Models and methods for hybrid system identification
a systematic survey

Moradvandi, Ali; Lindeboom, Ralph E.F.; Abraham, Edo; De Schutter, Bart

DOI
10.1016/j.ifacol.2023.10.1553

Publication date
2023

Document Version
Final published version

Published in
IFAC-PapersOnLine

Citation (APA)
Moradvandi, A., Lindeboom, R. E. F., Abraham, E., & De Schutter, B. (2023). Models and methods for hybrid system identification: a systematic survey. In H. Ishii, Y. Ebihara, J. Imura, & M. Yamakita (Eds.), *IFAC-PapersOnLine* (2 ed., pp. 95-107). (IFAC-PapersOnLine; Vol. 56, No. 2). Elsevier. https://doi.org/10.1016/j.ifacol.2023.10.1553

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 95–107

Models and methods for hybrid system identification: a systematic survey ⋆

Ali Moradvandi *,** **Ralph E.F. Lindeboom** ** **Edo Abraham** **
Bart De Schutter *

* *Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands.*
** *Department of Water Management, Delft University of Technology, Delft, The Netherlands,*
(e-mails: a.moradvandi, r.e.f.lindeboom, e.abraham, b.deschutter@tudelft.nl)

Abstract: Dynamical systems and processes that either exhibit non-smooth behaviours (e.g. through logic control or natural phenomena) or work in different modes of operation are usually represented using hybrid systems models, i.e. mathematical models that combine continuous dynamics with discrete-event dynamics. Identification of a hybrid system includes finding switching patterns and identification of model parameters to obtain a data-driven model. This survey paper provides a systematic review of models (how to parameterize the system) and methods (how to identify unknown parameters) proposed for hybrid system identification with an exposition of recent advances and developments, and further research directions.

Keywords: Hybrid systems; system identification; piecewise-affine systems; switched systems; jump systems.

1. INTRODUCTION

Intertwined continuous and discrete behaviors can be represented as hybrid dynamical, which can either model systems with non-smooth behaviors or approximate systems with nonlinearities. The discrete behavior can be represented as a switching pattern (mode) using a finite number of values as countable state variables that orchestrates over corresponding continuous subsystems (submodels). Hybrid system identification is a twofold problem: 1) estimating the parameters of submodels and 2) determining the switching patterns. While estimation of parameters of the model is the objective in classical system identification, hybrid system identification also requires the estimation of switching patterns. In other words, the hybrid system identification problem, where the switching mechanisms are known, reduces to the conventional system identification.

The first step is the determination of the modelling structure, i.e. model parameterization. The structure of submodels and switching mechanisms can be parameterized based on: 1) identifying what model structure is well-fitted to capture the dynamical behavior of the system, and 2) whether the main purpose of system identification is prediction or model-based control. These decisions delineate the arena of hybrid system identification and once a parameterization has been chosen, a methodology can be chosen to solve the identification problem.

Parameterized submodels can be classified into three main groups: input-output (Piga et al. (2020b)), state space (Du et al. (2021b)), and probabilistic models (Breschi et al. (2019)). The input-output model complexity ranges from Auto-Regressive exogenous (ARX) models (Du et al. (2018)) to the more complex Box-Jenkins (BJ) models (Piga et al. (2020b)) and nonlinear models (Bianchi et al. (2020b)). Complexity in the structure of a parameterized model to include dynamical noises and disturbances, and delay is a trend in recent years. State-space models, which are a more control-oriented model structure, have been widely discussed for linear (Sefidmazgi et al. (2016)), affine (Du et al. (2021b)) or nonlinear (Du et al. (2021a)) representations. Furthermore, hybrid model representations in a probabilistic setting have recently drawn attention, since they can describe parametric uncertainties and external disturbances, and the number of parameters can be adapted as more data is collected (Piga et al. (2020a)).

Similarly, the switching mechanism can be represented differently, depending on the system dynamics, switching behaviors, and purpose of modelling. Switching logics can be generally classified in three groups: polyhedral partitions (linearly or affinely partitioned) (Breschi et al. (2016b)), random switching (Liu et al. (2021)), and event-driven (deterministic (Basiri et al. (2018))). For hybrid system identification, the switching logic is mostly modelled either randomly (state-independent) or piecewise affinely (state-dependent). In the first part of this survey, a systematic discussion on parameterization of hybrid systems will be made, and various classes of hybrid systems for identification will be reviewed.

⋆ The research was conducted under the project SARASWATI2.0, WP4 (automation and control), which is funded by the European Horizon 2020 Framework Programme (grant agreement number 821427).

10.1016/j.ifacol.2023.10.1553

As a second aspect, besides system description, new recently-developed identification methods are reviewed and categorized into four groups. Taking into account the groups discussed in the previous survey by Garulli et al. (2012), most identification approaches fit in the so-called optimization-based framework. Clustering-based (Du et al. (2020)) and probabilistic (Chen et al. (2020a,b)) methods are the other identification methods that are widely discussed in recent publications. The other methods that cannot be covered by the aforementioned classes are algebraic, bounded-error-based, continuous-time, and neural network approaches. An innovative trend is either combination or generalization of the approaches (Tang and Dong (2020); Liu et al. (2022); Piga et al. (2020b)) for either dealing with new forms of parameterization or improving efficiency of computational burden and accuracy of the methods. The theory and mathematics of various well-established hybrid system identification methods have been discussed by Lauer and Bloch (2018), which was an inspiration for systematically modifying and expanding the categorization in this survey in comparison with Garulli et al. (2012) and Lauer and Bloch (2018). Moreover, discussing the probabilistic parametrization and the associated likelihood-based methods are two additions to the aforementioned reviews. Furthermore, the detailed specification of switching patterns, specifically piecewise affine is reviewed in both the sections on parameterization and solution method of this survey.

The paper is organized as follows. Models of hybrid systems are presented in Section 2. We discuss input-output models (Section 2.1), state-space models (Section 2.2), switching mechanisms (Section 2.3), and probabilistic models (Section 2.4), and a comparison in Section 2.5. A review of methods is given in Section 3. This section consists of optimization-based methods (Section 3.1), clustering-based methods (Section 3.2), likelihood-based methods (Section 3.3), other methods (Section 3.4), and a comparison of important papers (Section 3.5). In the last section, conclusions are drawn to show current and future research directions.

2. MODELS OF HYBRID SYSTEMS: PARAMETERIZATION

2.1 Input-output models

A quite wide spectrum of hybrid systems can be represented in input-output (I/O) form as

$$y_k = f_{q_k}(x_k) + \varepsilon_k, \tag{1}$$

where $x_k \in \mathbb{R}^{n_x}$ is the regressor or input, $y_k \in \mathbb{R}^{n_y}$ is the output, and $\varepsilon_k \in \mathbb{R}^{n_y}$ is a noise vector, in which k denotes the index (e.g. timestep) of the sequence, and $q_k \in \{1, ..., N\}$ is the switching signal that determines which vector field, f_{q_k}, is active at timestep k. The vector field f_{q_k} can be either linear or nonlinear. In case of linear functionality of hybrid systems, the vector field can be represented by $f_i(x_k) = x_k^T \theta_i$, in which $q_k = i$ (i-th mode).

The Switched Auto-Regressive eXogenous (SARX) model is the simplest and widely-used parameterization of hybrid systems for the identification problem. SARX models are a combination of several different ARX models defined as submodels. Discrete-time Single-Input Single-Output (SISO) SARX dynamical system can be expressed as

$$y_k = x_k^T \theta_{q_k} + \varepsilon_k, \tag{2}$$

where $\tilde{x}_k = [y_{k-1}, ..., y_{k-n_a}, u_{k-1}, ..., u_{k-n_b}]^T \in \mathbb{R}^{n_a+n_b}$ and $x_k = [\tilde{x}_k^T, 1]^T \in \mathbb{R}^{n_a+n_b+1}$ are the regressor and the extended regressor vectors, respectively, in which n_a and n_b are the orders of the system, the constant "one" is to account for different offsets of the modes, and y_k and u_k are the output and input at timestep k, respectively. $q_k \in \mathcal{S} = \{1, ..., N\}$ denotes the active mode at timestep k and $\theta_i = [a_{i1}, ..., a_{in_a}, b_{i1}, ..., b_{in_b}, c_i]^T \in \mathbb{R}^{n_a+n_b+1}$ represents the parameter vector for the i-th mode, in which $i \in \mathcal{S}$.

