2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Handling of Time Delays in MISO Processes with Regularized Finite Impulse Response Models

Christopher Illg[1], Tarek Kösters[1] and Oliver Nelles[1]

Abstract— **Data driven system identification is the technique for learning models from input/output data. To increase the robustness of the model estimation, prior knowledge can be incorporated, the so-called gray-box identification. In finite impulse response (FIR) models, prior knowledge of the process under investigation can be introduced by regularization. In the regularization term basic impulse response characteristics such as smoothness and exponentially decaying behavior can be incorporated. For estimation of time-delay systems, the novel impulse response and time-delay preserving (IRDP) regularization matrix is proposed. In this contribution this method is extended to the estimation of multiple input single output (MISO) processes and is compared to other state-of-the-art approaches. A linear process with four inputs and different input dynamics and time-delays is investigated. The focus of the evaluation is placed on model quality, time-delay estimation, and computation time. The simulation results point out the superiority of the novel regularization approach in comparison to state-of-the-art methods.**

I. Introduction

Dynamic models are an important basis for many tasks such as prediction, simulation, optimization, control, fault detection and diagnosis. System identification is one way to obtain these models. For example, finite impulse response (FIR) models can be utilized to predict the output from given input data. The main advantages of FIR models are, their inherent stability and robustness with respect to wrong model orders as well as time delays. Moreover, they are linear in their parameters, whereby an efficient estimation with the least squares (LS) method can take place [1]. In addition, the *output error* configuration allows an unbiased parameter estimation [2]. The major drawback is the high variance error. In contrast, autoregressive with exogenous input (ARX) and output error (OE) models typically exhibit a low variance error. However, ARX models have to deal with a bias error due to the *equation error* configuration and the parameters of OE models have to be optimized nonlinearly.

Another disadvantage of ARX models arises when considering multivariate processes: Different inputs share the same denominator polynomial (see Fig. 2c). Therefore, the order of ARX models has to be chosen significantly higher than the orders of the true sub-processes for each input dynamics [3].

To get more robust identification methods, data-driven methods can be combined with prior knowledge. This results in a gray-box approach. Recent publications propose introducing prior knowledge of the investigated process into the estimation of FIR models by regularization.

[1]all authors are with the Research Group Automatic Control – Mechatronics, University of Siegen, 57076 Siegen, Germany, christopher.illg@uni-siegen.de

By applying this concept, the high flexibility of FIR models is limited and therefore, the number of effective parameters can be significantly reduced. Therefore, the high variance error can also be overcome [4].

In this paper, we will discuss a special class of processes called time-delay systems. There exist some state-of-the-art approaches which can determine the time delay, like area and moment methods or higher order statistics methods [5]. These approaches perform a two-step procedure to obtain a good quality model: First, the time delay is determined and the signal is compensated for it. Second, a model without time delay is estimated. In contrast, there are also one-step methods that carry out the time-delay estimation simultaneously with the model estimation. For this, a candidate set of ARX or OE models with different time delays are estimated and the best model is selected with the help of validation data. However, due to the combinatorial nature of this approach in the multiple input single output (MISO) case, many models have to be estimated.

The estimation of time-delay systems with regularized FIR models with the standard regularization schemes in [6], [7] will fail, since they have one limitation: The smoothness and the exponentially decaying behavior assumptions of the impulse response are not fulfilled. Therefore, in previous work a novel regularization matrix is introduced which also allows the estimation of time-delay systems [8]. It utilizes multiple kernels to perform the estimation of the time delay and the model in one step.

The goal of this paper is to extend our proposed approach in [8] to identify MISO time-delay systems. Moreover, this method is compared to other established state-of-the-art approaches:

- Two-stage procedure, first compensating for time delay and then identifying regularized FIR models,
- One-step procedure with ARX models,
- One-step procedure with OE models.

The paper is structured as follows. First, Sect. II deals with regularized FIR models and the novel impulse response and time-delay preserving (IRDP) regularization matrix. It is explained, how to estimate the time delay with a multiple kernel approach and how to deal with the hyperparameter optimization. Afterwards, the extension of the proposed approaches and of other state-of-the-art approaches to the MISO case is addressed in Sect. III. The different model types are evaluated and compared on a MISO test process with four inputs in Sect. IV. Finally, the findings are concluded in Sect. V.

