Proceedings of the 2004 IEEE
International Conference on Control Applications
Taipei, Taiwan, September 2-4, 2004

Comparison of three procedures for the identification of hybrid systems[1]

J.H.G. Niessen[2] A. Lj. Juloski[2,5] G.Ferrari-Trecate[4] W.P.M.H. Heemels[2,3]

Abstract— **In this paper three recently proposed procedures for the identification of Piece-Wise AutoRegressive eXogenous (PWARX) models are compared. Quantitative measures for the quality of the obtained models are proposed. Using one dimensional examples specific behaviors of the methods are pointed out. An experimental example is considered as well.**

I. Introduction

In this paper we study three recently proposed procedures for the black-box identification of discrete time hybrid systems in piecewise affine (PWA) form. More precisely, the studied procedures identify piecewise ARX (PWARX) models, which are a hybrid generalization of the classical ARX models. PWARX models are obtained when the regressor space is partitioned into a finite number of convex polyhedral regions and a single ARX model is valid over each region.

We compare the clustering-based procedure [1], the greedy procedure [2] and the algebraic procedure [3], [4]. These identification algorithms are briefly summarized in section III. A formal analysis of the properties of the greedy and algebraic procedures is not available. Some features of the clustering-based procedure have been analyzed theoretically in [5]. However, the study of specific cases can help to better understand properties of the methods in practical situations. In order to compare the procedures and asses the quality of obtained models we propose several quantitative measures in section III.

II. Problem statement

All the procedures considered identify piece-wise ARX (PWARX) models of the form:

$$y(k) = f(x(k)) + e(k), \tag{1}$$

where $e(k)$ is the noise term and the piece-wise affine (PWA) map $f(\cdot)$ is defined as:

$$f(x) = \begin{cases} [x' \quad 1]\,\theta_1 & \text{if } x \in \mathcal{X}_1, \\ \quad\vdots & \\ [x' \quad 1]\,\theta_s & \text{if } x \in \mathcal{X}_s. \end{cases} \tag{2}$$

In (2) $x(k)$ is a vector of regressors defined as

$$x(k) \triangleq \begin{array}{l}[y(k-1)\,y(k-2)\,\ldots\,y(k-n_a) \\ \quad u(k-1)\,u(k-2)\,\ldots\,u(k-n_b)]',\end{array} \tag{3}$$

where k is the time index and $y, u \in \mathbb{R}$ are the outputs and the inputs of the system, respectively. For $i = 1, \ldots s$, $\theta_i \in \mathbb{R}^{n+1}$ is a parameter vector (PV) with $n = n_a + n_b$.

[1]This work is partially supported by STW/PROGRESS grant ESS.5173

[2]Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, E-mail: {j.h.g.niessen, a.juloski, w.p.m.h.heemels}@tue.nl

[3] also with Embedded Systems Institute, PO Box 513, 5600MB Eindhoven, The Netherlands

[4]INRIA, Domaine de Voluceau, Rocquencourt - B.P.105, 78153, Le Chesnay Cedex, France, E-mail: Giancarlo.Ferrari-Trecate@inria.fr

[5]Corresponding author, tel: +31.40.2473142

The bounded regressor space $\mathbb{X}$ is partitioned in s convex polyhedral regions $\{\mathcal{X}_i\}_{i=1}^s$, i.e.

$$\bigcup_{i=1}^{s} \mathcal{X}_i = \mathbb{X} \subset \mathbb{R}^n \quad \text{and} \quad \mathcal{X}_i \cap \mathcal{X}_j = \varnothing, \tag{4}$$

with $i \neq j$.

The general identification problem reads as follows: given the dataset $\mathcal{N} = \{(x(k), y(k))\}_{k=1}^N$ reconstruct the PWA map $f(\cdot)$, i.e. determine the PVs $\{\theta_i\}_{i=1}^s$ and the polyhedral partitions $\{\mathcal{X}_i\}_{i=1}^s$. Apart from the identification data each procedure requires some tuning parameters which will define the structure of the identified models. For all the procedures, the orders of ARX models n_a, n_b need to be provided beforehand. The number of submodels s should be given, but some methods determine it automatically.

Identification of PWARX-models is a challenging problem since it involves the estimation of both the PVs $\{\theta_i\}_{i=1}^s$ and the regions of the regressor space $\{\mathcal{X}_i\}_{i=1}^s$ on the basis of the available dataset $\mathcal{N}$. In case that regions of the regressor space are known a priori the problem complexity reduces to that of a linear system identification problem, since the datapoints can be classified to corresponding data subsets $\{\mathcal{F}_i\}_{i=1}^s$, and a linear system identification method can be applied to each of these subsets of $\mathcal{N}$ [1].