Jump Box-Jenkins (JBJ) models have a more general and flexible structure for representing hybrid systems in comparison with the mentioned auto-regressive (AR) models, since they include both the moving average and auto-regressive terms to model dynamics of disturbances and noises. For this class of systems, the output is noise-corrupted, while the noise term is a dynamical noise. The noise-corrupted output, y_k, can be written with respect to the noise-free output, $\hat{y}_k$, and the noise term, v_k as follows:

$$y_k = \hat{y}_k + \varepsilon_k, \tag{3}$$

where the noise-free output, $\hat{y}_k$, is modelled based on the input u_k as follows:

$$\hat{y}_k = G(q^{-1}, \theta_i) u_k, \tag{4}$$

and noise term, v_k, is written as follows:

$$\varepsilon_k = H(q^{-1}, \theta_i) v_k, \tag{5}$$

where $G(q^{-1}, \theta_i)$ and $H(q^{-1}, \theta_i)$ are the linear filters in which θ_i denotes the i-th mode at timestep k. These linear filters are rational functions of the time shift operator q^{-1} (i.e. $q^{-d} x_k = x_{k-d}$ for $d \in \mathbb{Z}$) as written below

$$G(q^{-1}, \theta_i) = \frac{B(q^{-1}, \theta_i)}{A(q^{-1}, \theta_i)} = \frac{b_{i1} q^{-1} + ... + b_{in_b} q^{-n_b}}{1 + a_{i1} q^{-1} + ... + a_{in_a} q^{-n_a}}, \tag{6a}$$

$$H(q^{-1}, \theta_i) = \frac{C(q^{-1}, \theta_i)}{D(q^{-1}, \theta_i)} = \frac{1 + c_{i1} q^{-1} + ... + c_{in_c} q^{-n_c}}{1 + d_{i1} q^{-1} + ... + d_{in_d} q^{-n_d}}, \tag{6b}$$

where n_a, n_b, n_c, and n_d are the orders of the system. The parameters vector that should be identified for the i-th submodel in a compact form is as follows:

$$\theta_i = [a_{i1}, ..., a_{in_a}, b_{i1}, ..., b_{in_b}, c_{i1}, ..., c_{in_c}, d_{i1}, ..., d_{in_d}]^T. \tag{7}$$

Substituting (6a) and (6b) into (4) and (5) results in

$$\hat{y}_k = (1 - A(q^{-1}, \theta_i)) \hat{y}_k + B(q^{-1}, \theta_i) u_k = x_{k1}^T \theta_{i1}, \tag{8a}$$

$$\varepsilon_k = (1 - C(q^{-1}, \theta_i)) \varepsilon_k + D(q^{-1}, \theta_i) v_k = x_{k2}^T \theta_{i2} + v_k, \tag{8b}$$

where x_{k1} and x_{k2} are the regressor vector defined as follows:

$$x_{k1} = [-\hat{y}_{k-1}, ..., -\hat{y}_{k-n_a}, u_{k-1}, ..., u_{k-n_b}]^T \in \mathbb{R}^{n_a+n_b}, \tag{9a}$$

$$x_{k2} = [-\varepsilon_{k-1}, ..., -\varepsilon_{k-n_c}, v_{k-1}, ..., v_{k-n_d}]^T \in \mathbb{R}^{n_c+n_d}, \tag{9b}$$

$$x_k = [x_{k1}^T, x_{k2}^T]^T \in \mathbb{R}^{n_a+n_b+n_c+n_d}, \tag{9c}$$

and the parameters vector can be expressed as

Table 1. Switching input-output linear models.

Name	$G(q^{-1},\theta_i)$	$H(q^{-1},\theta_i)$	Reference
SFIR	$B(q^{-1},\theta_i)$	1	Liu et al. (2021)
SOE	$\frac{B(q^{-1},\theta_i)}{A(q^{-1},\theta_i)}$	1	Goudjil et al. (2017b)
SARX	$\frac{B(q^{-1},\theta_i)}{A(q^{-1},\theta_i)}$	$\frac{1}{A(q^{-1},\theta_i)}$	Du et al. (2018)
Delay-SARX	$\frac{B(q^{-1-\tau_i},\theta_i)}{A(q^{-1},\theta_i)}$	$\frac{1}{A(q^{-1},\theta_i)}$	Chen et al. (2017)
EIV-SARX	$\frac{B(q^{-1},\theta_i)}{A(q^{-1},\theta_i)}$	$\frac{1}{A(q^{-1},\theta_i)}$	Ozbay et al. (2019)
SARMAX	$\frac{B(q^{-1},\theta_i)}{A(q^{-1},\theta_i)}$	$\frac{C(q^{-1},\theta_i)}{A(q^{-1},\theta_i)}$	Hojjatinia et al. (2020)
SBJ	$\frac{B(q^{-1},\theta_i)}{A(q^{-1},\theta_i)}$	$\frac{C(q^{-1},\theta_i)}{D(q^{-1},\theta_i)}$	Piga et al. (2020b)

$$\theta_{i1} = [a_{i1}, ..., a_{in_a}, b_{i1}, ..., b_{in_b}]^T \in \mathbb{R}^{n_a+n_b}, \tag{10a}$$

$$\theta_{i2} = [c_{i1}, ..., c_{in_c}, d_{i1}, ..., d_{in_d}]^T \in \mathbb{R}^{n_c+n_d}, \tag{10b}$$

$$\theta_i = [\theta_{i1}^T, \theta_{i2}^T]^T \in \mathbb{R}^{n_a+n_b+n_c+n_d}, \tag{10c}$$

Therefore, according to (3), the model can be parameterized for the identification problem considering (2) as follows:

$$\begin{aligned} y_k &= x_{k1}^T\theta_{i1} + x_{k2}^T\theta_{i2} + v_k \\ &= x_k^T\theta_i + v_k. \end{aligned} \tag{11}$$

According to the definitions of $G(q^{-1},\theta_i)$, $H(q^{-1},\theta_i)$, $A(q^{-1},\theta_i)$, $B(q^{-1},\theta_i)$, $C(q^{-1},\theta_i)$, and $D(q^{-1},\theta_i)$ in (6a) and (6b), other classes of hybrid systems can also be defined. The switched Finite Impulse Response (SFIR) model is the simpler than SARX. The switched Auto-Regressive Moving-Average with eXogenous input (SARMAX) model is a well-structured model representing hybrid systems subject to disturbances. The switched Output-Error (SOE) model is a suitable hybrid model structure for systems subject to output measurement noise. The Error-in-Variable SARX (EIV-SARX) model is a class of hybrid systems, where input measurements are also corrupted with noise. Moreover, time-delayed models have drawn attention, since several real-world applications are subject to delays. The parameterization of the these submodels are presented in Table 1.

Besides the switching linear I/O systems, identification of switched nonlinear I/O systems has also been addressed in the literature. The switched nonlinear ARX (SNARX) model is a type of switched I/O nonlinear system that is represented by a finite set of nonlinear maps of ARX model. Considering (1), the nonlinear map, i.e. f_{q_k}, can be expressed as either polynomial expansion of all monomials of x_k up to a given order or any other (nonlinear) basis function as follows:

$$f_{q_k}(x_k) = \sum_{j=1}^{n} \vartheta_{ij}\varphi_j(x_k), \tag{12}$$

where $\varphi_j(x_k)$, $j = 1, \ldots, n$, is a nonlinear regressor, and $\vartheta_i \in \mathbb{R}^n$ is the parameter vector of the i-th submodel. The vector field, f_{q_k}, can also be expressed by a Takagi-Sugeno (TS) model or a Neural Network (NN). TS models and NNs, however, are not nonlinear in principle and the nonlinearities come from membership and activation functions in TS models and NNs, respectively. Considering weighted Gaussian membership function and element-wise sigmoid with hyperbolic tangent activation functions, TS-SARX and NN-SARX are nonlinear models. The block-oriented model is represented by Wiener and Hammerstein structure, that consists of a Hammerstein block and a Wiener block in series with a linear block in between. The switched nonlinear system in this form is formulated based on SARX as a middle linear block. These input-output nonlinear switching models are summarized in Table 2.

Table 2. Switching input-output nonlinear models.

Name	Basis	Nonlinearity	Reference
SNARX	SARX	basis function	Bianchi et al. (2020b)
TS-SARX	SARX	membership function	Wagner and Kroll (2014)
NN-SARX	SARX	activation function	Brusaferri et al. (2020)
WH-SARX	SARX	Wiener, Hammerstein	Wang et al. (2019) Zhang et al. (2016)

2.2 State-space models

Switched State-Space (SS) models provide a more meaningful representation for physical applications in comparison with switched I/O models. Moreover, most control approaches and dynamical analysis rely on SS models as simple and compact form for theoretical developments. A general form of hybrid model in SS structure is described by

$$\begin{cases} x_{k+1} &= F_{q_k}(x_k, u_k) + w_k \\ y_k &= G_{q_k}(x_k, u_k) + v_k \end{cases} \tag{13}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^p$, and $y_k \in \mathbb{R}^l$ are the continuous state, input and output of the system, respectively, $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^l$ are noise terms, $q_k \in \{1, ..., N\}$ is the switching signal that determines which vector fields, F_{q_k} and G_{q_k}, are active at timestep k. The vector fields, $F_q : \mathbb{R}^{n+p} \to \mathbb{R}^n$ and $G_q : \mathbb{R}^{n+p} \to \mathbb{R}^l$ can be either linear (and affine) or nonlinear.

Linearization of nonlinear form of (13) around operating points yields affine models. "Affine" systems can be represented in "linear" forms (i.e. without affine or bias constant), if the equilibrium points are known. The more accurate the approximation of a complex system, the higher the number of submodels required, which hinders the identification problem due to the increasing number of modes. Switched nonlinear systems in SS form can also be rewritten as a linear combination of the basis functions. A SS forms represent a more informative model for a system in comparison with input-output models. Table 3 provides an overview of state-space models.

