Data Mining of Historic Data for Process Identification ⋆

Daniel Peretzki [a], Alf J. Isaksson [a,b], André Carvalho Bittencourt [a], Krister Forsman [c]

[a] *Linköping University, Department of Electrical Engineering, SE-581 83 Linköping, Sweden*

[b] *ABB AB, Corporate Research, SE-721 78 Västerås, Sweden*

[c] *Perstorp AB, SE-284 80 Perstorp, Sweden*

Abstract

Performing experiments for system identification is often a time-consuming task which may also interfere with the process operation. With memory prices going down, it is more and more common that years of process data are stored (without compression) in a history database. The rationale for this work is that in such stored data there must already be intervals informative enough for system identification. Therefore, the goal of this project was to find an algorithm that searches and marks intervals suitable for process identification (rather than performing completely automatic system identification). For each loop, 4 stored variables are required; setpoint, manipulated variable, process output and mode of the controller.

The proposed method requires a minimum of knowledge of the process and is implemented in a simple and efficient recursive algorithm. The essential features of the method are the search for excitation of the input and output, followed by the estimation of a Laguerre model combined with a chi-square test to check that at least one estimated parameter is statistically significant. The use of Laguerre models is crucial to handle processes with deadtime without explicit delay estimation. The method was tested on three years of data from more than 200 control loops. It was able to find all intervals in which known identification experiments were performed as well as many other useful intervals in closed/open loop operation.

Key words: Data Mining, Data Segmentation, System Identification, Excitation, Condition numbers, Laguerre filters

1 INTRODUCTION

In the process industry, models are relevant for different purposes, such as optimizing production, improving control performance and supervision. In many occasions the task of building a process model is complemented with an identification procedure, where the model parameters are identified from measured data. Performing dedicated experiments for system identification is often a time-consuming task which may also interfere with the process operation. Nowadays, it is common to store measurement data from the plant operation (without compression) in a history database. Such data is a very useful source of information about the plant, and might contain suitable data to perform process identification. Due to the size of such databases, searching for data intervals suitable for identification is a challenging task. Preferably, this task should be supported by a data scanning algorithm that automatically searches and marks data intervals of interest.

Relatively little can be found in the literature that directly addresses this problem. In [1], a data removal criterion is presented that uses the singular value decomposition (SVD) technique for discarding data which is only noise dependent and leads to a bigger mean square error (MSE) of the estimated model parameters. Horch introduced in [2] a method for finding transient parts of data after a setpoint change, specifically targetting the identification of time delay [3]. In [4], the authors discuss persistence of excitation for on-line identification of linear models but do not deal with finding intervals of data that are persistently exciting. Data mining techniques have been proposed to give a fully automated modeling and identification, based solely on data. In [5], the authors proposed a method to discover the topology of a chemical reaction network, whereas in [6] a method is proposed to find the dynamical model that generated the

⋆ Patent pending.

Email addresses: d.peretzki@gmx.net (Daniel Peretzki), alf.isaksson@liu.se (Alf J. Isaksson), andrecb@isy.liu.se (André Carvalho Bittencourt), krister.forsman@perstorp.com (Krister Forsman).

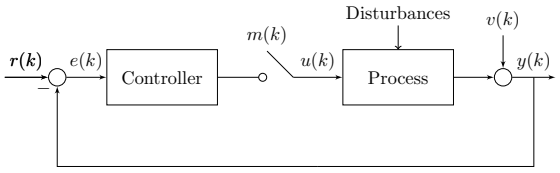


Fig. 1. Control loop.

data using symbolic regression. In [7], the authors consider the use of historical data to achieve process models for inferential control, some guidelines are suggested on how to select intervals of data to build models, but no algorithm is proposed with this objective.

Process plants have specific characteristics that make a fully automated modeling and identification a challenging task, see e.g. [8]. This work focuses on models that are suitable for the design/tuning of low order controllers like PI and PID. For this purpose, it is clear from the above considerations that a fully automated identification is challenging. Instead, **the objective** of this work is to develop a data mining algorithm that retrieves intervals of data from a historic database that are suitable for process identification. The method outputs the intervals together with a quality indicator. The user can then decide on the model and identification method to be used.