III. The compared procedures

In this section we briefly discuss the clustering-based procedure [1], the greedy procedure [2] and the algebraic procedure [3], [4]. The basic steps that each method performs are: estimation of the PVs $\{\theta_i\}_{i=1}^s$, classification of the datapoints and estimation of the corresponding regions $\{\mathcal{X}_i\}_{i=1}^s$.

The first two steps are performed in a different way by each procedure, as discussed in the sequel, while the estimation of the regions can be done in the same way for all methods. The basic idea is as follows. Having the data points that are attributed to sets $\mathcal{F}_i$ and $\mathcal{F}_j$, we are looking for a separating hyperplane in the regressor space $\mathbb{X}$ described by:

$$M_{ij}' x = m_{ij}, \tag{5}$$

where M_{ij} is a vector, and m_{ij} is a scalar, so that for each $x(k) \in \mathcal{X}_i$, $M_{ij}' x(k) \leq m_{ij}$, and for each $x(k) \in \mathcal{X}_j$ $M_{ij}' x(k) > m_{ij}$. If such a hyperplane can not be found (i.e. the dataset is not linearly separable) we are interested in a generalized separating hyperplane which minimizes the number of misclassified datapoints. The method we use for estimating the separating hyperplanes in this paper is Multicategory Robust Linear Programming (MRLP). This method can solve the classification problem with more than two data classes. For a detailed discussion on MRLP see [6].

A. Clustering-based procedure

The clustering-based procedure [1] is based on the rationale that data points close to each other in regressor space are likely to belong to the same partition and the same ARX model. The main steps of the procedure are:

- For each data pair $(x(k), y(k))$ a local dataset (LD) $\mathcal{C}_k$ is built containing its $c-1$ nearest datapoints[1] in the regressor space $\mathbb{X}$. LDs that only contain data points belonging to a single subsystem are referred to as *pure* LDs, while LDs containing data generated by different subsystems are called *mixed* LDs.
- Calculate θ_k^{LS} for each LD using least squares on $\mathcal{C}_k$ and compute the mean m_k of $\mathcal{C}_k$. Each datapoint $(x(k), y(k))$ is thereby mapped onto the feature vectors $\xi_k = [(\theta_k^{LS})', m_k']'$.
- Cluster points $\{\xi_k\}_{k=1}^N$ in s clusters $\mathcal{D}_i$ by minimizing a suitable cost function
- Since the mapping of the datapoints onto the feature space is bijective, the data subsets $\{\mathcal{F}_i\}_{i=1}^s$ can be built using the clusters $\{\mathcal{D}_i\}_{i=1}^s$. The PVs $\{\theta_i\}_{i=1}^s$ are estimated from data subsets $\mathcal{F}_i$ by least squares.

The parameters s and c are the tuning knobs of this procedure.

B. Greedy procedure

In the greedy procedure [2] a bound $\delta > 0$ on the prediction error is selected. Each data pair $(x(k), y(k))$ should satisfy

$$| y(k) - \varphi'(k)\theta_i | \leq \delta, \quad \forall k = 1, \ldots, N, \tag{6}$$

for some θ_i, where $\varphi(k) = [x'(k) \quad 1]'$. The system of inequalities (6) is in general infeasible for a single PV θ, hence, it should be split in the minimal number of feasible subsystems $\bar{s}$, and corresponding PVs $\{\theta_i\}_{i=1}^{\bar{s}}$ should be determined. Unfortunately, the problem of splitting (6) into a minimal number of feasible systems is NP-hard. The main steps of the procedure are:

- A suboptimal version of a MIN-PFS algorithm proposed in [7] is used to partition the infeasible system (6) into $\bar{s}$ feasible subsystems.
- A refinement algorithm is repeatedly applied to the previously found set of parameter vectors $\{\theta_i\}_{i=1}^{\bar{s}}$ and data subsets $\{\mathcal{F}_i\}_{i=1}^{\bar{s}}$. Submodels are merged if their parameter vectors are 'similar'. The similarity measure between θ_i and θ_j is defined as,

$$\alpha_{i^*,j^*} = \min_{1 \leq i < j \leq \hat{s}} \|\hat{\theta}_i^{(t)} - \hat{\theta}_j^{(t)}\| / \min\{\|\hat{\theta}_i^{(t)}\|, \|\hat{\theta}_j^{(t)}\|\}, \tag{7}$$

  and they are considered similar if $\alpha_{i^*,j^*} < \alpha$. The subset $\mathcal{F}_i$ will be discarded in case the cardinality of subset $\mathcal{F}_i$ with respect to the cardinality of $\mathcal{N}$ is less then the bound β. The refinement steps are repeated until the parameter updates between iteration t and $t+1$ becomes negligible, i.e. until the termination condition:

$$\|\hat{\theta}_i^{(t+1)} - \hat{\theta}_i^{(t)}\| / \|\hat{\theta}_i^{(t)}\| \leq \gamma, \quad \forall_i = 1, \ldots, \hat{s}, \tag{8}$$

  is satisfied.