II. Time Delay Estimation with Regularized Finite Impulse Response Models

The output $y(k)$ of a linear single input single output (SISO) process without direct feedthrough can be approximated by an nth-order FIR model. The model output $\hat{y}(k)$ is calculated by the convolution of the delayed inputs $u(k-i)$ and the model parameters θ_i, $i = 1, \ldots, n$

$$\hat{y}(k) = \sum_{i=1}^{n} \theta_i u(k-i) = \underline{x}(k)^T \underline{\theta} \,, \qquad (1)$$

with $\underline{x}(k) = \begin{bmatrix} u(k-1) & u(k-2) & \cdots & u(k-n) \end{bmatrix}^T$ and $u(k) = 0$, for $k \leq 0$. Hereby, the parameter vector $\underline{\theta}$ can be interpreted as an approximation of the impulse response coefficients. The model order n has to be chosen high enough to cover all coefficients of the true impulse response which deviate significantly from zero [1]. Since the model is linear in the parameters, the estimation of the model parameters $\underline{\theta}$ can be done via an LS method. However, the main drawback of FIR models is the high variance error. By introducing regularization, this problem can be overcome [6].

By restricting the flexibility of the model, the effective number of parameters can be drastically reduced: $n_{\text{eff}} \ll n$. This is done by adding a penalty term into the estimation procedure which yields the new objective [7]

$$J = \sum_{k=1}^{N} \left(y(k) - \hat{y}(k)\right)^2 + \lambda \underline{\theta}^T \underline{R}(\underline{\eta}) \, \underline{\theta} \,, \qquad (2)$$

with the regularization matrix $\underline{R}$ the regularization strength λ and the hyperparameter vector $\underline{\eta}$. Through the regularization matrix, prior knowledge is incorporated into the parameter estimation by linking consecutive parameters. With $\lambda = 0$, no prior knowledge is applied and the parameter estimation results in the unregularized solution. A high λ weights prior knowledge more heavily. Therefore, the task is to find an appropriate tradeoff for λ. The solution of (2) can be found in [4] by

$$\underline{\hat{\theta}} = \left(\underline{X}^T \underline{X} + \lambda \underline{R}(\underline{\eta})\right)^{-1} \underline{X}^T \underline{y} \,, \qquad (3)$$

with $\underline{X} = \begin{bmatrix} \underline{x}(1) & \underline{x}(2) & \cdots & \underline{x}(N) \end{bmatrix}^T$ and $\underline{y} = \begin{bmatrix} y(1) & y(2) & \cdots & y(N) \end{bmatrix}^T$. The effective number of the parameters can be calculated from $n_{\text{eff}} = \text{tr}(\underline{S})$ with $\underline{S} = \underline{X}\left(\underline{X}^T \underline{X} + \lambda \underline{R}(\underline{\eta})\right) \underline{X}^T$. An easier interpretation of the regularization matrix $\underline{R}$ results from its Cholesky decomposition

$$\underline{\theta}^T \underline{R} \, \underline{\theta} = \underline{\theta}^T \underline{F}^T \underline{F} \, \underline{\theta} = ||\underline{F} \, \underline{\theta}||^2 \,. \qquad (4)$$

This allows the interpretation of the regularization as a filtering of the parameters $\underline{\theta}$ with the filter matrix $\underline{F}$ [11]. Therefore, it enables the incorporation of the prior knowledge directly via the filter matrix $\underline{F}$.

A. Impulse Response Preserving (IRP) Matrix

In [7] a novel regularization method is contributed, which assumes a transfer function of the process under investigation. Thus, the prior knowledge of the process is restricted to a linear n_{IRP}th-order transfer function. However, for the method in [7] only the denominator polynomial $A(z)$ is required. Here, a transfer function with $n_{\text{IRP}} = 2$

$$G(z) = \frac{B(z)}{a_0 + a_1 z^1 + z^2} \,. \qquad (5)$$

is chosen as an example. An extension to lower- and higher-order transfer functions is straightforward. The filter matrix $\underline{F}_{\text{IRP}}(\underline{a}) \in \mathbb{R}^{(n-n_{\text{IRP}}) \times n}$ of the second-order impulse response preserving (IRP$_2$) matrix is defined as [7]

$$\underline{F}_{\text{IRP}}(\underline{a}) = \begin{bmatrix} a_0 & a_1 & 1 & 0 & \cdots & 0 \\ 0 & a_0 & a_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_0 & a_1 & 1 \end{bmatrix} . \qquad (6)$$