1.1 Problem Formulation

Consider the control loop in Fig. 1, at time k the operation mode $m(k)$ is either manual or automatic. In manual mode, the input to the process $u(k)$ is decided by the user. In automatic mode, $u(k)$ is given by the controller. The controller is driven by the control error $e(k)$, formed from the setpoint $r(k)$ subtracted by the measured process output $y(k)$, which is corrupted by noise $v(k)$. It is considered that the system can be described by an unknown model $\mathcal{M}(\theta)$, which is a function of the parameters θ.

A collection of data $Z^N = [Z(1)^T, \cdots, Z(N)^T]^T$ is available, where $Z(k) = [m(k),\ r(k),\ u(k),\ y(k)]$. The objective is to find time intervals $\Delta = [k_{\text{init}},\ k_{\text{end}}]$ where the data in Z^N is suitable to perform identification of the process parameters.

For its practical use, the following characteristics are sought:

I **Minimal knowledge about the plant is required.** That is, none (or little) input is expected from the user.
II **The resulting algorithm should process the data quickly.** For example, a database containing data from a month of a large scale plant operation should not take longer than a few minutes to be processed.
III **For each interval found, a numeric measure of its quality should be given.** This can be used by the user in order to select which intervals to use for identification.

In order to achieve both I and II, some simplifying assumptions are taken:

Assumption 1.1 (SISO) *The complex interconnections present in a plant are disregarded and it is assumed that only SISO control loops are to be estimated.*

Assumption 1.2 (Linear models) *It is assumed that the process can be well described by a linear model $\mathcal{M}(\theta)$.*

For a process in operation there are mainly two scenarios to hope for that may result in data informative enough for system identification (see [9,2]):

- The process is operating in manual mode and the input signal $u(k)$ is varied enough to be exciting the process.
- The controller is in automatic and there are enough changes in the setpoint $r(k)$ to make identification possible.

As a consequence, the method developed below is treating these two cases separately.

2 SYSTEM IDENTIFICATION PRELIMINARIES

Consider first the case of open loop operation. Even under Assumptions 1.1 and 1.2, there are many model structures and numerical identification methods possible. Since we have a clear requirement on low computational complexity it is natural to focus on model structures based on linear regression:

$$y(k) = \varphi^T(\bar{Z}^{k-1})\ \theta + v(k), \tag{1}$$

where the data $\bar{Z}^{k-1}$ contains past inputs $u(k-1), \ldots,\ u(1)$ and outputs $y(k-1), \ldots,\ y(1)$. The vector φ is a regressor and its choice defines the model structure $\mathcal{M}$ of the process. The n dimensional vector θ contains the unknown parameters and $v(k)$ is white noise with variance γ.

A common choice of identification approach is the prediction error method, where the prediction error $\varepsilon(k,\theta) = y(k) - \varphi^T(\bar{Z}^{k-1})\theta$ is minimized according to some criterion. Using a least squares criterion, the estimate is given by

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{k=1}^{N} \left[y\left(k\right) - \varphi^T\left(\bar{Z}^{k-1}\right) \theta \right]^2, \tag{2}$$

which has the solution, [9],

$$\hat{\theta}_N = \hat{R}_N^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(\bar{Z}^{k-1}) y(k), \tag{3}$$

$$\hat{R}_N \triangleq \frac{1}{N} \sum_{k=1}^{N} \varphi(\bar{Z}^{k-1}) \varphi(\bar{Z}^{k-1}), \tag{4}$$

The feasibility of this solution depends on whether the *information matrix* $\hat{R}_N$ is invertible.