Datapoints that do not satisfy (6) for any of the estimated parameter vectors are not classified and will be marked as *infeasible*. Datapoints satisfying (6) for more than one estimated parameter vector will be marked as *undecidable*.

The tuning parameters of this procedure are α, β, γ and δ. These parameters implicitly determine the estimated number of subsystems $\bar{s}$, returned by the procedure for a given dataset $\mathcal{N}$.

[1] according to the Euclidean distance

C. Algebraic procedure

The method proposed in [3], [4] approaches the problem of identifying a PWARX model as an algebraic geometric problem. In the noiseless case ($e = 0$) the data pair $(x(k), y(k))$ satisfies the equality

$$y(k) - \varphi'(k)\theta_i = 0 \tag{9}$$

for a suitably chosen PV θ_i. Hence, the equality

$$\prod_{i=1}^{s} (y(k) - \varphi'(k)\theta_i) = 0, \tag{10}$$

always holds.

In [3] it is shown that in the noiseless case the number of subsystems s can be determined as a rank of a suitably constructed matrix L_s. In the noisy case the rank of the matrix L_s is considered to be r if $\sigma_{r+1}/\sigma_r < \varepsilon$, where σ_i is the i-th singular value of L_s.

The algorithm to compute PVs $\{\theta_i\}_{i=1}^s$ from the system of equations (10) is described in the paper [4]. This procedure involves finding the roots of the second order polynomial, that in the noisy case may become complex-valued. Hence, the procedure may terminate without yielding the model. Data pairs $(x(k), y(k))$ are attributed to the submodel λ satisfying the rule:

$$\lambda(k) = \arg \min_{1 \leq i \leq \hat{s}} (y(k) - \varphi(k)'\theta_i)^2. \tag{11}$$

The only tuning parameter of the procedure is ε.

D. Quality measures

Since our aim is to compare the procedures, some quantitative measures for the quality of the identification results are introduced. These measures will capture the accuracy of the estimated PVs $\{\hat{\theta}_i\}_{i=1}^s$ and the accuracy of the estimated partitions $\{\hat{\mathcal{X}}_i\}_{i=1}^s$.

When the model that generated the data is known, one can measure the accuracy of the identified PV through the quantity:

$$\Delta_\theta = \max_{1 \leq i \leq s} \left(\min_{1 \leq j \leq s} \frac{\|\hat{\theta}_i - \theta_j\|_2}{\|\theta_j\|_2} \right), \tag{12}$$

where $\hat{\theta}_i$ are the reconstructed PVs and θ_j are the PVs of the generating model. This measure is only applicable for the cases where the number of submodels is the same for the generating and identified model. Δ_θ is zero for the perfect estimates, and increases as the estimates worsen.

A sensible quality measure for the estimated regions is much harder to define. For the case where $n = 1$ and $s = 2$ we propose the following index:

$$\Delta_\mathcal{X} = \left| \frac{m_{12}}{M_{12}} - \frac{\hat{m}_{12}}{\hat{M}_{12}} \right|. \tag{13}$$

where M_{12}, m_{12}, $\hat{M}_{12}$, $\hat{m}_{12}$ are the coefficients of the separating hyperplanes, defined in (5), of the original and reconstructed model, respectively.

An overall quality measure which is also applicable when the generating model is not known is provided by the sum of squared residuals (one step ahead prediction errors):

$$\hat{\sigma}_e^2 = \frac{1}{s} \sum_{i=1}^{s} \frac{\mathrm{SSR}_{\mathcal{F}_i}}{|\mathcal{F}_i|}, \tag{14}$$

where the set $\mathcal{F}_i$ contains the datapoints classified to submodel i and the sum of squared residuals (SSR) of submodel i is defined as:

$$\mathrm{SSR}_{\mathcal{F}_i} = \sum_{x(k)\in\mathcal{F}_i} (y(k) - [x(k)'\ 1]\theta_i)^2.$$

The value of the estimated model is considered acceptable if $\hat{\sigma}_\varepsilon^2$ is small and/or near the expected noise of the identified system.