2.3 Switching mechanisms

It is also worth to discuss switching mechanisms to review one of the important model group, i.e. piecewise affine models. Switching behavior is determined by the switching mechanism. The switching signal is defined as discrete state, i.e. $q_k \in \mathcal{S} = \{1, ..., N\}$ that determines which submodel is active at timestep k. There are a variety of switching mechanisms: exogenous, deterministic, state-driven, event-driven, time-driven, and completely random. In terms of hybrid system identification, according to what has been discussed in the literature, the switching mechanism can be considered as either continuous-state-dependent or -independent. The behaviour of an application (i.e. the source of switching) determines the mechanism of the switching signal.

Table 3. Switching state-space models.

Name	type of F_i	type of G_i	Reference
Switched affine input-to-state	affine	-	Du et al. (2021b)
Switched linear state-space	linear	linear	Sefidmazgi et al. (2016)
Switched affine state-space	affine	linear	Rui et al. (2016)
Switched nonlinear input-to-state	basis function	-	Du et al. (2021a)

The PieceWise Affine (PWA) model is a class of hybrid systems where in the discrete state depends on the continuous state. The simple and flexible structure as well as the universal properties of PWA model for approximation of any nonlinear functions with any accuracy has drawn researchers' attention to develop identification methods based on this model. In principle, the switching space in PWA model is partitioned into some regions based on the continuous states. According to (2), the mode of q_k for the i-th active mode is written based on the regressor vector, x_k, belonging to i-th set, $\boldsymbol{\chi}_i$, of a polyhedral partition $\{\boldsymbol{\chi}_i\}_{i\in\mathcal{S}}$ of the regressor space $\boldsymbol{\chi}$ as

$$q_k = i \Longleftrightarrow x_k \in \boldsymbol{\chi}_i. \tag{14}$$

In other words, a PWA function, $f : \boldsymbol{\chi} \rightarrow \mathbb{R}^{n_b}$, approximates a nonlinear function with sufficient number of modes defined by a set of polyhedra, $\boldsymbol{\chi}_i$. The partition region $\boldsymbol{\chi}_i$, can be defined based on linear classifiers as follows, which is commonly-used formulation for the partition domain. The bounded polyhedron $\boldsymbol{\chi}_i$, can be represented by a hyperplane matrix $\mathcal{H}_i \in \mathbb{R}^{n_{p_i}\times(n_x+1)}$, in which n_{p_i} is the number of hyperplanes of the corresponding partition:

$$\boldsymbol{\chi}_i = \{x \in \mathbb{R}^{n_x} : \mathcal{H}_i \begin{bmatrix} x \\ 1 \end{bmatrix} \leq 0\} \tag{15}$$

A few types of model have been used for hybrid systems with polyhedral partition as switching mechanism, which is given in Table 4. Voronoi-type partition (seeds generators), scheduling-variable space (Linear Parameter Varying (LPV)), time-partitioned region, , and input-to-state form of partition are the models of the switching mechanism. Therefore, different classes of hybrid systems such as PWFIR, PWARX (Breschi and Mejari (2020)), PWAOE (Mejari et al. (2020b)), PWA LPV-ARX (Mejari et al. (2018)), PWNL (Liu et al. (2022)), and state-space PWA (Rui et al. (2016)) models can be derived according to the switching pattern parameterization.

Besides the deterministic representation of PWA partitions, the random probability distribution can be represented over the discrete state model as a switching signal using the Dirichlet process. Moreover, for the state-independent domain, Markov switching is the other class of switching mechanism widely used. The orderly-switching pattern, in the form of a Markovian jump model, determines the switching between the submodels independent of the past (except the immediate one) information. The modes of a hybrid system can also be expressed as an event-based model, then the model is so-called event-driven. In this way, hybrid systems can mathematically model the physical "environment". The summary of mode-switching models is given in Table 4.

2.4 Probabilistic models

Probabilistic models (also so-called non-parametric model) are models that take a probability distribution into account to describe the process. This kind of the model representation, that can be expressed in various settings, is suited for the systems with available priori physical knowledge. Generally, the distribution of the output in the discrete-time form can be expressed as follows:

$$y_k \mid (x_k, \theta_{I_k}, \sigma^2_{I_k}, v_{I_k} : I_k = i) \sim \mathcal{D}(\theta_i^T x_k, \sigma_i^2, v_i) \tag{16}$$

where v_i and σ_i^2 denote degree of freedom and the variance of the distribution, respectively. Moreover, I_k is the Markov chain and $\mathcal{D}$ can be any form of distribution like Gaussian distribution and t-distribution to represent the distribution of the parameters (θ_{I_k}) based on given regressor vector (x_k). Other parameterization are derived to include some features such as delays and missing data. Various classes of hybrid systems that can be parameterized in this way range from SFIR and SARX to PWARX.

Moreover, the stable spline kernel is a way of stochastic modelling of hybrid systems via a Bayesian network. The stable spline kernel modelling can be used for PWFIR and PWARX models. The distribution of the model in the form of linear stable spline kernel can be expressed as follows:

$$y_k \mid (\{\theta_i\}_{i=1}^M, \lambda, \alpha, \sigma^2, \{\omega_k\}_{k=1}^N) \sim \mathcal{N}(\Phi_k\theta_k, \frac{1}{\sigma^{-2}}I) \tag{17}$$

where λ and α are the scalar factor and the stability parameter, respectively as the stable spline hyperparameters, ω_k denotes other hyperparameters associated with the classification, Φ_k is the matrix with rows selected by the input regressor vector with respect to the type of the input regressor chosen based on the type of the model, and σ^{-2} and I are the noise precision and the identity matrix with the appropriate dimension, respectively.

Furthermore, the submodels also can be formulated in the stochastic setting with the deterministic derivations. In this way, the a priori prediction error, $\epsilon_{k|k-1}$ and the a posteriori prediction error, $\epsilon_{k|k}$ can be expressed as

$$\epsilon_{k|k-1} = Y_k - \hat{Y}_{k|k-1} \tag{18a}$$

$$\epsilon_{k|k} = Y_k - \hat{Y}_{k|k} \tag{18b}$$

where $Y_k \in \mathbb{R}^{N\times 1}$ is the collection of outputs for all submodels at timestep k, $\hat{Y}_{k|k-1}$ and $\hat{Y}_{k|k}$ are the a priori and the a posteriori estimations, respectively. The partitions can also be represented deterministically. The representations of different model structure in the literature based on the above discussion have been summarized in Table 5.

2.5 Applications of models

The advantage of the more complex models in terms of structure complexity like SBJ and EIV-SARX models is the accuracy in prediction due to consideration of

Table 4. Mode-switching models.

Switching mechanism		Model	Reference
Polyhedral partition	Linear classifier	Regressor-vector dependent	Breschi et al. (2016b)
	Center generator	Voroini-type	Bako and Yahya (2019)
	Scheduling-variable space	Soft function	Mejari et al. (2020c)
	Batch time-based	Time-partitioned	Xu et al. (2018)
	State-space	State-input-vector dependent	Du et al. (2021b)
	Dirichlet	Probabilistic	Wågberg et al. (2015)
Random	Arbitrary	Non-modelled	Liu et al. (2021)
	Markov	Probabilistic	Chen et al. (2020a)
Event-driven		Inferred mathematical	Basiri et al. (2018)

Table 5. Representation of hybrid systems in probabilistic settings.

Model structure	Representation feature	Reference(s)
SFIR	Missing measurement included	Liu et al. (2021)
SARX	Hammerstein nonlinearity included	Ma et al. (2019)
	Delay included	Chen et al. (2020b)
	Based on t-distribution	Fan et al. (2017)
	Missing measurement included	Chen et al. (2020a)
	Error-based posteriori prediction	Goudjil et al. (2016)
SOE	Error-based posteriori prediction	Goudjil et al. (2017a)
PWFIR	Based on kernel hyperparameters	Pillonetto (2016)
PWARX	Hierarchical Bayesian	Wågberg et al. (2015)
	Based on kernel hyperparameters	Scampicchio and Pillonetto (2018)
	Error-based posteriori prediction	Yahya et al. (2020)
PWA	Bayesian inference	Piga et al. (2020a)
SBJ	Posterior distribution	Breschi et al. (2019)

either dynamical disturbances and noises or errors in the parameters.

Moreover, an advantage of switching patterns represented by polyhedron is regressor-dependent switching that can project physical behaviors of the process such as different metabolic stages (Wang et al. (2020)), different operating modes (Song et al. (2020)), and different phases of a batch process (Xu et al. (2018)). Various forms of polyhedral partition may not have a specific advantage unless a specific problem for classification is defined, e.g. Bako and Yahya (2019) in which the centers of the partition map are intuitively interpreted as operating points. Arbitrary patterns can also be used for a system that has no information on switching instants (Zhang et al. (2018)).