Assume that the true system is described by a linear regression with true parameters θ_0, and with a true noise variance γ_0. Then, as $N \to \infty$, the estimate $\hat{\theta}_N$ will be asymptotically normally distributed, [9]. More precisely

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \in As\mathcal{N}(0, P), \quad P \triangleq \gamma_0 \left[\lim_{N \to \infty} \hat{R}_N \right]^{-1}. \tag{5}$$

This means that for finite number of data, N,

$$\hat{\theta}_N \sim \mathcal{N}(\theta_0, P_N), \tag{6}$$

where an estimate of the covariance matrix $\hat{P}_N$ is given by

$$\hat{P}_N = \frac{1}{N} \hat{\gamma}_N \left[\hat{R}_N \right]^{-1}, \quad \hat{\gamma}_N = \frac{1}{N} \sum_{k=1}^{N} \varepsilon^2(k, \hat{\theta}_N). \tag{7}$$

The matrix $\hat{R}_N$ therefore determines the quality of the estimate $\hat{\theta}_N$. Notice that $\hat{R}_N$ is a function of the data Z^N as well as the model structure $\mathcal{M}$. In order to make the covariance small, $\hat{R}_N$ should be made large in some sense. This idea is explored, for instance, in experiment design (choosing u) for system identification. A data set Z^N is therefore suitable for identification of the model structure $\mathcal{M}$ if it is such that the matrix $\hat{R}_N$ is large. Another important piece of information in (5) is the size of $\hat{\gamma}_N$, i.e. how small the optimal prediction errors are. To test how informative an interval of data is in Sec. 3 we define quantities that relate to these measures of identification quality.

However, before defining test quantities, two other relevant issues should be addressed. The first is related to the solution of $\hat{\theta}_N$ and the related quantities. A solution based on explicitly forming the inverse as in (3) is not numerically well conditioned. In the proposed method this is overcome with a numerical solution based on QR factorization, which is presented in Sec. 2.1. The second issue is related to the fact that any condition based on $\hat{R}_N$ demands knowledge of the model structure $\mathcal{M}$, which is assumed to be unknown by the Requirement I. An idea to circumvent this is to use a model structure flexible enough to explain the input-output relation of a large variety of processes. Such issues are addressed in Sec. 2.2.

2.1 Solution to the Least Squares Problem based on QR factorization

An equivalent formulation of (2) is

$$\hat{\theta}_N = \arg\min_{\theta} \|Y - \Phi\theta\|_2^2, \tag{8}$$

where

$$Y^T = \left[y^T(1) \ \cdots \ y^T(N) \right], \Phi^T = \left[\varphi(1) \ \cdots \ \varphi(N) \right]. \tag{9}$$

Because the norm in the minimization is not affected by an orthonormal transformation, apply the QR factorization as $[\Phi \ Y] = QR$, where Q is orthonormal, $QQ^T = I$. R is a matrix of the form

$$R = \begin{bmatrix} R_0 \\ \cdots \\ 0 \end{bmatrix}, \quad R_0 = \begin{bmatrix} R_1 & R_2 \\ 0 & R_3 \end{bmatrix}, \tag{10}$$

where the matrix R_0 is square upper triangular of dimension $n+1$ and R_3 is a scalar. Applying the orthonormal transformation Q^T from the left in (8), then

$$\|Y - \Phi\theta\|_2^2 = \left\| Q^T (Y - \Phi\theta) \right\|_2^2 = \tag{11}$$

$$\left\| \begin{bmatrix} R_2 \\ R_3 \end{bmatrix} - \begin{bmatrix} R_1\theta \\ 0 \end{bmatrix} \right\|_2^2 = \|R_2 - R_1\theta\|_2^2 + |R_3|^2, \tag{12}$$

Hence, the $\hat{\theta}_N$ which minimizes (12) is given by the solution to

$$R_1 \hat{\theta}_N = R_2, \tag{13}$$

Since R_1 is upper triangular, solving for $\hat{\theta}_N$ in (13) is easier and better numerically conditioned than using (3). Similarly,

$$\hat{\gamma}_N = \frac{1}{N} \|Y - \Phi\hat{\theta}_N\|_2^2 = \frac{1}{N} |R_3|^2. \tag{14}$$

Furthermore

$$\hat{R}_N = \frac{1}{N} \Phi^T \Phi = \frac{1}{N} R_1^T R_1. \tag{15}$$

2.2 A Flexible Model Structure

Even restricting to linear regression, there are many potential model structures to choose from. A common

choice is the so-called ARX model structure leading to regressors φ containing sequences of past inputs $u(k-1-d), \ldots, u(k-n_u-d)$ and outputs $y(k-1), \ldots, y(k-n_y)$, where d corresponds to the process delay and n_u, n_y are the model orders. For processes with a nonzero delay (commonly found in the process industry), the delay d needs to be known, in order to form the regressor, which is not the case. Since deadtime estimation is in itself a complicated matter (see e.g. [10]), alternative model structures are considered.