Models with good one-step ahead prediction properties may perform poorly in simulation. To measure the model performance in simulation we propose to use the averaged Sum of the Squared simulation Errors ($\mathrm{SSE}_{\mathrm{sim}}$),

$$\mathrm{SSE}_{\mathrm{sim}} = \frac{1}{N-n} \sum_{k=n+1}^{N} (y(k) - \hat{y}(k))^2, \qquad (15)$$

where $\hat{y}(k)$ is the output of the simulation obtained by building $x(k)$ from the real inputs and previously estimated outputs. The idea behind (15) is that poorly estimated regions may increase the simulation error, since these poor estimates may lead to wrong choices of the next submodel.

When doing experimental identification $\hat{\sigma}_\varepsilon^2$ and $\mathrm{SSE}_{\mathrm{sim}}$ are useful for selecting acceptable models from a set of identified models obtained by using the procedures with different tuning parameters and estimates of the system orders.

IV. Intersecting Hyperplanes

In the previous section we have highlighted that the procedures use different approaches for parameter estimation and classification of the datapoints. The clustering-based procedure assumes that certain geometrical properties are present in the data set, while the greedy and algebraic procedure do not use this assumption. This may lead to wrong estimates of the separating hyperplanes. Namely, if the hyperplanes over the regressor space defined by PVs θ_i and θ_j intersect over $\mathcal{X}_j$, datapoints may be wrongly attributed to the data subset $\mathcal{F}_i$. In order to illustrate this problem we consider an example where the PVs of the real system virtually intersect over the regressor space $\mathbb{X}$. Consider the PWARX model $y(k) = f(x(k)) + e(k)$ where f is defined as:

$$f(x) = \begin{cases} [x \quad 1]\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} & \text{if } x \in [-2.5,\ 0] \\ [x \quad 1]\begin{bmatrix} -1 \\ 2 \end{bmatrix} & \text{if } x \in (0,\ 2.5] \end{cases} \qquad (16)$$

We generated regressors whose 80% is uniformly distributed over the regressor space $\mathbb{X} = [-2.5,\ 2.5]$ and the remaining 20% over $[0.85,\ 1.15]$, so that the virtual intersection is excited thoroughly. A normally distributed noise e with zero mean and variance $\sigma_e^2 = 0.005$ is added to $f(x(k))$. The results are plotted in the figures 1 and 2.

From figure 1 it is seen that the clustering-based procedure, as expected, does not experience problems with the intersecting PVs. Both the PVs and separating point are estimated accurately.

The greedy procedure is used with an error bound of $\delta = 3\sigma_e$, where σ_e is the standard deviation of the noise e. The result of the greedy procedure depicted in figure 1 shows that the datapoints within the error bound δ for both estimated PVs are not classified since they are considered undecidable. Figure 2 shows the result for the greedy

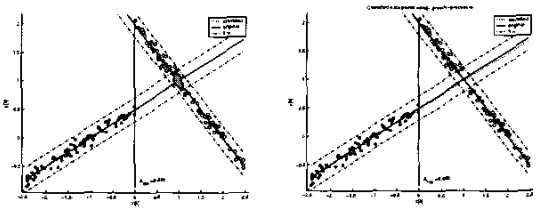

Fig. 1. **left**: Clustering-based procedure yielding $\Delta_\theta = 0.0045$ and $\Delta_\mathcal{X} = 0.000$ **right**: Greedy procedure (not classifying undecidable datapoints) yielding $\Delta_\theta = 0.0334$ and $\Delta_\mathcal{X} = 0.000$

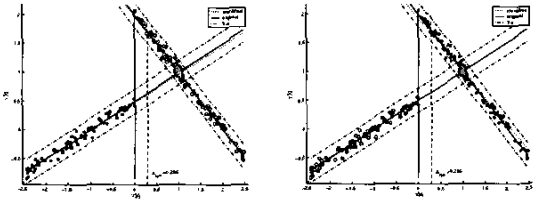

Fig. 2. **left:** Greedy procedure (classifying datapoints to subsystems which generates the smallest prediction error) yielding $\Delta_\theta = 0.0334$ and $\Delta_\mathcal{X} = 0.285$ **right:** Algebraic procedure (classifying datapoints to subsystems which generates the smallest prediction error) yielding $\Delta_\theta = 0.0453$ and $\Delta_\mathcal{X} = 0.280$

procedure when classifying the datapoints to the subsystem generating the smallest prediction error, as in the algebraic procedure. It is seen that classification errors influence the estimation of the separating point. Figure 2 shows that the algebraic procedure experiences problems for estimating the separating point due to classification errors.

