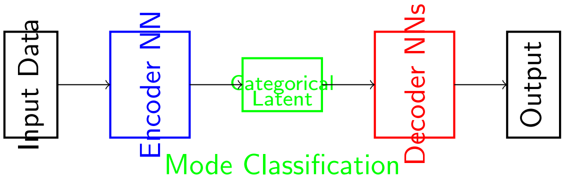
- Neural networks for complex flow dynamics approximation
- Foundation models pre-trained on diverse physical systems
- Transfer learning for CPS-specific fine-tuning
- Grey-box modeling: interpretability + expressiveness

Neural Network-Based PWA Identification

Variational Autoencoder Approach

Key Idea: Use VAE with specialized structure where:

- Encoder: Provides categorical latent variables representing discrete modes
- Decoder: Set of neural networks, each corresponding to a local submodel
- Latent space: Interpretable representation of hybrid system modes



Advantages:

- Automatic feature learning
- Joint mode classification and parameter estimation

PWA Decomposition of Neural Networks

Reverse Problem

Objective: Given a trained neural network, compute a low-complexity PWA function that closely approximates it.

F-16 Jet Benchmark

Successfully demonstrated on nonlinear system identification benchmark using F-16 aircraft data:

- Original neural network: High complexity, black-box
- PWA approximation: Interpretable, suitable for controller synthesis
- Performance: Maintains approximation accuracy with reduced complexity

Benefits:

- Combines machine learning applicability with PWA controller synthesis
- Interpretable approximation of neural network decisions
- Enables formal verification and analysis

Online and Distributed Identification

Online Identification Challenges

Requirements:

- Real-time parameter adaptation
- Active mode recognition during operation
- Handling of unknown switching sequences
- Computational efficiency for embedded systems

Recent Solutions

Integral Concurrent Learning Yang et al. 2021:

- Online estimation of continuous-time PWA systems
- Simultaneous identification of: number of subsystems, switching sequence, parameters
- Concurrent learning identifiers with theoretical guarantees

Distributed Identification:

Method Selection Guidelines

Scenario	Method	Advantage	Limitation
Unknown structure	DAINARX	Threshold-free	Template needed
Fast interpretable	FaMoS	Decision trees	Linear only
Complete HA needed	LearnHA	Resets & inputs	First-order
Online learning	HySynth	ϵ-guarantees	Affine only
Clean classical data	Algebraic	Closed-form	Noise sensitive

Hybrid Approach

Combine methods: e.g., Classical initialization + Modern refinement

Real-world Applications

Industrial CPS

Pick-and-place: Juloski et al. 2004

- Electronic component placement
- Multiple operating modes

Automotive: Borrelli et al. 2006

- Traction control systems
- Engine/transmission control

Other Domains

Computer Vision:

- Motion segmentation
- Video analysis

Biology: Ferrari-Trecate 2007

- Gene regulatory networks
- Metabolic pathways

Success Factors

Proper method selection, careful tuning, validation, domain expertise integration.

Grand Challenges from the Survey

The Noise and Uncertainty Problem

- Most methods assume clean, noise-free data
- Real sensors have errors and disturbances
- Noise creates spurious mode switches
- Need robust methods with noise modeling

The Template Conundrum

- Advanced methods require user-provided templates
- Contradicts "black-box" identification premise
- Need automatic structure discovery
- Goal: Fully automated model learning

Scalability and Formal Guarantees

- **Curse of dimensionality**: Industrial CPS with dozens of variables/modes

Future Research Directions

Deep Learning Integration

- Neural networks enable arbitrary nonlinearities
- Black-box nature conflicts with verification
- Requires massive datasets
- Solution: Hybrid interpretable approaches

Foundation Models & Digital Twins

- Pre-trained on diverse physical systems
- Fine-tune with small CPS datasets
- Grey-box modeling paradigm
- High-fidelity digital replicas

Theoretical Development

- Persistence of excitation for hybrid systems

Open Problems

1. **Optimal input design**: How to design experiments that maximize information for hybrid system identification?
2. **High-dimensional systems**: Scalable methods for systems with hundreds of states and inputs?
3. **Distributed identification**: How to identify large-scale networked hybrid systems?
4. **Safety and verification**: How to ensure identified models are safe for control applications?
5. **Online adaptation**: How to continuously adapt models as system behavior changes?
6. **Multi-modal data**: Integration of different data types (time series, images, text) for hybrid system identification?