Another appealing model structure is Finite Impulse Response (FIR) models, resulting in regressors only containing lagged values of the input u. In the presence of deadtime an FIR model will lead to close to zero estimates of the leading parameters. The drawback with FIR is that a process with slow dynamics requires very many parameters leading to a very large, and impractical, size of $\hat{R}_N$.

Luckily, there are model structures available that combine the advantages of ARX and FIR models, without suffering from their drawbacks. One such model structure is the Laguerre model:

$$y(k) = \sum_{i=1}^{n} \theta_i \; L_i(q,\alpha) u(k) \tag{16}$$

where $L_i(q,\alpha)$ is the Laguerre filter

$$L_i(q,\alpha) = \frac{\sqrt{1-\alpha^2}}{q-\alpha} \left(\frac{1-\alpha q}{q-\alpha} \right)^{i-1}. \tag{17}$$

Hence, the Laguerre filter of order i consists of one low-pass filter cascaded with $(i-1)$ first-order all pass filters, which acts effectively as a delay approximation of order $i-1$. The substitution of the delay operator with a Laguerre filter has important characteristics that makes it a more suitable choice than an FIR model [11].

The parameter α determines the transient response of the low pass filter. It controls the settling time of the first Laguerre output $L_1(q,\alpha)$ and should be set as equal to the largest time constant in the system [12]. The maximum delay $\bar{d}$ a Laguerre model can explain can be found by comparing a Padé approximation of a delay with the all pass part of the Laguerre filters [10,13] and is given by

$$\bar{d} = -2(n-1)T_s / \log \alpha \tag{18}$$

where T_s is the sampling interval. If the real pole α and order n are selected properly, then the Laguerre model can efficiently approximate a large class of linear systems [2]. In general, the performance of the identification is relatively insensitive to the choice of α [14].

Plants with an integrator

Because a Laguerre model has a finite gain at low frequencies, it is not a good approximation for plants with an integrator. This is in fact an important limitation of Laguerre models since many processes in process industry (most notably level control) have an integrating behavior. To overcome this, it is *assumed known* whether a plant has an integrator or not. For integrating processes, a Laguerre model between the integrated input $\bar{u}(k)$ and output sequences is considered instead:

$$y(k) = \sum_{i=1}^{n} \theta_i L_i(q,\alpha) \bar{u}(k), \quad \bar{u}(k) = \frac{u(k)}{1-q^{-1}}. \tag{19}$$

3 DATA FEATURES FOR PROCESS IDENTIFICATION

Based on the previous section, the testing of three data features with increasing computational complexity and theoretical justification are presented below.

3.1 Variability in the data

The first test of potential excitation is to check that there is an actvitity in the input and output signals at all. An empirical and simple solution is therefore to monitor the signals' variability over time k. Changes in the signal variances can be used with this purpose.

3.2 Numerical Conditioning of $\hat{R}_N$

A more theoretically based approach is to monitor $\hat{R}_N$. From (3), the least squares problem solution is well posed only if $\hat{R}_N$ is invertible. A near singular matrix $\hat{R}_N$ might occur when the input data is not exciting enough to fit a model of order n. The condition number, $\kappa(M)$ of a matrix M is defined as the ratio of the largest and smallest singular values of M. The accuracy with which the model parameters can be estimated are related to the condition number [1]. An information matrix with condition number close to 1 means that the least squares problem is numerically well-conditioned. If a QR algorithm is used, then the relationship $\hat{R}_N = \Phi^T \Phi = R_1^T R_1$ gives

$$\kappa(\hat{R}_N) = \kappa(R_1)^2, \tag{20}$$

This is easily seen by taking the SVD of R_1 and comparing the singular values of R_1 and the ones of $\hat{R}_N = R_1^T R_1$.