This example illustrates data classification problems that may occur when the hyperplanes parameterized by PVs θ_i in the PWARX model intersect, and the data is classified without taking geometric properties into account. By neglecting 'undecidable' data points the greedy procedure overcomes the problem in this example, but it is not clear if this is true in general. The algebraic procedure, as well as the greedy procedure, is not able to provide the correct estimate of the model, if the classification rule (11) is used.

V. Overestimation of Model Orders

All the considered identification algorithms assume the system orders n_a and n_b to be known exactly, but in practice this is seldom the case. In order to investigate the effects of overestimating the model orders we consider a one-dimensional autonomous model of the form $y(k) = f(x(k)) + e(k)$ with orders $n_a = 1$ and $n_b = 0$ where the PWA map $f(\cdot)$ is defined as:

$$f(x) = \begin{cases} [x \quad 1]\begin{bmatrix} 2 \\ 10 \end{bmatrix} & \text{if } x \in [-10,\ 0], \\ [x \quad 1]\begin{bmatrix} -1.5 \\ 10 \end{bmatrix} & \text{if } x \in (0,\ 10], \end{cases} \qquad (17)$$

and the noise e is normally distributed, with zero mean and variance $\sigma_e^2 = 0.01$. To generate the dataset for identification values of x are obtained by uniform sampling in the interval $[-10, 10]$. The dataset is depicted in figure 3.

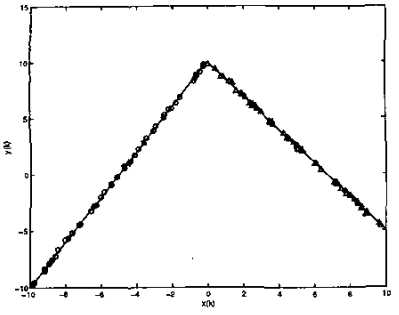

Fig. 3. Dataset for the one-dimensional model

When $x(k) = y(k-1)$ the time series generated by (17) has the following form[2]:

$$y(0), f(y(0)), f^2(y(0)), \dots f^k(y(0)), \dots \qquad (18)$$

where $f^k(\cdot)$ denotes k-times iterated function f over it's argument. Hence, for $\tilde{n}_a = 1, \tilde{n}_b = 0$ the data pairs have the form $(x, f(x))$, for $\tilde{n}_a = 2, \tilde{n}_b$=0 data pairs take the form $((f(x), x), f^2(x))$, and for some $\tilde{n}_a$, $\tilde{n}_b = 0$ data pair has the form $((f^{\tilde{n}_a-1}, \dots x), f^{\tilde{n}_a})$. Since the model (17) does not depend on input u, when $\tilde{n}_b > 0$ the regressor part that depends on the input was filled with random values uniformly distributed in the interval $[-10, 10]$. Identification procedures were subsequently applied to regressors formed in this way. Note that for overestimated model orders, the correct model is obtained by setting to zeroes the entries in θ_i, M_{ij}, m_{ij} on positions corresponding to superfluous elements in the regressor.

We did not manage to identify the model using algebraic procedure when the model orders were overestimated, because the procedure returned complex numbers for parameter values. Figure 4 shows the values of the criterion $\hat{\sigma}_\varepsilon^2$ on the logarithmic scale, for models with different model orders identified by the clustering-based and the greedy procedure, respectively.

From figure 4 it is seen that the clustering procedure identifies the model with $\hat{\sigma}_\varepsilon^2$ value close to the noise in the system for true system orders, but that the performance rapidly deteriorates when the model order is overestimated. The problem with the overestimated order lies in the assumption that datapoints close to each other in the regressor space belong to the same subsystem. When overestimating the order of the model regressor is extended with elements which do not contain relevant information for the estimation of the subsystems, but change the distance between the regressors. If the true distance is denoted by d_0, the distance between the extended regressors is $d_e^2 = d_0^2 + d_*^2$, where d_*^2 is due to the added elements, and contains no useful information. Depending on the true and overestimated model orders d_* can easily be of the same or higher order of magnitude as d_0.