In addition, for control-orientated purposes, state-space representation of the submodels and the partition region (Du et al. (2021b)) has the advantage over others because of its structure for observer and controller design.

For applications with no prior knowledge on their structures, non-parametric models have an advantage because of probability distribution over parameters and noises to probabilistic model uncertainties (Fan et al. (2017)). The range of complexity in probabilistic structures is the same as input-output models (Liu et al. (2021); Breschi et al. (2019)). Different advantages of model representation in probabilistic settings are based on including system restrictions in practice, such as missing measurement (Chen et al. (2020a)) and time delay (Chen et al. (2020b)). Probabilistic models in stable spline kernel also have advantage as they mitigate the difficulty of the model order selection based on defining hyperparameters (Scampicchio and Pillonetto (2018)).

In case of existing time-varying relation between input and output measurements, scheduling variable can be introduced to extend linear time-invariant models to linear time-varying models (Mejari et al. (2020c)).

Nonlinear models have the ability of representing a nonlinear system with fewer number of parameters in a wider range. Polynomial nonlinear models are discussed by Bianchi et al. (2021) and non-parametric piecewise nonlinear models are addressed by Mazzoleni et al. (2021). These kinds of hybrid systems are a new trend in the field of system identification.

3. METHODS OF HYBRID SYSTEM IDENTIFICATION

The method used to solve the identification problem depends on the parameterization used to model the hybrid system. Subsystem identification and switching rule detection can be performed either separately, or jointly. Identification of hybrid systems, therefore, needs classification of dataset into some clusters, and estimation of the parameters of submodels for each cluster. The systematic classification of the methods has been reviewed below.

3.1 Optimization-based methods

The generic identification problem can be formulated in an optimization framework as follows:

$$\underset{\mathcal{X}_{i,k},\theta_i}{\text{minimize}} \sum_k \sum_i \mathcal{X}_{i,k}(y_k - x_k^T\theta_i)^2, \tag{19a}$$

$$\text{s.t.} \sum_i \mathcal{X}_{i,k} = 1, \forall k \tag{19b}$$

where y_k is the actual system output, x_k and θ_i are the regressor and parameter vectors, respectively, which can be constructed based on the model parameterization discussed in Section 2, and $\mathcal{X}_{i,k} \in \{0,1\}$ denotes the submodel activation for all timesteps. The goal of the optimization problem is to minimize the error of the identified output and actual output based on the available measurements i.e. the measured outputs $\{y_k\}_{k=1}^{n}$, and the measured inputs, $\{u_k\}_{k=1}^{n}$, where n is the number of measurement. It is, however, an NP-hard optimization problem, a reformulation can be made for different classes of model parameterization to make the problem computationally feasible.

Wang et al. (2019) have solved the optimization problem based on the least squares criterion for switched Hammerstein ARX models with a long-horizon and a different horizon iteration to detect the process rapid changes. Hu et al. (2015) have proposed a reformulation of the cost function in order to identify the subsystem parameters of a SARX model based on the least geometric mean squares algorithm, which is followed by a neural network classifier to label the training data.

A sum-of-norm regularized convex optimization problem for SARX models has been discussed by Hartmann et al. (2015), which is combined with Expectation-Maximization (EM) approach to cluster preliminary estimates and formulate a quadratic program to complete switching detection and parameter identification procedure. The application of EM has been also used by Tang and Dong (2020) as a first step of the two-step approach to solve a convex optimization problem for simultaneous clustering and identification. The proposed approach has been compared with a non-convex optimization algorithm proposed by Lauer (2013) and a recently-developed general optimization-based approach addressed by Yuan et al. (2019). Moreover, Xiujun et al. (2020) have developed a weighted multi-innovation the least squares algorithm for Hammerstein SARX models, which is based on EM approach for clustering.

An optimization problem has been formulated for a class of nonlinear switched ARX (SNARX) models and solved in an iterative way by Bianchi et al. (2020b). While the structure of the nonlinear model is characterized in a probability setting, a randomized method is employed to address the formulated combinatorial optimization. Since prior sample-mode assignment is required, Bianchi et al. (2021) have addressed this problem based on a two-stage randomized approach in a general framework with no a priori limit on the number of switching time instants by using a cost function that alternates between parameter and mode identification. The other heuristic approach for solving a typical heterogeneous optimization problem efficiently is similar to the discussed two-stage iterative approaches for both SARX and PWARX models proposed by Bianchi et al. (2020a).

Combining the prediction error method with a coordinate descent approach has been taken into account to address the identification problem of SARMAX by Breschi et al. (2018) in both batch and recursive ways. A close-to-optimal four-step solution has been also proposed by Amaldi et al. (2016) for the mixed-integer linear programming to fit a k-piecewise affine model with a piecewise linear separability problem. Domain partitioning based on multi-category linear classification and submodel fitting have been addressed simultaneously to guarantee solutions of the k-hyperplane clustering problem.

Paoletti et al. (2019) have formulated the identification problem of PWA models in the framework of bi-level programming, in which data classification and partition estimation are addressed in the upper level and subsystem parameters are identified in the lower level based on a prediction error criterion. An optimization-based method has been formulated by Breschi and Mejari (2020) for structure selection and identification of PWARX models, using regularization-based shrinking strategies within a coordinate-descent identification method to determine the parameters of the submodels along with their structures.

If the partition of the PWA model is considered as a Voronoi type, the least harmonic mean approach can be employed that has been discussed by Bako and Yahya (2019). Moreover, identification of time-partitioned PWA-OE models has been tackled by Xu et al. (2018) for batch processes in the framework of optimization-based algorithms. Similarly, identification of PWA-OE models has been also discussed by Mejari et al. (2020a), that a recursive bias-correction scheme to correct the bias in the ordinary least square method has been presented. While simultaneous clustering and parameter estimation are achieved within the first stage by applying bias-corrected the least squares, partitioning the regressor space is obtained via a convex optimization problem known as multi-category discrimination.

Recursive multiple least squares for simultaneous clustering and parameter estimation has been proposed by Breschi et al. (2016b) for PWA models. A linear multi-category discrimination algorithm has been considered via a Newton-like approach and an averaged stochastic gradient descent for solving the unconstrained optimization problem for batch and recursive ways. The proposed method has been extended for LPV-ARX models with linear partitioning by Breschi et al. (2016a). The convex optimization problem has been solved using a sparse estimation approach as a likelihood-based methodology in stochastic framework by Mattsson et al. (2016).

Another two-stage optimization-based method has been discussed with an iterative regularized moving-horizon approach by Naik et al. (2017) for PWARX models, and Mejari et al. (2020c) for switched LPV-ARX models. Active modes and the parameters of the submodels are optimally and recursively found by solving small-size mixed-integer quadratic-programming problems and polyhedral partitions are identified using linear multi-category discrimination.

A quite general jump model, which is PWA models with hidden Markov jumps, has been formulated in an optimization-based framework by Bemporad et al. (2018) and solved by alternating between minimizing a loss function of fitting submodel parameters with a generalized k-means algorithm and minimizing a discrete objective function for determining active modes.

3.2 Clustering-based methods

Clustering aims to divide a dataset into different subsets based on how similar they are to one another. This idea

Table 6. Optimization-based approaches.

Method	Switching mechanism	Hybrid model	Reference(s)
Different horizons least squares	Arbitrary (slow switching)	Hamm-SARX	Wang et al. (2019)
Combinatorial optimization	Arbitrary	SNARX	Bianchi et al. (2020b) Bianchi et al. (2021)
Constrained optimization	Arbitrary	SARX/PWARX	Bianchi et al. (2020a)
Prediction error method	Arbitrary/Markov	SARMAX	Breschi et al. (2018)
Discrete optimization	Linear classifier	PWA	Amaldi et al. (2016)
Nested optimization	Linear classifier	PWA	Paoletti et al. (2019)
Regularization-based optimization	Linear classifier	PWARX	Breschi and Mejari (2020)
Least harmonic mean approach	Center generator (Voronoi)	PWA	Bako and Yahya (2019)
Separable nonlinear least-squares	Time-based	PWA-OE	Xu et al. (2018)
Bias-correction approach	Linear classifier	PWA-OE	Mejari et al. (2020a)
Multiple least squares	Linear classifier Scheduling-variable space	PWA LPV-ARX	Breschi et al. (2016b) Breschi et al. (2016a)
Sparse estimation approach	Linear classifier	PWARX	Mattsson et al. (2016)
Regularized moving-horizon approach	Linear classifier Scheduling-variable space	PWARX LPV-ARX	Naik et al. (2017) Mejari et al. (2020c)
Fitting algorithm	Regression models (PWA) Statistical models (Markov jump)		Bemporad et al. (2018)
Sum-of-norm regularized optimization	Arbitrary	SARX	Hartmann et al. (2015)
EM-based sparse method	Linear classifier	PWARX	Tang and Dong (2020)
Weighted multi-innovation least squares	Arbitrary	Hamm-SARX	Xiujun et al. (2020)

is very close to the hybrid system identification problem. Papers of this section have contributions to clustering, even they are combined with other methods.