Ultimate Vision: Push-Button Synthesis

Goal: Raw noisy sensor data $\rightarrow$ Formal hybrid automaton with verifiable safety guarantees.
Requires: Synthesis of DAINARX principles + PAC learning theory + automated verification.

Summary

Key Takeaways

- **CPS imperative**: "Model-first" engineering creates "learn-first" necessity for safety-critical systems
- **Methodological evolution**: From fragile heuristics to robust mathematical principles
- **Modern state-of-art**: DAINARX (threshold-free), FaMoS (interpretable), LearnHA (complete)
- **Future integration**: Foundation models + formal verification for automated CPS design

Evolution Timeline

1. **Classical Foundation (2007)**: Paoletti et al. four procedures, PWA/SARX focus
2. **Threshold-Free Revolution**: DAINARX model fittability principle
3. **Specialized Methods**: FaMoS (speed), LearnHA (completeness), HySynth (guarantees)
4. **Future Vision**: Foundation models + automated verification

Questions and Discussion

Further Reading

Classical Reference:

- Paoletti et al. 2007

Recent Surveys:

- Lauer and Bloch 2023
- Chen, Ljung, et al. 2024

Software Tools:

- Hybrid Identification Toolbox (HIT) - Ferrari-Trecate
- Various implementations available on GitHub

References I

- Amaldi, Edoardo and Marco Mattavelli (2002). "The MIN PFS problem and piecewise linear model estimation". In: *Discrete Applied Mathematics* 118.1-2, pp. 115–143.
- Bennett, Kristin P and Olvi L Mangasarian (1992). "Robust linear programming discrimination of two linearly inseparable sets". In: *Optimization Methods and Software* 1.1, pp. 23–34.
- Borrelli, Francesco et al. (2006). "An MPC/hybrid system approach to traction control". In: *IEEE Transactions on Control Systems Technology* 14.3, pp. 541–552.
- Chen, Tianshi, Michael Sebek Andersen, et al. (2021). "Piecewise affine system identification based on centralized and distributed strategies". In: *Mathematical Problems in Engineering* 2021.
- Chen, Tianshi, Lennart Ljung, and Koen Tiels (2024). "Deep networks for system identification: A survey". In: *Automatica* 161, p. 111489.
- Cortes, Corinna and Vladimir Vapnik (1995). "Support-vector networks". In: *Machine Learning* 20.3, pp. 273–297.

References II

- Ferrari-Trecate, Giancarlo (2007). "Hybrid identification methods for the reconstruction of genetic regulatory networks". In.
- Ferrari-Trecate, Giancarlo and Marco Schinkel (2003). "Conditions of optimal classification for piecewise affine regression". In: *Hybrid Systems: Computation and Control*, pp. 188–202.
- Juloski, Aleksandar, WPMH Heemels, and Giancarlo Ferrari-Trecate (2004). "Data-based hybrid modelling of the component placement process in pick-and-place machines". In: *Control Engineering Practice* 12.10, pp. 1241–1252.
- Lauer, Fabien and Gérard Bloch (2023). "Models and methods for hybrid system identification: a systematic survey". In: *IFAC-PapersOnLine* 56.2, pp. 7577–7589.
- Lu, Lei et al. (2020). "A PWA model identification method for nonlinear systems using hierarchical clustering based on the gap metric". In: *Computers & Chemical Engineering* 140, p. 106895.

References III

- Ma, Yi and René Vidal (2005). "Identification of deterministic switched ARX systems via identification of algebraic varieties". In: *Hybrid Systems: Computation and Control*, pp. 449–465.
- Paoletti, Simone et al. (2007). "Identification of hybrid systems: a tutorial". In: *European Journal of Control* 13.2-3, pp. 242–260.
- Vidal, René et al. (2003). "An algebraic geometric approach to the identification of a class of linear hybrid systems". In: *Proceedings of the 42nd IEEE Conference on Decision and Control*. IEEE, pp. 167–172.
- Yang, Xiaodi, Jan Dufek, and Farshad Arvin (2021). "Online identification of piecewise affine systems using integral concurrent learning". In: *IEEE Transactions on Automatic Control* 66.9, pp. 4377–4384.