The better the transfer function $G(z)$ fits the actual dynamics of the process, the more accurate the estimation of the model parameters will be. Therefore, the choice of coefficients $\underline{a} = [a_0 \; a_1]^T$ is important concerning the resulting FIR model. Since the regularization only penalizes the FIR parameters (and the strength depends on λ), the choice of the transfer function order n_{IRP} is not as crucial here as for the ARX or OE model order.

For stable processes, the latter parameters of the impulse response always decay exponentially towards zero. If the true order is not known, exponential weighting is suggested in [7], which reduces the regularization strength of the first parameters and allows for more flexibility. The weighted filter matrix is defined as $\underline{\tilde{F}}_{\text{IRP}} = \underline{W}_{\text{exp}} \underline{F}_{\text{IRP}}$, with the exponential weighting matrix $\underline{W}(\alpha) \in \mathbb{R}^{(n-n_{\text{IRP}}) \times (n-n_{\text{IRP}})}$

$$\underline{W}_{\text{exp}}(\alpha) = \text{diag}\left(\alpha^{-0}, \alpha^{-\frac{1}{2}}, \alpha^{-1}, \cdots, \alpha^{-\frac{n-n_{\text{IRP}}-1}{2}}\right) . \qquad (7)$$

Here, the hyperparameter α can either be optimized or chosen heuristically.

This approach is also called the single kernel method, because only one regularization matrix $\underline{R}$ with one corresponding λ is utilized.

B. Impulse Response and Time-Delay Preserving (IRDP) Matrix

The IRP matrix introduced in Section II-A is not able to estimate time-delay systems well, as they do not correspond to the assumed transfer function $G(z)$. Thus, only low regularization strengths can be applied. Therefore, in [8] a novel regularization matrix, called *impulse response and time-delay preserving* (IRDP) matrix, has been proposed. The impulse response is divided into two parts, the part containing the time-delay parameters and part with the dynamic parameters, see Fig. 1. The IRDP matrix is defined as

$$\underline{R}_{\text{IRDP}}(d, \beta, \underline{a}, \alpha) = \begin{bmatrix} \beta \underline{I}_d & \underline{0} \\ \underline{0} & \underline{\tilde{R}}_{\text{IRP}}(\underline{a}) \end{bmatrix} , \qquad (8)$$

with the identity matrix $\underline{I}_d \in \mathbb{R}^{d \times d}$ and $\underline{\tilde{R}}_{\text{IRP}} \in \mathbb{R}^{(n-d) \times (n-d)}$. The parameter β scales the ridge regression term for the time-delay parameters and the higher β, the more the parameters are pushed towards zero. This regularization

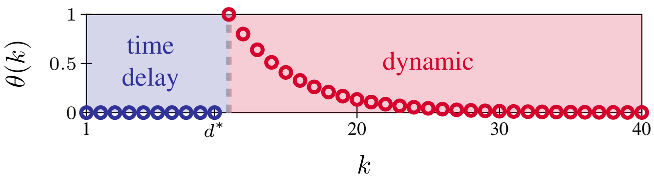


Fig. 1: Separation of the impulse response into time-delay and dynamic parameters. The process has a time delay of $d^* = 10$.

matrix can only be applied, if the time delay d is known a priori. In practice, this is oftentimes not the case. Therefore, a multiple kernel method is proposed to identify the time delay directly together with the other hyperparameters and avoid a mixed integer optimization.