In figure 4 we see that the greedy procedure generates σ_ε^2's which remains approximately constant. This can be explained by the pre-selected error bound δ. Hence, the value of σ_ε^2 depends on the chosen value for δ. However, overestimation of the model orders has other consequences for the procedure. In figure 4 we see that the number of

[2]modulo the error term e

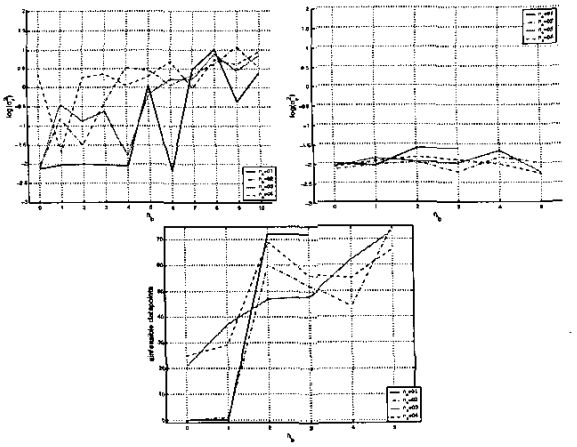

Fig. 4. **upper left:** $\hat{\sigma}_\varepsilon^2$ for the clustering procedure with $s = 2$ and $c = 20$ **upper right:** $\hat{\sigma}_\varepsilon^2$ for the greedy procedure with $\delta = 0.3$, $\alpha = 0.15$, $\beta = 0.10$ and $\gamma = 1 \cdot 10^{-5}$ **lower:** number of infeasible datapoints for the greedy procedure with $\delta = 0.3$, $\alpha = 0.15$, $\beta = 0.10$ and $\gamma = 1 \cdot 10^{-5}$

points which are marked as infeasible by the procedure grows rapidly when overestimating model orders. The first step of the procedure (suboptimal partitioning of the infeasible system of inequalities (6)) results in the large number of subsystems with small cardinality, which are then rejected by the refinement algorithm, and permanently marked as infeasible. It is hard to analytically explain the effects that occur when overestimating the system order in each of the procedures. However, when using these algorithms for systems with unknown orders, the observed behaviors may provide useful guidelines. In fact, by running the procedures for several order estimates and looking at values of σ_ε^2 for the clustering procedure and at the number of infeasible datapoints for the greedy procedure can help in choosing the acceptable estimate of the system orders.

VI. Effects of noise

In this section we study effects of noise e on the identification procedures. The first issue of interest is the effect that different realizations of noise with the same statistical properties have on the identification results. The second issue is how statistical properties of noise influence identification results.

To shed some light on these issues we designed an experiment with the PWARX model of section V (see (17)). For this model we generated a noiseless dataset of 100 datapoints. The procedures are applied 100 times on this dataset, after adding a different realization of normally distributed noise with zero mean and variance σ_ε^2 to the outputs $y(k)$. For each identified model the index Δ_θ is computed. In this way an approximate distribution of Δ_θ for each σ_ε^2 can be constructed. One such distribution obtained for the clustering-based procedure in given in figure 5. For each such distribution we computed its mean and variance, and plotted them in figure 6 as functions of σ_ε^2 for all three procedures.

From figure 6 we can conclude that the clustering procedure is the most robust with respect to noise. This can be attributed to the number of steps that are taken to make the procedure more robust as discussed in [1]. The greedy procedure is somewhat less robust to noise than

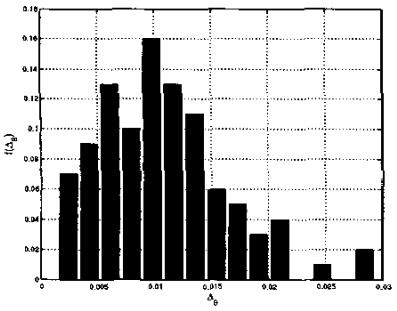

Fig. 5. Approximate distribution of Δ_θ over 100 runs using the clustering-based procedure with $c = 10$ and different realizations of noise with $\sigma_\eta^2 = 0.075$.

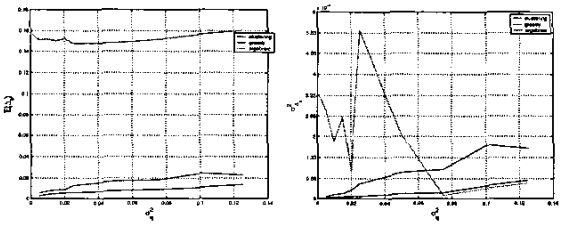

Fig. 6. Means (left) and variances (right) of the Δ_θ distributions for several variances of noise σ_η^2

the clustering procedure. From the plot of variances we see that the results of the greedy procedure show higher dependance on the realization of noise, in comparison to the clustering procedure. From the figures we see that the algebraic procedure performs poorly, in comparison to the other two procedures. However, the results are also very sensitive to the specific realization of noise.

VII. Experimental example

In this section we show the results of the identification of the component placement process in pick-and-place machines. A detailed description of the process and the experimental setup can be found in [8], [9].