3.3 Statistical Significance of $\hat{\theta}_N$

The first two tests proposed above do not consider the actual correlation between the input and output sequences,

and whether it appears as it is possible to fit a linear model between them. A more conclusive test is to in fact compute $\hat{\theta}_N$ and check whether any of the estimated parameters are significantly non-zero.

According to the null hypothesis ($\theta_0 = 0$) and (5), the estimate $\hat{\theta}_N$ is asymptotically normal $\mathcal{N}(0, P)$, i.e.

$$\hat{\mathcal{X}}_N \triangleq \hat{\theta}_N^T \hat{P}_N^{-1} \hat{\theta}_N \in \mathcal{X}_n^2 \tag{21}$$

where the degrees of freedom n for the chi-square distribution is the dimension of the parameter vector θ. Hence checking whether $\hat{\theta}_N$ is non-zero, rejecting the null hypothesis, corresponds to comparing the quantity $\hat{\mathcal{X}}_N$ with the ones in a table of the chi-square distribution.

If the QR solution is used, then

$$\begin{aligned}\hat{\mathcal{X}}_N &= \hat{\theta}_N^T \hat{P}_N^{-1} \hat{\theta}_N = [R_1^{-1} R_2]^T \frac{N}{\hat{\gamma}_N} \hat{R}_N [R_1^{-1} R_2] \\ &= \frac{1}{\hat{\gamma}_N} R_2^T R_2 = \frac{N}{|R_3|^2} R_2^T R_2 = \left\| \frac{\sqrt{N}}{|R_3|} R_2 \right\|_2^2 .\end{aligned}$$

Notice that the quantity $\hat{\mathcal{X}}_N$ behaves directly proportional to the statistical significance of the the parameters. In fact, this quantity can be used to compare the quality of different data sets (larger meaning better).

4 METHOD OUTLINE

At this point, it is possible to define a method to search for suitable data to perform process identification. Laguerre models are used between the input-output data and the idea is to monitor the data features described in the previous section to consider whether the data is relevant. The signals are first scaled and the operating points are removed since linear models are being considered.

The features discussed in Sec. 3 have very different computational complexity. In order to avoid excessive computations, the features are computed conditionally in a cascade of events as

Compute variances of $u(k)$ and $y(k)$, $\mathsf{v}_u(k)$ and $\mathsf{v}_y(k)$.
if $\mathsf{v}_u(k)$ and $\mathsf{v}_y(k)$ are large **then**
  Compute $\kappa(\hat{R}_N)$.
  if $\kappa(\hat{R}_N)$ is large **then**
    Compute $\hat{\mathcal{X}}_N$.
    if $\hat{\mathcal{X}}_N$ is large **then**
      Mark data interval as useful.
    end if
  end if
end if

The ordering also considers that the variances check is more easily satisfied than the condition number, and that the statistical significance is the most demanding test.

Since we are interested in finding relevant changes in the data, it is natural that the features are computed recursively, over a window of data or in a forgetting filter scheme. An efficient implementation can be achieved with an exponential moving average (EMA). For instance, the variance of the output signal, $\mathsf{v}_y(k)$, can be estimated as

$$\begin{aligned}\hat{\mathsf{v}}_y(k) =& \frac{2 - \lambda_{\mathsf{m},y}}{2} \Big[\lambda_{\mathsf{v},y} \left[y(k) - \mathsf{m}_y(k) \right]^2 \\ &+ (1 - \lambda_{\mathsf{v},y})\, \mathsf{v}_y(k-1) \Big] \qquad (22a) \\ \hat{\mathsf{m}}_y(k) =& \lambda_{\mathsf{m},y}\, y(k) + (1 - \lambda_{\mathsf{m},y})\, \mathsf{m}_y(k-1) \qquad (22b)\end{aligned}$$

where m_y is the estimate of the mean and $0 < \lambda_{\mathsf{v},y}, \lambda_{\mathsf{m},y} < 1$ are tuning parameters which control the effective size of the window. A moving average is also used to update the information matrix recursively.