C. Multiple Kernel Method

In [9] a multiple kernel method has been proposed. This method is modified in [8] which allows for a time-delay estimation. Therefore, n_{MK} matrices $\underline{R}_{\text{IRDP}}$ with different time delays are applied in the estimation procedure. The $\underline{R}_{\text{IRDP}}$ matrices are built with $\underline{\delta} = [d_{\text{min}}, d_{\text{min}}+1, \dots, d_{\text{max}}]$. Each regularization matrix is multiplied with a separate regularization strength λ_i and summed up to

$$\underline{R}_{\text{MK}}(\underline{\lambda}, \beta, \underline{a}, \alpha) = \sum_{i=1}^{n_{\text{MK}}} \lambda_i \underline{R}_{\text{IRDP}}(\delta_i, \beta, \underline{a}, \alpha)\,. \tag{9}$$

The number of multiple kernels n_{MK} and thus the considered time delays as well as the ridge parameter β have to be set a priori. Since the regularization strength λ is already included in (9), the model parameters are estimated with

$$\hat{\underline{\theta}} = \left(\underline{X}^T \underline{X} + \underline{R}_{\text{MK}}(\underline{\lambda}, \beta, \underline{a}, \alpha)\right)^{-1} \underline{X}^T \underline{y}\,. \tag{10}$$

To extract the time delay $\hat{d}$ from the different regularization matrices after identification, the assumed time delays δ_i are weighted with their corresponding λ_i:

$$\hat{d} = \text{round}\left(\frac{\sum_{i=1}^{n_{\text{MK}}} \lambda_i \delta_i}{\sum_{i=1}^{n_{\text{MK}}} \lambda_i}\right)\,. \tag{11}$$

D. Hyperparameter Optimization

To determine the hyperparameters of the IRDP kernel $\underline{a}$, α, and $\underline{\lambda}$, a hyperparameter optimization can be performed. In [8] a study is carried out which demonstrates the superiority of the generalized cross-validation (GCV) error over the marginal likelihood (ML) function for the IRDP kernel. Therefore, it is considered in the following investigations. In [12] the GCV error is defined as

$$J_{\text{GCV}} = \frac{1}{N} \sum_{k=1}^{N} \left(\frac{y(k) - \hat{y}(k)}{1 - \text{tr}(\underline{S})/N}\right)^2, \tag{12}$$

with the smoothing matrix $\underline{S} = \underline{X}\left(\underline{X}^T \underline{X} + \lambda \underline{R}\right)^{-1} \underline{X}^T$. The objective J_{GCV} has to be minimized to determine the optimal hyperparameters and estimate the time delay.

III. Estimation of Time Delay MISO Systems

There exist several methods for estimating time-delay systems in the SISO case. They can be applied with minor modifications also for MISO processes. In Fig. 2 the model structures of FIR, OE and ARX models with exemplary two inputs are illustrated. Since the time delay of each input j is contained in $B_j(q)$, it can be estimated separately. In the following, four methods are described to estimate the different polynomials $A_j(q)$ and $B_j(q)$.

A. Multiple Kernel Regularized FIR (MKR-FIR) Models

The output of a MISO FIR model becomes

$$\hat{y}(k) = \sum_{j=1}^{n_{\text{u}}} \underline{x}_j(k)^T \underline{\theta}_j\,, \tag{13}$$

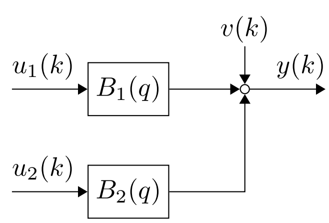
where n_{u} is the number of inputs and each vector $\underline{\theta}_j, \underline{x}_j(k) \in \mathbb{R}^{n_j \times 1}$ can have a separate FIR model order n_j. The model parameters are estimated as described in (10) with a stacked regressor matrix

$$\underline{X} = \begin{bmatrix} \underline{X}_{u_1} & \underline{X}_{u_2} & \cdots & \underline{X}_{u_{n_{\text{u}}}} \end{bmatrix}, \tag{14}$$

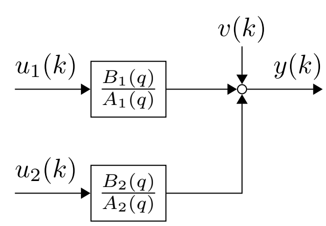
where $\underline{X}_{u_j}$ denotes the regressor matrix of the jth input and the new regularization matrix with multiple kernels

$$\underline{R}_{\text{MK}}(\underline{\eta}_{\text{MK}}) = \begin{bmatrix} \underline{R}_{\text{MK},1} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{R}_{\text{MK},2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \underline{0} \\ \underline{0} & \cdots & \underline{0} & \underline{R}_{\text{MK},n_{\text{u}}} \end{bmatrix}. \tag{15}$$