A hierarchical clustering method based on the gap metric has been proposed by Wang et al. (2020). Similarly, simultaneous submodel and optimal operation region partition estimation have been addressed based on output-error minimization in order to improve accuracy by Song et al. (2020). In this approach, local models are initially found with the least squares and clustering of local models and parameters identification are done based on the initialization using the feature vectors and weighted least squares, respectively.

Bounded-switching clustering has been discussed by Sefidmazgi et al. (2015) for SARX systems and Sefidmazgi et al. (2016) for switched state-space systems to convert the non-convex optimization problem into a binary integer programming problem using an innovative clustering method. However, the problem still includes optimization, by using bounded-switching technique, it has been easily formulated and solved by least squares and subspace identification by Sefidmazgi et al. (2015) and Sefidmazgi et al. (2016), respectively.

For PWA models, a semi-supervised clustering approach has been proposed by Du et al. (2020) to obtain the number of submodels, the initial clustered dataset, and the corresponding parameters of each submodels. The output of the clustering stage is used for a modified self-training Support Vector Machine (SVM) algorithm to identify the polyhedral partitions and the parameters of submodels. Moreover, a self-adaptive clustering algorithm has been addressed by Sellami et al. (2016). The sequential estimation procedure of the switching signal is based on an unsupervised self-adaptive classification algorithm. The core of the proposed approach is clustering based on three steps consisting of cluster creation, online cluster adaptation, and cluster fusion. Hure and Vasak (2017) developed a clustering-based identification algorithm for PWA models based on feature vectors and clustering them in which the k-mean++ algorithm is adapted for initialization. Feature vector transformation is introduced to reduce and in some cases omit partitioning in some dimensions.

Li et al. (2016) have proposed a subspace clustering approach that removes the requirement on convex regions in the conventional k-mean clustering. A block-diagonal matrix permutation algorithm is the proposed subspace algorithm to reduce the computational complexity in handling arbitrarily shaped regions. Another subspace clustering approach has been proposed by Li and Liu (2017), who employ a spectral clustering algorithm with a relaxed-permutation structure. The spectral clustering method has also been addressed by Zhang et al. (2018) for EIV-SARX models. Based on the proposed method, data points are partitioned into subsets and a manifold distance between the dynamics of each segment is computed via a Riemannian distance-like function to assign segments to clusters, and finally a common identification method is used to identify parameters of each cluster. Another subspace clustering algorithm for identification of EIV-SARX models has been presented by Ozbay et al. (2019), using sum-of-squares polynomial with Christoffel's functions to perform singular value decomposition independently of the number of data points.A subspace clustering algorithm for state-space switched systems has been proposed by Lopes et al. (2016). The hybrid Kalman filter as an interacting multiple model algorithm is used for reclassification to assign the original dataset to a specific mode and to refine model estimation at the end of the procedure.

A prototype-shaped clustering-based algorithm has been proposed by Wagner and Kroll (2014) for partitioning nonlinear Takagi-Sugeno systems. The identification process includes fuzzy c-means and Gustafson-Kessel algorithms for clustering and identification.Another application of fuzzy c-means clustering as an efficient unsupervised partitioned technique has been discussed by Shah and Adhyaru (2014) for PWARX models. The number of submodels is estimated by the proposed fuzzy clustering approach and submodel parameters are identified by weighted least squares approach based on a fuzzy distance weight matrix. Likewise, an incremental c-regression approach has been addressed by Blazic and Skrjanc (2020) as an online identification procedure. Furthermore, while PCA-guided k-Means clustering approach is a conventional clustering approach in which clusters are derived in a PCA-guided process, a fuzzy PCA-guided clustering technique has been proposed by Khanmirza et al. (2016) as a modified robust clustering method.

Classification and clustering with evolutionary algorithms for estimating switching patterns modeled by a Gaussian mixture before parameter identification with weighted and extended least squares algorithms is one of the innovative methods to address the identification problem of PWARX and PWARMAX models proposed by Barbosa et al. (2019). Classical first-order algorithms such as mirror descent algorithm or Nesterov's optimal scheme can be employed to solve the reformulated determination problem of the regions as a multi-class classification, which is discussed by Jianwang and Ramirez-Mendoza (2020). In this work, parameter estimation has been addressed via zonotope parameter identification.

A constrained clustering approach for time-partitioned PWARX model has been developed by Liu et al. (2022). The clustering optimization problem has been formulated by imposing the complete and non-overlapping partition constraints and it has been efficiently solved by employing a greedy iterative approach.

For PWNL models, a semi-supervised clustering setting has been proposed by Mazzoleni et al. (2021), which is based on a data augmentation strategy to deal with a situation when unsupervised data are not basically provided. This work is a pioneer of piecewise nonlinear regression in the domain of hybrid system identification.

3.3 Likelihood-based methods

Likelihood-based methods are formulated based on the models represented in the probabilistic form. Expectation-maximization is one of the well-and-widely-studied algorithms, which not only can be used for clustering (as reviewed before), but also for the maximum-likelihood estimation. The EM algorithm consists of two steps: E-step and M-step. Considering the unknown parameter vector, Θ, defined based on the model structure, and observed and unobserved dataset, C_{obs} and C_{uno}, the E-step calculates the conditional expectation of the log-likelihood function (known as Q-function) formulated as follows:

$$Q(\Theta \mid \Theta_{\text{old}}) = E_{C_{\text{uno}}|(C_{\text{obs}},\Theta_{\text{old}})} \left(\log P(C_{\text{uno}}, C_{\text{obs}} \mid \Theta_{\text{old}})\right) \quad (20)$$

where Θ_{old} is the parameter set calculated in the previous iteration. Besides the E-step, the M-step maximizes the Q-function with respect to parameter set written as follows:

$$\Theta = \arg\max_{\Theta} Q(\Theta \mid \Theta_{\text{old}}) \quad (21)$$

Rui, Ardeshiri, and Bazanella (2016) have proposed a framework based on the EM algorithm to identify the parameters of the model represented in PWA state-space form. A cumulative distribution function is employed to compute the probability of each submodel based on the measured samples at that timestep, and then the latent discrete state is estimated using a Kalman smoother for the computed submodel, and finally parameters are identified based on the maximization of a surrogate function for the likelihood. Fan et al. (2017) have addressed a robust identification problem of the model parameterized by a hidden Markov ARX model using EM algorithm in which student's t-distribution is imposed to the noise model for more accurate estimation. The extension from a batch to a recursive EM algorithm has been proposed by Chen et al. (2020b) for delayed SARX models to identify the parameters of the submodels, the Markov chain transition, and the time delays simultaneously.

Variational Bayesian (VB) method is another Bayesian optimization-based strategy that is used to approximate high dimensional posterior distributions instead of point-wise estimations of the parameters. The VB scheme is a more general approach in comparison with the EM approach because of approximation of parameter densities.

The identification problem of Markov-switching Hammerstein ARX models has been addressed via the VB approach by Ma et al. (2019). Estimating the unknown number of submodels and switching signals as well as approximating the distributions of the unknown submodels parameters have been tackled for SFIR models by Liu et al. (2021). Estimation of parameters distributions and point-estimation of the transition probabilities of switched Markov ARX models and construction of missing measurements have been discussed by Chen et al. (2020a) under the VB framework. Similarly, a robust VB approach for SARX models with a combination of an adjusted-tail t-distribution to deal with contaminated data with outliers has been proposed by Lu et al. (2016). Furthermore, the application of a VB approach for Markov SARX models with slowly varying time delay has been extended to identify the parameter distributions of submodels besides the transition probability matrix and unknown delays by Chen and Liu (2019).

Kernel-based stable spline is another algorithm for non-parametric models in Bayesian framework. The hybrid systems in the form of stable spline kernel can be identified by optimizing marginal likelihood via a stochastic simulation scheme. Pillonetto (2016) has proposed this approach for identification of hybrid systems. The two-step kernel-based stable spline procedure consists of data classification and distribution in the marginal likelihood optimization by exploiting the Bayesian interpretation of regularization, and reconstruction of subsystems. While the performance of the proposed method has been assessed through a Markov chain Monet Carlo approach by Pillonetto (2016), the Gibbs sampling scheme has been employed by Scampicchio and Pillonetto (2018). Scampicchio et al. (2018) have also extended it for nonlinear hybrid systems, which is capable of automatic discrimination among linear and nonlinear submodels.

In addition, non-parametric representation of hierarchical PWARX models in the Bayesian framework with respect to Dirichlet clustering properties provides probabilistic predictions with confidence intervals, which has been addressed by Wågberg et al. (2015) within a Gibbs sampling process. Furthermore, two Rao-Blackwillized sampling algorithms in batch and recursive manners for PWA models

Table 7. Clustering-based approaches.