The algorithm flowchart for open-loop data is shown in Fig. 2, where $\mathbf{L}_k = [L_1(q,\alpha)u(k), \cdots, L_n(q,\alpha)u(k)]$, $U = [u(1), \cdots, u(N)]$ and η are thresholds. After loading/scaling the data and removing the operating points, the sample where the input is first changed, k_0 is searched to avoid unnecessary computations. If the plant is integrating, the input signal is integrated. Several quantities used are computed from k_0 to N; the Laguerre outputs $\mathbf{L}$, the regression matrix Φ and the variance estimates of the input and output (computed with an EMA).

The algorithm enters in a loop, searching for excitation in the data k_0 to N. Before any criteria are checked, it is required that $\mathsf{v}_{L_1}(k)$ exceeds a minimum threshold, indicating that excitation might have started. This sample is marked as a candidate for the start of an interval k_{init}. The first check for data excitation is then performed using the estimated variances. Only if passed, the information matrix $\hat{R}_k$ is update using an EMA, followed by checking if the condition number of $\hat{R}_k$ is small enough.

Finally, if all tests so far were successful, the statistical significance of the estimates are compute using QR factorization and the interval is marked up to the current sample, $\Delta = [k_{\text{init}}, k]$. If any test fails, the algorithm moves to the next sample and continues until the data is over. In case any useful data was found, the algorithm outputs the interval Δ and the value of $\hat{\mathcal{X}}_{k_{\text{init}}:k}$ as an estimate of the data quality.

There are a total of 11 design parameters: The order of the Laguerre model, n, and its pole α, the moving average filters coefficients $[\lambda_{L_1}, \lambda_{\mathsf{v},y}, \lambda_{\mathsf{m},y}, \lambda_R]$ and the

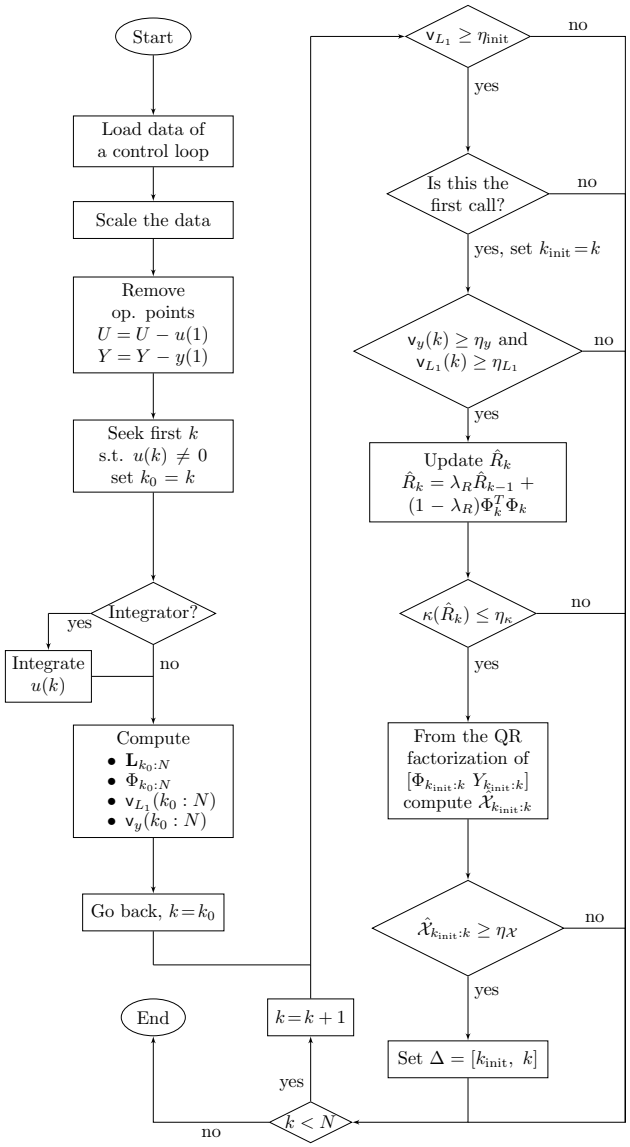


Fig. 2. Algorithm flowchart.

thresholds $[\eta_{\text{init}}, \eta_{L_1}, \eta_y, \eta_\kappa, \eta_{\mathcal{X}}]$. Notice that the same values of these design parameters are used, for any type of control loop or operational mode.