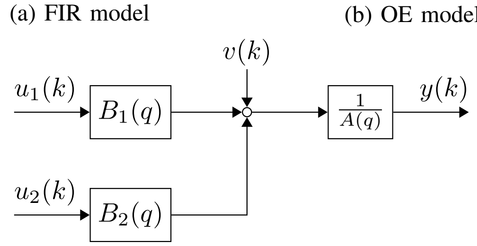
The hyperparameter vector $\underline{\eta}_{\text{MK}}$ contains all $\underline{a}_j, \alpha_j$ and $\underline{\lambda}_j$ for each input. Different hyperparameter choices are possible for each input depending on the different dynamics (also different orders of the assumed transfer function $G(z)$ in (5)). Therefore, it is also possible to assume different time delays $\underline{\delta}_j$ for each input separately. Thus, the number of



(a) FIR model



(b) OE model



(c) ARX model

Fig. 2: Model structures of FIR, OE and ARX models for MISO systems with exemplary two inputs.

hyperparameter is depended on the assumed input dynamics and time delays.

B. Single Kernel Regularized FIR (SKR-FIR) Models

By estimation of time-delay systems with a single kernel method, proposed in Sect. II-B the time delay $\hat{d}_j$ for each input has to be known or determined a priori. This results in a two-step procedure. The time delay of each input is determined by the *CUSUM* detector proposed in [5]. Firstly, the unregularized impulse response coefficients are estimated and the internal algorithm parameters of the CUSUM detector are set to $h_{\text{std}} = 2$ and $v_{\text{std}} = 1$ according to [13].

Afterwards there are two estimation possibilities: (i) Estimate only the parameters from $\hat{d}$ (setting the time-delay parameters to zero) or (ii) the estimation of all parameters with the IRDP matrix (8). The second estimation method is more robust with respect to wrongly estimated time delays and can slightly compensate wrong time delays $\hat{d}$ by a lower regularization strength. Therefore, in the following the second estimation method is pursued for the MISO case.

With stacked regressor (14) and the regularization matrix with single kernels

$$\underline{R}(\underline{\eta}_{\text{SK}}) = \begin{bmatrix} \underline{R}_{\text{IRDP},1}(\hat{d}_1) & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{R}_{\text{IRDP},2}(\hat{d}_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \underline{0} \\ \underline{0} & \cdots & \underline{0} & \underline{R}_{\text{IRDP},n_{\text{u}}}(\hat{d}_{n_{\text{u}}}) \end{bmatrix}. \tag{16}$$

the parameters can be estimated via (3). By externalizing the determination of the time delays $\underline{d}$, the number of hyperparameters can be reduced. For the regularization strength only one λ_j instead of n_{MK} for each input has to be optimized.

C. ARX Models

Each input of ARX models has an individual numerator polynomial $B_j(q)$, but there exist only one denominator polynomial $A(q)$. Therefore, the order of $A(q)$ has to be chosen in such a way that the different poles of all input dynamics can be realized in it. By a suitable choice of the zeros, the not required poles of the respective input dynamics can be compensated. To integrate the time delay into the ARX model, it has to be specified in advance. The output of the ARX model $\hat{y}$ is defined as

$$\begin{aligned} \hat{y}(k) = -\sum_{i=1}^{m} a_i y(k-i) + \sum_{i=\hat{d}_1+1}^{\hat{d}_1+m} b_{i,1} u_1(k-i) + \cdots \\ + \sum_{i=\hat{d}_{n_{\text{u}}}+1}^{\hat{d}_{n_{\text{u}}}+m} b_{i,n_{\text{u}}} u_{n_{\text{u}}}(k-i)\,, \end{aligned} \tag{17}$$

with an order m of the ARX model. Here, the order for all numerator and the denominator polynomials is chosen to be identical. Only the time delay $\hat{d}_j$ is chosen independently for each input. Thus, a time consuming search for the individual time delays is required, which increases combinatorially with the number of inputs and the time delays to be considered. Nevertheless, this can still be done efficiently since the ARX models can be estimated with an LS method.

TABLE I: Four different transfer functions: a first-order, a second-order, a oscillating second-order, and a third-order transfer function with different time delays d_j.