The pick-and-place machine is used for automatically placing electronic components on a Printed Circuit Board (PCB). To study the placement process, an experimental setup was made, for which a scheme is shown in figure 7.

The setup consists of the mounting head, from an actual pick-and-place machine, which is fixed above the impacting surface. Mounting head contains a vacuum pipette that carries the component (mass M), which is connected to the casing via spring c_1, and moved by the electrical motor (force $\vec{F}$). Friction effects (linear friction d_1, dry friction f_1) are present. The chosen design of the impacting surface simulates the elasticity and damping properties of the PCB (spring c_2, damper d_2), but some parasitic effects are also present (dry friction f_2).

The control variable is the input voltage to the motor u, and the measured variable is the position of the pipette y. This experimental setup has two main modes of operation: *free mode*, when the component is not in contact with the impacting surface, and *impact mode*, when they are

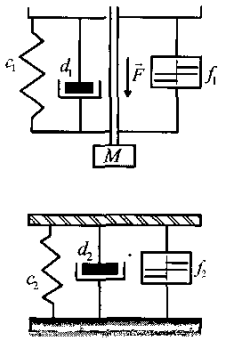

Fig. 7. Schematic representation of the experimental setup

in contact. Note that the sensor of contact between the component and the impacting surface is not available.

A dataset consisting of 750 samples is collected. The dataset is divided into two overlapping sets of 500 points, the first set is used for identification, and the second for validation. All three procedures where applied for several order estimates and with different tuning parameters. The procedures were executed for all the combinations of these orders and tuning parameters. The proposed quality measures σ_ε^2 and SSE_{sim} were used to choose acceptable identified models for which the simulations were plotted. The best identified model was then chosen by visual inspection.

For the clustering-based procedure figure 8 shows the simulation based on the validation dataset for the best model obtained. In the upper panel of the figure measured output y_{id} and the simulated output y_{sim} are depicted. The input signal u is plotted in the middle panel and the lower panel shows which of the identified subsystems is active at each time instant. It turns out that the best models are obtained for high values of c. The same was observed in [8]. A possible explanation is the following: because of the presence of dry friction neither the free nor the impact mode are linear, but with large LD's the effects of dry friction can be 'averaged out' as a process noise. Note that the difference between the measured and simulated responses, which is due to unmodeled dry friction, is clearly visible, e.g. on the time interval $[225, 300]$.

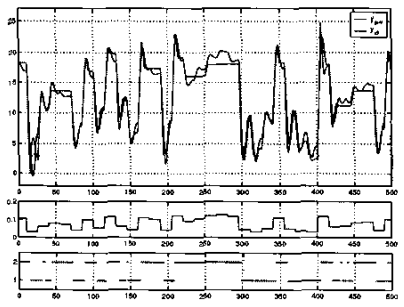

Fig. 8. Simulation of the PWARX model generated by the clustering procedure with $n_a = 2$, $n_b = 2$, $s = 2$ and $c = 90$ for the validation data set with $\text{SSE}_{sim} = 1.98$

As the number of modes s for the greedy procedure is not fixed, in order to identify two models, the right combination of the parameters α, γ and δ has to be found.

For the initial error bound δ we used $3\hat{\sigma}_\varepsilon \approx 1$, obtained from the clustering-based procedure, assuming that this value would be a good estimate for the variance of the measurement noise. Executing the greedy procedure with δ's in the vicinity of this $3\hat{\sigma}_\varepsilon$ resulted in identified models with only one parameter vector, and a large number of infeasible points. Therefore, we had to lower the error bound to $\delta = 0.25$. For this value of δ the procedure identified a model that distinguishes two subsystems. Model identified with this δ had a smaller value of $\hat{\sigma}_\varepsilon^2$ for the identification dataset than the model identified with the clustering-based procedure. However, when looking at the values of SSE_{sim} for the identification dataset and the validation dataset the model identified with the clustering-based procedure performs better. This observation can be attributed to the fact that the greedy procedure is less robust to noise and therefore the estimated parameter vectors are sensitive to the noise realization. The simulation of the validation dataset for the best identified model is shown in the figure 9.

Due to the severe noise and nonlinearities (e.g. dry friction) the algebraic procedure returned complex numbers as the estimated parameter values.