Method	Switching mechanism	Hybrid model	Reference(s)
Hierarchical clustering	Scheduling-variable space	PWARX	Song et al. (2020) Wang et al. (2020)
Bounded-switching clustering	Arbitrary	SARX Switched SS	Sefidmazgi et al. (2015) Sefidmazgi et al. (2016)
Semi-supervised clustering	Linear classifier	PWARX	Du et al. (2020)
Unsupervised clustering	Arbitrary	SARX	Sellami et al. (2016)
Enhanced k-means++	Center generator	PWARX	Hure and Vasak (2017)
Subspace clustring	Arbitrary	bi-model PWL Switched affine EIV-SARX Switched SS	Li et al. (2016) Li and Liu (2017) Zhang et al. (2018) Ozbay et al. (2019) Lopes et al. (2016)
Fuzzy clustering	Linear classifier Linear classifier	PWA Takagi-Sugeno PWARX	Wagner and Kroll (2014) Shah and Adhyaru (2014)
Fuzzy PCA-guided clustering	Linear classifier	PWARX	Khanmirza et al. (2016)
Genetic-based clustering	Linear classifier	PWARX/PWARMAX	Barbosa et al. (2019)
Multi-class classification	Arbitrary	PWARX	Jianwang and Ramirez-Mendoza (2020)
Constrained K-means clustering	Batch time-based	PWARX	Liu et al. (2022)
Greedy semi-supervised	Linear classifier	PWNL	Mazzoleni et al. (2021)

represented in a Bayesian setting have been addressed by Piga et al. (2020a). The parameters of regressor-space partition formulated based on marginal posterior are approximated via Markov chain Monte Carlo sampling for offline learning and particle filters for online learning in batch and recursive ways. The identification problem of SBJ models has been tackled by Breschi et al. (2019); Piga et al. (2020b) using a maximum-a-posterior estimation approach. Embedding the prediction error algorithm in the likelihood framework tailored by stochastic Markov chains within a coordinate ascent method enables the identification procedure to iteratively and computationally effective (due to a suboptimal moving-horizon approach) alternate between local parameter identification and mode sequence reconstruction.

3.4 Other methods

Algebraic methods: Recent trends show a combination of algebraic methods with clustering (e.g. subspace clustering) and optimization-based algorithms. The set membership identification problem has been addressed for SARX models with prior information on the number of submodels by Ozay et al. (2015), using an algebraic procedure and combining it with a polynomial function of the unknown noise to recast the problem into constrained rank minimization form, which is a convex optimization problem. Likewise, matrix rank minimization along with an iterative partial matrix shrinkage algorithm has been presented by Konishi (2015) for identification of SRAX models. Hojjatinia et al. (2020) have proposed a similar approach but for cases with a very large number of samples affected by large levels of noise for both SARX and SARMAX models. Similarly, a non-convex optimization problem has been computational efficiently solved using an algebraic procedure and a polynomial optimization approach with sparse reformulation of the problem to jointly identify a kernel-based mapping and the corresponding continuous-state evolution of Wiener SARX models by Zhang et al. (2016).

Furthermore, an iterative algebraic geometric approach has been proposed by Nazari et al. (2016), which is built upon stochastic hybrid decoupling polynomial construction and it is shown that the problem of the linear regression can be transferred into stochastic hybrid decoupling polynomial. An algebraic procedure by constructing Hankel-like matrices and performing singular value decomposition of the Hankel matrices results in parameter estimates, which has been discussed by Sarkar et al. (2019) for switched state-space models.

Outer Bounding Ellipsoid (OBE) methods: OBE type algorithms are set membership real-time identification algorithms under assumption of unknown-but-bounded noises or disturbances. The OBE algorithm has been presented for SARX models by Goudjil et al. (2016) PWARX models by Yahya et al. (2020), and SOE models by Goudjil et al. (2017b).A two-step algorithm that proposed by Du et al. (2018) for SRAX models is another method similar to OBE. . Data assignment is carried out based on incorporating both the residual error and an upper bound of the subsystem estimation error, which is followed by a randomized algorithm to update simultaneously the parameters of the submodels.

Continuous-time identification methods: The literature reviewed above is for hybrid systems represented in discretized form, while there can be found some literature on continuous-time hybrid system identification methods. The concurrent learning technique, which has been developed in a recursive manner for PWA state space models by Kersting and Buss (2017), has been proposed to identify online continuous-time system dynamics, using the recorded and current data concurrently for adaption.

Due to necessity of state derivatives for the concurrent learning technique, extended integral concurrent learning identifier has been presented by Du et al. (2021a). The two-stage online identification of switched state space models consists of recognition of the active modes based on the projection matrix inspired by a recursive projection subspace method followed by an integral concurrent learning technique for identification of the system dynamics. Likewise, using an integral concurrent learning method for continuous-time PWA state space models has been addressed by Du et al. (2021b). Polyhedral regions are estimated by solving an optimization problem based on the parameter identification and mode recognition steps.

Table 8. Likelihood-based approaches.

Method	Switching mechanism	Hybrid model	Reference(s)
Expectation-Maximization	Linear classifier	PWA-SS	Rui et al. (2016)
	Markov	SARX	Fan et al. (2017)
	Markov	Delay-SARX	Chen et al. (2020b)
Variational Bayesian	Markov	Hamm-SARX	Ma et al. (2019)
	Arbitrary	SFIR	Liu et al. (2021)
	Markov	SARX	Chen et al. (2020a)
	Markov	Delay-SARX	Chen and Liu (2019)
Kernel-based stable spline	Linear classifier	PWFIR/PWARX	Pillonetto (2016)
	Linear classifier	PWFIR/PWARX	Scampicchio and Pillonetto (2018)
	Arbitrary	Nonlinear dynamics	Scampicchio et al. (2018)
Bayesian	Dirichlet	PWARX	Wågberg et al. (2015)
Maximum-a-posterori	Linear classifier	PWA	Piga et al. (2020a)
	Markov	SBJ	Piga et al. (2020b) Breschi et al. (2019)

A continuous-time identification method has been proposed by Goudjil et al. (2020) in which consistent submodel outputs are constructed based on a sum of sinusoids as an appropriate input signal and then parameter vector estimation is carried out by conventional continuous-time identification methods under assumption of a given number of submodels. Furthermore, Keshvari-Khor et al. (2018) have proposed an identification method for continuous-time switched state space models. The advantage of this approach is detection of switching time between two sampling instants even with low-rate sampled data.

Neural network (NN) methods: Due to the capability of NNs to represent nonlinear systems in a simpler structure with precision in approximation, Yang et al. (2017) have proposed a way to use and train NNs in modelling of nonlinear ARMA models in the form of hybrid systems, which is called multiple NARMA-L2 model. In addition, Brusaferri et al. (2020) have proposed Mixture of Expert (MoE) NN architecture that can represent the feature of hybrid systems in a NN structure for one-step-ahead identification of SARX models. MoE layer along with a gated recurrent units with softmax output plays a role as neural switching machine, while feed-forward NNs have been chosen as a structure to represent continuous dynamics. Moreover, a method to decompose an NN into a PWA model by using weight pruning to reduce the number of linear classifier has been discussed by Robinson (2021).

3.5 Research trends in methods

Recently, heuristic approaches combine different methods. For instance, Mejari et al. (2020b) have discussed a recursive manner of the least squares method with bias correction to deal with unknown noises based on the clustering method. However, this online method estimates the submodel parameters and the switching signal as well as unknown noise variance with high accuracy, but clustering leads to misclassification that should be dealt with. Wang et al. (2020) have also addressed weighted least squares method, but with improvement on accuracy of clustering by introducing gap metric to find similar measurement and minimize the number of the submodels. Moreover, a combination of cluster-based algorithm and self-training SVM algorithm has been proposed by Du et al. (2020). In this method, the clustering outputs initialize the modified SVM algorithm that reduces the computational complexity and increases the precision of partitioning. Furthermore, using algebraic procedure as a starting point and combing it with a polynomial optimization is another example of a heuristic combined method to address high dimensional dataset affected by large level of noise by Hojjatinia et al. (2020).

Generalization is the other important trend. Using a maximum-a-posteriori algorithm by Piga et al. (2020b) in a general way in terms of fitting SBJ models with time varying coefficients and equivalent state space models is a pioneer, and relaxation of user-dependent regularization of hyperparameters and the derivation of confidence intervals are for further extension and generalization. Similarly, the proposed method by Piga et al. (2020a) has been derived for PWA models, and it can be extended for polynomial nonlinear models in this trend. The priority of these papers is using a maximum-a-posteriori method that also obtains the distribution of the model parameters and the predicted output.

Moreover, VB methods have drawn attention since they also provide the distribution of the parameters and solve the optimization problem effectively. With less generality but taking missing measurements and delays into account for two simple model classes, i.e. SFIR and SARX, Liu et al. (2021) and Chen and Liu (2019) have discussed the application of VB methods, which should be extended for other types of hybrid systems to complete this chain of trends.

4. CONCLUSIONS

As discussed, hybrid system identification is an active research field, and it can be used for a wide range of real-world applications for modeling and control. In this survey a systematic review on models and methods has been proposed to show the state-of-the-art in hybrid system identification. Current and future research directions include the following objectives to:

- increase and generalize model complexity in terms of parameterization and find a method to solve its identification problem in a computationally efficient way,

Table 9. Other approaches.