Transfer function	Parameter
$G_1 = \dfrac{q^{-1}q^{-d_1}(1-p_1)}{(1-p_1 z^{-1})}$,	$p_1 = 0.92$, $d_1 = 8$
$G_2 = \dfrac{q^{-1}q^{-d_2}(1-p_1)(1-p_2)}{(1-p_1 z^{-1})(1-p_2 z^{-1})}$	$p_1 = 0.9$, $p_2 = 0.7$ $d_2 = 7$
$G_3 = \dfrac{q^{-1}q^{-d_3}(1-p_1)(1-p_2)}{(1-p_1 z^{-1})(1-p_2 z^{-1})}$	$p_{1,2} = 0.65 \pm 0.65i$, $d_3 = 5$
$G_4 = \dfrac{q^{-1}q^{-d_4}(1-p_1)(1-p_2)(1-p_3)}{(1-p_1 z^{-1})(1-p_2 z^{-1})(1-p_3 z^{-1})}$	$p_1 = 0.9$, $p_2 = 0.7$, $p_3 = 0.1$, $d_4 = 2$

D. OE Models

For OE models each input dynamics can be estimated by its own numerator and denominator polynomials $B_j(q)$ and $A_j(q)$. The model output of OE models $\hat{y}$ is given by

$$\begin{aligned} \hat{y}(k) = -\sum_{i=1}^{m_1} a_{i,1} y(k-i) + \sum_{i=\hat{d}_1+1}^{\hat{d}_1+m_1} b_{i,1} u_1(k-i) + \cdots \\ -\sum_{i=1}^{m_{n_{\text{u}}}} a_{i,n_{\text{u}}} y(k-i) + \sum_{i=\hat{d}_{n_{\text{u}}}+1}^{\hat{d}_{n_{\text{u}}}+m_{n_{\text{u}}}} b_{i,n_{\text{u}}} u_{n_{\text{u}}}(k-i)\,. \end{aligned} \tag{18}$$

The model orders $m_j \ll m$ of the respective input dynamics can be chosen independently, as it was already the case for the assumed transfer function of regularized FIR models. The same intensive search for the time delay has to be done, but the parameters have to be estimated via nonlinear optimization. Thus, this leads to a much more time-consuming model estimation.

IV. Results

In this section, the four models types – MKR-FIR, SKR-FIR, ARX, and OE models – are compared. Herefore, a MISO process with different time delays d_j is investigated.

A. Simulation Setup

The output of the MISO process can be calculated by

$$y(k) = G_1 u_1(k) + G_2 u_2(k) + G_3 u_3(k) + G_4 u_4(k)\,, \tag{19}$$

with the four inputs $u_j(k)$. The four transfer functions are defined in Table I.

Four independent pseudo random binary signals (PRBS) with $N = 1\,000$ data samples are provided as excitation. The output is disturbed by white Gaussian noise such that a signal-to-noise-ratio (SNR) of $10\,\text{dB}$ results. To test the robustness of the methods regarding time-delay estimation and the model quality, a Monte Carlo study with 100 runs with different noise realizations is performed.

For the MKR-FIR as well as the ARX and OE model time delays in the range $\underline{\delta} = [0, 1, \cdots, 15]$ are assumed for each input. This corresponds to $16^4 = 65\,536$ models which have to be estimated for the ARX and OE time-delay estimation. For all FIR models $n_j = 90, \forall j \in [1, 2, 3, 4]$ is assumed. The order of the numerator and denominator of the ARX model is chosen as $m = 8$ to be able to include the poles of each input, since all inputs use the same denominator polynomial. By choosing $B_i(q)$ appropriately, the non-required poles are shortened. The OE model order of each input is set to $m = 2$, because each input has a separate denominator polynomial and all input dynamics can be described quite well with second-order transfer functions. Furthermore, the correct model order should not be chosen in this study, since typically, the exact model order is not known in real world applications. For better comparability, the order of assumed transfer functions for the regularization methods is chosen the same as for the OE models ($n_{\text{IRP}} = 2$ for each input dynamic). In [8] the GCV error turns out to be the best quality criterion for hyperparameter tuning, and therefore it is applied here. For the parameter $\beta = 1\,000$ is chosen heuristically. For the estimation of ARX and OE models, the training data set is split. Only the first $80\,\%$ of the training data is taken for training and the other $20\,\%$ for model selection (time-delay estimation).