4.1 Closed Loop Data

For data collected in automatic mode, we first search for excitation in the setpoint, i.e. change u to r in the first 2 tests. Then, for the final statistical test, $\hat{\mathcal{X}}_\Delta$ is computed for a tentative model between u to y, because at the end of the day it is an input-out model we aim to identify.

5 TEST DATA EVALUATION

To test the developed method, historic data from a chemical plant was used. It contains data from 211 control loops of density, flow, concentration, level, conductivity, temperature and pressure types. The loops have considerably different dynamics but most of them can be modeled as a first order model and a delay, and it is not hard to tell beforehand which ones that will have an integrator. The delays can vary up to 10 min. The data is mainly from closed loop operation, but there is also open loop data. The database contains data of nearly 37 months of operation, sampled every 15s, in almost 1.1G samples and requires 6.7GB of memory to be stored.

Loops where any Δ was found by mode	
closed loop	143 (67.7%)
open loop	185 (87.7%)
both	190 (90.1%)
Average length of Δ's found (samples) by mode	
closed loop	102.8
open loop	125.3
both	114.1
Average Δ's found by loop type	
Density	239
Flow	660
Concentration	84
Level	130
Conductivity	0
Temperature	35.3
Pressure	100

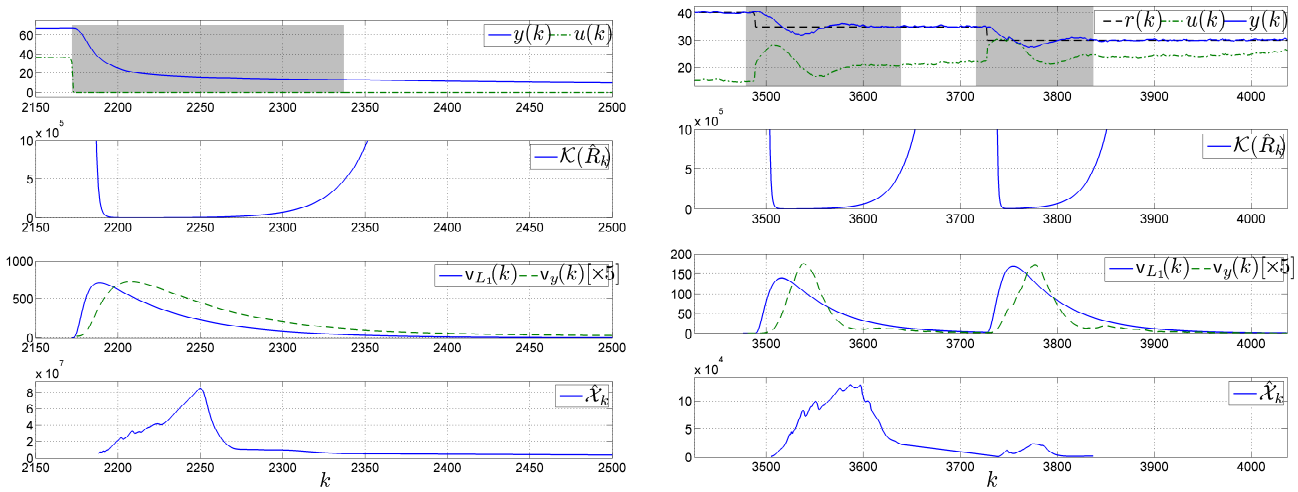
Table 1
Some quantities characterizing the performance of the method when applied to the test data.

The proposed method is applied to these data, taking approximately 1.5h to process it all. Table 1 summarizes the results. In total, about 1.5% of data was found to be useful for process identification.