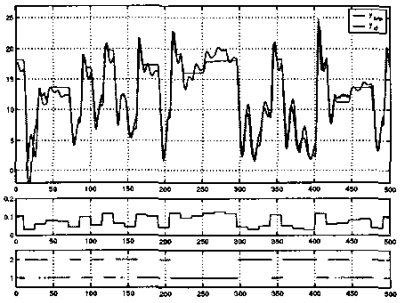

Fig. 9. Simulation of the PWARX model generated by the greedy procedure with $n_a = 2$, $n_b = 2$, $\delta = 0.25$, $\alpha = 0.15$, $\beta = 0.05$ and $\gamma = 5 \cdot 10^{-6}$ for the validation data set with $\text{SSE}_{sim} = 2.15$

VIII. Conclusions and Recommendations

In this paper we discussed three procedures for the identification of PWARX models. The clustering-based procedure, proposed in [1], reduces the problem of classifying the data to an optimal clustering problem. The greedy procedure, proposed in [2], approaches the problem as that of partitioning an infeasible system of linear inequalities into a minimum number of feasible subsystems. The algebraic procedure, proposed in [3], [4], treats the problem in an algebraic geometric fashion.

By applying the algorithms on suitably constructed one-dimensional problems we have emphasized some of the behaviors that are intrinsic to the procedures. In section IV we have shown that approaches that neglect geometrical properties present in the data set (such as greedy and algebraic approach) may experience data classification errors which may cause problems in the estimation of the regions. The greedy procedure tries to solve this problem by neglecting certain datapoints - this helped in the considered example, but it is not clear whether this is true in general. The straightforward classification of datapoints as proposed in [3] suffers the most from classification errors. A possible way to overcome this problem may lie in the combination of the approaches provided by all three procedures: e.g. the undecidable data points may be classified to the same mode as the closest feasible data point. This issue requires further research.

In section V we generated the identification data with an 1-dimensional PWARX system, and then we tried to identify models with overestimated orders. All of the procedures experienced problems with this example. The clustering procedure is not able to identify proper models since the distance information becomes corrupted when adding non-relevant information to the regressors. The greedy procedure has troubles estimating subsystems for which a sufficient number of datapoints can be classified. We did not manage to identify a model with the algebraic procedure when overestimating the orders.

In section VI we illustrated the effects of noise are on the accuracy of the estimated parameter vectors. We studied how sensitive the obtained results are to the specific realizations of noise with the same variance, as well as the effects of varying the variance. In this respect the clustering-based procedure is the procedure that was the most robust to noise. The method produced the most accurate estimates and was the least sensitive w.r.t. the realization of noise. The greedy procedure was less accurate and less consistent in estimating parameter vectors. Our experiments show that the algebraic procedure is very sensitive to noise. We stress that the observed behaviors are specific for the studied examples, and that for conclusive statements the analysis for the more general cases is needed.

In section VII we applied the procedures on a dataset obtained from an experimental setup of a pick-and-place machine discussed in [8]. Using physical insights and trial-and-error to tune the procedures, the models of the process were successfully identified with clustering-based and greedy procedures. However,we found that all three procedures lack clear guidelines on selecting tuning parameters.

References

[1] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine and hybrid systems," *Automatica*, vol. 39, no. 2, pp. 205–217, Feb 2003.

[2] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, "A greedy approach to identification of piecewise affine models," in *Proc. 6th Int. Workshop on Hybrid Systems: Computation and Control*, O. Maler and A. Pnueli, Eds., vol. 2623, Prague, Czech. Springer-Verlag, Berlin Heidelberg 2003, 3-5 April 2003, pp. 97–112.

[3] R. Vidal, S. Soatto, and S. Sastry, "An algebraic geometric approach for identification of linear hybrid systems," in *42nd IEEE Conference on Decision and Control*, 2003.

[4] R. Vidal, Y.Ma, and S.Sastry, "Generalized principal component analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[5] G. Ferrari-Trecate and M. Schinkel, "Conditions of optimal classification for piecewise affine regression," in *Proc. 6th International Workshop on Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, O. Maler and A. Pnueli, Eds. Springer-Verlag, 2003, vol. 2623, pp. 188–202.

[6] K. Bennett and O. Mangasarian, "Multicategory discrimination via linear programming," *Optimization Methods and Software*, vol. 3, pp. 27–39, 1993.

[7] E. Amaldi and M. Mattavelli, "The min pfs problem and piecewise linear model estimation," *Discrete Applied Mathematics*, vol. 118, pp. 115–143, April 2002.

[8] A. Juloski, W. Heemels, and G. Ferrari-Trecate, "Identification of an industrial hybrid system," in *IFAC Conference on the Analysis and Design of Hybrid Systems (ADHS 03)*, S. Engell, H. Gueguen, and J. Zaytoon, Eds. Saint-Malo, France, 16-18 june 2003.

[9] ——, "Data-based hybrid modelling of the component placement process in pick-and-place machines," *Control Engineering Practice*, to appear.