Method category	Hybrid model	Algorithm	Reference(s)
Algebraic	SARX	Set membership	Ozay et al. (2015)
		Matrix rank minimization with partial matrix shrinkage	Konishi (2015)
		Geometric approach with stochastic hybrid decoupling polynomial	Nazari et al. (2016)
	SARX/SARMAX	Veronese-map-embedded GPCA with polynomial optimization	Hojjatinia et al. (2020)
	Wiener SARX	Kernel-based mapping with polynomial optimization	Zhang et al. (2016)
	Switched SS	SVD based on Hankel-like matrices	Sarkar et al. (2019)
OBE	SARX	Real-time set membership algorithm based on unknown-but-bounded noises	Goudjil et al. (2016)
	PWARX		Yahya et al. (2020)
	SOE		Goudjil et al. (2017b)
Continuous-time	PWA-SS	Concurrent learning	Kersting and Buss (2017)
		Integral concurrent learning	Du et al. (2021b)
	Switched SS	Integral concurrent learning	Du et al. (2021a)
	Continuous input-output	Conventional continuous-time identification	Goudjil et al. (2020)
	Switched SS	Conversion of discrete-time to continuous-time	Keshvari-Khor et al. (2018)
NN	Multiple NARMA-L2	One-step-ahead	Yang et al. (2017)
	SARX	Mixture of expert with softmax output	Brusaferri et al. (2020)
	PWA	Neural network decomposition	Robinson (2021)

- extend the probabilistic model parameterization for other classes of hybrid systems and find a solving algorithm for it,
- parameterize the switching domain in a nonlinear form,
- take practical problems such as delay and missing measurement into account for other types of hybrid systems,
- solve the mis-classification problem with innovative clustering approaches, and
- explore nonlinear hybrid system identification to increase accuracy, while decrease the number of modes for nonlinear processes with a wide range of operation.

In summary, this survey has highlighted that the methodological focus has slightly changed from the use of innovative optimization and novel clustering approaches towards problem generalization, multi-method combination, and the use of novel probabilistic methods. Moreover, other important issues related to hybrid system identification such as experiment design and identifiability can be reviewed in future research.

REFERENCES

Amaldi, E., Coniglio, S., and Taccari, L. (2016). Discrete optimization methods to fit piecewise affine models to data points. *Computers and Operations Research*, 75, 214–230.

Bako, L. and Yahya, O. (2019). Piecewise affine system identification: A least harmonic mean approach. In *2019 IEEE 58th Conference on Decision and Control*.

Barbosa, B.H., Aguirre, L.A., and Braga, A.P. (2019). Piecewise affine identification of a hydraulic pumping system using evolutionary computation. *IET Control Theory and Applications*, 13(9), 1394–1403.

Basiri, M.H., Thistle, J.G., and Fischmeister, S. (2018). A framework for inference and identification of hybrid-system models: Mixed event-/time-driven systems (METS). *IFAC-PapersOnLine*, 51(15), 287–292.

Bemporad, A., Breschi, V., Piga, D., and Boyd, S.P. (2018). Fitting jump models. *Automatica*, 96, 11–21.

Bianchi, F., Breschi, V., Piga, D., and Piroddi, L. (2021). Model structure selection for switched NARX system identification: A randomized approach. *Automatica*, 125, 109415.

Bianchi, F., Falsone, A., Piroddi, L., and Prandini, M. (2020a). An alternating optimization method for switched linear systems identification. *IFAC-PapersOnLine*, 53(2), 1071–1076.

Bianchi, F., Prandini, M., and Piroddi, L. (2020b). A randomized two-stage iterative method for switched nonlinear systems identification. *Nonlinear Analysis: Hybrid Systems*, 35, 100818.

Blazic, S. and Skrjanc, I. (2020). Hybrid system identification by incremental fuzzy c-regression clustering. In *2020 IEEE International Conference on Fuzzy Systems*.

Breschi, V., Bemporad, A., and Piga, D. (2016a). Identification of hybrid and linear parameter varying models via recursive piecewise affine regression and discrimination. In *2016 European Control Conference*.

Breschi, V. and Mejari, M. (2020). Shrinkage strategies for structure selection and identification of piecewise affine models. In *2020 59th IEEE Conference on Decision and Control*.

Breschi, V., Piga, D., and Bemporad, A. (2016b). Piecewise affine regression via recursive multiple least squares and multicategory discrimination. *Automatica*, 73, 155–162.

Breschi, V., Piga, D., and Bemporad, A. (2019). Maximum-a-posteriori estimation of jump Box-Jenkins models. In *2019 IEEE 58th Conference on Decision and Control*.

Breschi, V., Bemporad, A., Piga, D., and Boyd, S. (2018). Prediction error methods in learning jump ARMAX models. In *2018 IEEE Conference on Decision and Control*.

Brusaferri, A., Matteucci, M., Portolani, P., Spinelli, S., and Vitali, A. (2020). Hybrid system identification using a mixture of NARX experts with LASSO-based feature selection. In *2020 7th International Conference on Control, Decision and Information Technologies*.

Chen, X. and Liu, F. (2019). A variational Bayesian approach for identification of time-delay Markov jump autoregressive exogenous systems. *Circuits, Systems, and Signal Processing*, 39(3), 1265–1289.

Chen, X., Zhao, S., and Liu, F. (2017). Identification of time-delay Markov jump autoregressive exogenous systems with expectation-maximization algorithm. *International Journal of Adaptive Control and Signal Processing*, 31(12), 1920–1933.

Chen, X., Zhao, S., and Liu, F. (2020a). Identification of jump Markov autoregressive exogenous systems with missing measurements. *Journal of the Franklin Institute*, 357(6), 3498–3523.

Chen, X., Zhao, S., and Liu, F. (2020b). Online identification of time-delay jump Markov autoregressive exogenous systems with recursive expectation-maximization algorithm. *International Journal of Adaptive Control and Signal Processing*, 34(3), 407–426.

Du, Y., Liu, F., Qiu, J., and Buss, M. (2020). A semi-supervised learning approach for identification of piecewise affine systems. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 67(10), 3521–3532.

Du, Y., Liu, F., Qiu, J., and Buss, M. (2021a). A novel recursive approach for online identification of continuous-time switched nonlinear systems. *International Journal of Robust and Nonlinear Control*, 31(15), 7546–7565.

Du, Y., Liu, F., Qiu, J., and Buss, M. (2021b). Online identification of piecewise affine systems using integral concurrent learning. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 68(10), 4324–4336.

Du, Z., Balzano, L., and Ozay, N. (2018). A robust algorithm for online switched system identification. *IFAC-PapersOnLine*, 51(15), 293–298.

Fan, L., Kodamana, H., and Huang, B. (2017). Robust identification of switching Markov ARX models using EM algorithm. *IFAC-PapersOnLine*, 50(1), 9772–9777.

Garulli, A., Paoletti, S., and Vicino, A. (2012). A survey on switched and piecewise affine system identification. *IFAC Proceedings Volumes*, 45(16), 344–355.

Goudjil, A., Pouliquen, M., Pigeon, E., and Gehan, O. (2016). A real-time identification algorithm for switched linear systems with bounded noise. In *2016 European Control Conference*.

Goudjil, A., Pouliquen, M., Pigeon, E., and Gehan, O. (2017a). Identification algorithm for MIMO switched output error model in presence of bounded noise. In *2017 IEEE 56th Annual Conference on Decision and Control*.

Goudjil, A., Pouliquen, M., Pigeon, E., Gehan, O., and Bonargent, T. (2020). Continuous-time identification for a class of switched linear systems. In *2020 European Control Conference*.

Goudjil, A., Pouliquen, M., Pigeon, E., Gehan, O., and Targui, B. (2017b). Recursive output error identification algorithm for switched linear systems with bounded noise. *IFAC-PapersOnLine*, 50(1), 14112–14117.

Hartmann, A., Lemos, J.M., Costa, R.S., Xavier, J., and Vinga, S. (2015). Identification of switched ARX models via convex optimization and expectation maximization. *Journal of Process Control*, 28, 9–16.

Hojjatinia, S., Lagoa, C.M., and Dabbene, F. (2020). Identification of switched autoregressive exogenous systems from large noisy datasets. *International Journal of Robust and Nonlinear Control*, 30(15), 5777–5801.

Hu, Q., Fei, Q., Ma, H., Wu, Q., and Geng, Q. (2015). Identification for switched systems. *IFAC-PapersOnLine*, 48(28), 514–519.

Hure, N. and Vasak, M. (2017). Clustering-based identification of MIMO piecewise affine systems. In *2017 21st International Conference on Process Control*.

Jianwang, H. and Ramirez-Mendoza, R.A. (2020). Zonotope parameter identification for piecewise affine system. *Systems Science and Control Engineering*, 8(1), 232–240.

Kersting, S. and Buss, M. (2017). Recursive estimation in piecewise affine systems using parameter identifiers and concurrent learning. *International Journal of Control*, 92(6), 1264–1281.

Keshvari-Khor, H., Karimpour, A., and Pariz, N. (2018). Identification of continuous-time switched linear systems from low-rate sampled data. *IET Control Theory and Applications*, 12(14), 1964–1973.