The 3 year database contains stretches of data where we know that the control group at Perstorp conducted identification experiments, performed in manual mode with a sequence of steps in $u(k)$. All of these intervals and many others were found with the proposed algorithm. The models built using the data intervals selected by the new method were found very similar to those obtained during the identification experiments. The related $\hat{\mathcal{X}}_\Delta$ were also consistently large. Fig. 3 presents two examples of intervals found, in open/closed loop operation. It also illustrates the quantities used to infer the data quality to process identification.

6 CONCLUSIONS

From the quite extensive testing it seems that the method developed in this paper can successfully find intervals of data relevant for system identification. It requires minimal knowledge of the control loops, namely, whether the process is integrating or not. It is implemented efficiently in a recursive manner. A day's worth of data, from all loops in the test plant, takes less than



(a) Temperature, open loop.

(b) Density, closed loop

Fig. 3. The shaded regions are the found identification data intervals.

5s to be processed despite that we have not yet fully optimized the algorithm for computational speed.

The developed method is based on classical results from identification theory of linear systems. As an initial screening, it checks that input and output signals are varying at all. Then it forms an information matrix and checks its condition number. Finally, it estimates a provisional model using a Laguerre model structure. If the parameter estimates of this provisional model are found to be statistically non-zero the data interval is marked as potentially useful for system identification and a quality measure is also provided.

7 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support from the Swedish Foundation for Strategic Research (SSF) as part of the Process Industry Centre Linköping (PIC-LI). The authors would also like to thank Professor Lennart Ljung for encouraging and inspirational meetings early in the project.

References

[1] Carrette P, Bastin G, Genin Y, Gevers M. Discarding data may help in system identification. *IEEE Transactions on Signal Processing*. 1996;44(9):2300–2310.

[2] Horch A. Condition Monitoring of Control Loops. Ph.D. thesis, KTH, Signals, Sensors and Systems. 2000.

[3] Isaksson A, Horch A, Dumont G. Event-triggered deadtime estimation from closed-loop data. *American Control Conference, Arlington VA, June 25-27*. 2001;pp. 3280–3285.

[4] Green M, Moore J. Persistence of excitation in linear systems. *Systems & Control Letters*. 1986;7(5):351–360.

[5] Cho YJ, Ramakrishnan N, Cao Y. Reconstructing chemical reaction networks: data mining meets system identification. In: *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08. New York, NY, USA: ACM. 2008; pp. 142–150.

[6] Schmidt MD, Lipson H. Data-Mining Dynamical Systems: Automated Symbolic System Identification for Exploratory Analysis. *ASME Conference Proceedings*. 2008; 2008(48364):643–649.

[7] Amirthalingam R, Sung SW, Lee JH. Two-step procedure for data-based modeling for inferential control applications. *AIChE Journal*. 2000;46(10):1974–1988.
URL `http://dx.doi.org/10.1002/aic.690461010`

[8] Ng YS, Srinivasan R. Data Mining for the Chemical Process Industry. In: *Encyclopedia of Data Warehousing and Mining*, edited by Wang J, pp. 458–464. IGI Global. 2009;.

[9] Ljung L. *System identification: Theory for the user*. Prentice-Hall Englewood Cliffs, NJ, 2nd ed. 1998.

[10] Björklund S, Ljung L. A review of time-delay estimation techniques. In: *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 3. 2003; pp. 2502 – 2507 Vol.3.

[11] Wahlberg B. System Identification using Laguerre Models. *IEEE Trans on Automatic Control*. 1991;.

[12] Wang L, Cluett W. Building transfer function models from noisy step response data using the Laguerre network. *Chemical Engineering Science*. 1995;50(1):149–161.

[13] Isaksson M. A Comparison of Some Approaches to Time-Delay Estimation. *Master's Thesis ISRN LUTFD2/TFRT--5580--SE*, Department of Automatic Control, Lund University, Sweden. 1997.

[14] Park H, Sung S, Lee I, Lee J. On-line process identification using the Laguerre series for automatic tuning of the proportional-integral-derivative controller. *Ind Eng Chem Res*. 1997;36(1):101–111.