Khanmirza, E., Nazarahari, M., and Mousavi, A. (2016). Identification of piecewise affine systems based on fuzzy PCA-guided robust clustering technique. *EURASIP Journal on Advances in Signal Processing*, 2016(1).

Konishi, K. (2015). Multiple low rank matrix approach to switched autoregressive exogenous system identification. In *2015 10th Asian Control Conference*.

Lauer, F. and Bloch, G. (2018). Hybrid system identification: Theory and algorithms for learning switching models, vol. 478. *Cham, Switzerland: Springer*.

Lauer, F. (2013). Estimating the probability of success of a simple algorithm for switched linear regression. *Nonlinear Analysis: Hybrid Systems*, 8, 31–47.

Li, L., Done, W., and Ji, Y. (2016). A subspace approach to the identification of MIMO piecewise linear systems. In *2016 35th Chinese Control Conference*.

Li, L. and Liu, J. (2017). Subspace clustering on parameter estimation of switched affine models. In *2017 36th Chinese Control Conference*.

Liu, J., Xu, Z., Zhao, J., and Shao, Z. (2022). Identification of piecewise affine model for batch processes based on constrained clustering technique. *Chemical Engineering Research and Design*, 181, 278–286.

Liu, X., Yang, X., and Yu, M. (2021). Identification of switched FIR systems with random missing outputs: A variational Bayesian approach. *Journal of the Franklin Institute*, 358(1), 1136–1151.

Lopes, R.V., Ishihara, J.Y., and Borges, G.A. (2016). Identification of state-space switched linear systems using clustering and hybrid filtering. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 39(2), 565–573.

Lu, Y., Huang, B., and Khatibisepehr, S. (2016). A variational Bayesian approach to robust identification of switched ARX models. *IEEE Transactions on Cybernetics*, 46(12), 3195–3208.

Ma, J., Huang, B., and Ding, F. (2019). Parameter estimation of Markov-switching hammerstein systems using the variational Bayesian approach. *IET Control Theory and Applications*, 13(11), 1646–1655.

Mattsson, P., Zachariah, D., and Stoica, P. (2016). Recursive identification method for piecewise ARX models: A sparse estimation approach. *IEEE Transactions on Signal Processing*, 64(19), 5082–5093.

Mazzoleni, M., Breschi, V., and Formentin, S. (2021). Piecewise nonlinear regression with data augmentation. *IFAC-PapersOnLine*, 54(7), 421–426.

Mejari, M., Breschi, V., Naik, V.V., and Piga, D. (2020a). A bias-correction approach for the identification of piecewise affine output-error models. *IFAC-PapersOnLine*, 53(2), 1096–1101.

Mejari, M., Breschi, V., and Piga, D. (2020b). Recursive bias-correction method for identification of piecewise affine output-error models. *IEEE Control Systems Letters*, 4(4), 970–975.

Mejari, M., Naik, V.V., Piga, D., and Bemporad, A. (2018). Regularized moving-horizon PWA regression for LPV system identification. *IFAC-PapersOnLine*, 51(15), 1092–1097.

Mejari, M., Naik, V.V., Piga, D., and Bemporad, A. (2020c). Identification of hybrid and linear parameter-varying models via piecewise affine regression using mixed integer programming. *International Journal of Robust and Nonlinear Control*, 30(15), 5802–5819.

Naik, V.V., Mejari, M., Piga, D., and Bemporad, A. (2017). Regularized moving-horizon piecewise affine regression using mixed-integer quadratic programming. In *2017 25th Mediterranean Conference on Control and Automation*.

Nazari, S., Rashidi, B., Zhao, Q., and Huang, B. (2016). An iterative algebraic geometric approach for identification of switched ARX models with noise. *Asian Journal of Control*, 18(5), 1655–1667.

Ozay, N., Lagoa, C., and Sznaier, M. (2015). Set membership identification of switched linear systems with known number of subsystems. *Automatica*, 51, 180–191.

Ozbay, B., Camps, O., and Sznaier, M. (2019). Efficient identification of error-in-variables switched systems via a sum-of-squares polynomial based subspace clustering method. In *2019 IEEE 58th Conference on Decision and Control*.

Paoletti, S., Savelli, I., Garulli, A., and Vicino, A. (2019). A bilevel programming framework for piecewise affine system identification. In *2019 IEEE 58th Conference on Decision and Control*.

Piga, D., Bemporad, A., and Benavoli, A. (2020a). Rao-blackwellized sampling for batch and recursive Bayesian inference of piecewise affine models. *Automatica*, 117, 109002.

Piga, D., Breschi, V., and Bemporad, A. (2020b). Estimation of jump box-jenkins models. *Automatica*, 120, 109126.

Pillonetto, G. (2016). A new kernel-based approach to hybrid system identification. *Automatica*, 70, 21–31.

Robinson, H. (2021). Approximate piecewise affine decomposition of neural networks. *IFAC-PapersOnLine*, 54(7), 541–546.

Rui, R., Ardeshiri, T., and Bazanella, A. (2016). Identification of piecewise affine state-space models via expectation maximization. In *2016 IEEE Conference on Computer Aided Control System Design*.

Sarkar, T., Rakhlin, A., and Dahleh, M. (2019). Nonparametric system identification of stochastic switched linear systems. In *2019 IEEE 58th Conference on Decision and Control*.

Scampicchio, A., Giaretta, A., and Pillonetto, G. (2018). Nonlinear hybrid systems identification using kernel-based techniques. *IFAC-PapersOnLine*, 51(15), 269–274.

Scampicchio, A. and Pillonetto, G. (2018). A new model selection approach to hybrid kernel-based estimation. In *2018 IEEE Conference on Decision and Control*.

Sefidmazgi, M.G., Kordmahalleh, M.M., Homaifar, A., and Karimoddini, A. (2015). Switched linear system identification based on bounded-switching clustering. In *2015 American Control Conference*.

Sefidmazgi, M.G., Kordmahalleh, M.M., Homaifar, A., Karimoddini, A., and Tunstel, E. (2016). A bounded switching approach for identification of switched MIMO systems. In *2016 IEEE International Conference on Systems, Man, and Cybernetics*.

Sellami, L., Zidi, S., and Abderrahim, K. (2016). Identification of switched linear systems using self-adaptive SVR algorithm. In *2016 24th Mediterranean Conference on Control and Automation*.

Shah, A.K. and Adhyaru, D.M. (2014). Parameter identification of PWARX models using fuzzy distance weighted least squares method. *Applied Soft Computing*, 25, 174–183.

Song, C., Wang, J., Ma, X., and Zhao, J. (2020). A PWA model identification method based on optimal operating region partition with the output-error minimization for nonlinear systems. *Journal of Process Control*, 88, 1–9.

Tang, X. and Dong, Y. (2020). Expectation maximization based sparse identification of cyberphysical system. *International Journal of Robust and Nonlinear Control*, 31(6), 2044–2060.

Wågberg, J., Lindsten, F., and Schön, T.B. (2015). Bayesian nonparametric identification of piecewise affine ARX systems. *IFAC-PapersOnLine*, 48(28), 709–714.

Wagner, M. and Kroll, A. (2014). A method to identify hybrid systems with mixed piecewise affine or nonlinear models of Takagi-Sugeno type. In *2014 European Control Conference*.

Wang, J., Song, C., Zhao, J., and Xu, Z. (2020). A PWA model identification method for nonlinear systems using hierarchical clustering based on the gap metric. *Computers and Chemical Engineering*, 138, 106838.

Wang, Z., An, H., and Luo, X. (2019). Switch detection and robust parameter estimation for slowly switched hammerstein systems. *Nonlinear Analysis: Hybrid Systems*, 32, 202–213.

Xiujun, C., Hongwei, W., Lin, W., and Zhengqing, X. (2020). Identification of switched nonlinear systems based on EM algorithm. In *2020 39th Chinese Control Conference*.

Xu, Z., Huang, Y., Zhao, J., Song, C., and Shao, Z. (2018). Time-partitioned piecewise affine output error model for batch processes. *Industrial and Engineering Chemistry Research*, 57(5), 1560–1568.

Yahya, O., Lassoued, Z., and Abderrahim, K. (2020). Identification of PWARX model based on outer bounding ellipsoid algorithm. In *2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering*.

Yang, Y., Xiang, C., Gao, S., and Lee, T.H. (2017). Data-driven identification and control of nonlinear systems using multiple NARMA-l2 models. *International Journal of Robust and Nonlinear Control*, 28(12), 3806–3833.

Yuan, Y., Tang, X., Zhou, W., Pan, W., Li, X., Zhang, H.T., Ding, H., and Goncalves, J. (2019). Data driven discovery of cyber physical systems. *Nature Communications*, 10(1).

Zhang, X., Cheng, Y., Wang, Y., Sznaier, M., and Camps, O. (2016). Identification of switched Wiener systems based on local embedding. In *2016 IEEE 55th Conference on Decision and Control*.

Zhang, X., Sznaier, M., and Camps, O. (2018). Efficient identification of error-in variables switched systems based on Riemannian distance-like functions. In *2018 IEEE Conference on Decision and Control*.