To evaluate the model quality, an independent PRBS test signal is created and the normalized root mean square error (NRMSE) time-delay error

$$\Delta d_j = d_j - \hat{d}_j \tag{20}$$

is evaluated.

B. Comparison of the Proposed Methods

In the following the four methods are compared. In Fig. 3 the error on test data is illustrated. ARX models give the worst results. The MKR-FIR models are the best with respect to both mean and median error as well as the narrowness of the error distribution.

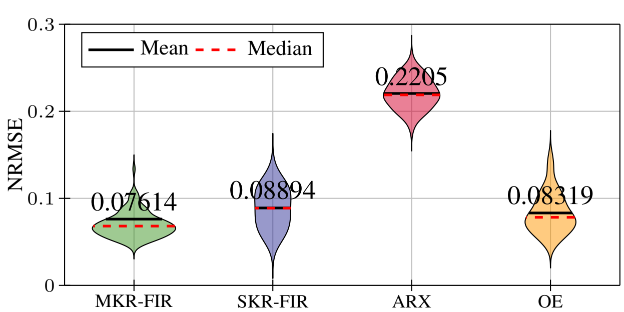


Fig. 3: Violine plot of the NRMSE for the different noise realizations on test data. The four different model types MKR-FIR, SKR-FIR, ARX, and OE models are investigated and the mean error is labeled.

In Fig. 4 the time-delay estimation errors are visualized for the different methods and inputs. Especially, for u_3 the time-delay estimation is very precise, what can be seen by a maximum estimation error of only one time step. Also for the first-order transfer function G_1, the regularized FIR models yield good results in the time-delay estimation. The variance of the time-delay estimation of ARX and OE models is significantly larger than for the regularized FIR models. Thereby, OE models provide for all inputs the correct median time delay. For ARX models, the median time-delay estimation error is zero only once. This in combination with the equation error configuration causes the low model quality.

For the two estimation methods – MKR-FIR and OE – 40 impulse responses are presented in Fig. 5. It demonstrates effect of the high variance in the time-delay estimation of the OE models (especially for G_1 and G_2). The errors of the sub-process G_3 are less visible due to the higher values of the coefficients (same gain), but they are similar in relative terms.

The computation time of the different model estimations can be seen in Fig. 6. The model types with the best model performance also take the most time for the estimation (MKR-FIR, OE). The SKR-FIR method requires the least computation time, since on the one hand the number of hyperparameters is quite low and on the other hand only two models have to be estimated – an unregularized FIR model to perform the time-delay estimation by the CUSUM detector and a regularized FIR model including the hyperparameter optimization to obtain a good model. The estimation of ARX models with an LS method always need the same time. Therefore, it has a small variance in computation time. However, it has to be executed quite often due to the combinatorial options of the time delays. Due to the nonlinear optimization required for the estimation of OE models, it takes significantly longer in comparison to the ARX model estimation, although the same number of models have to be estimated.

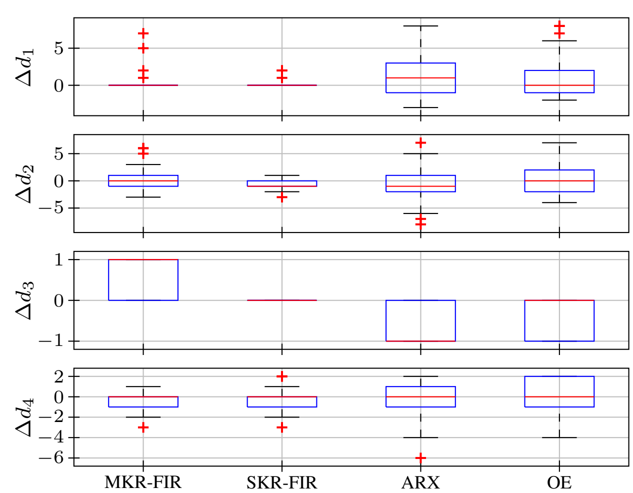


Fig. 4: Box plot of the time-delay estimation errors plotted for the different models and inputs.

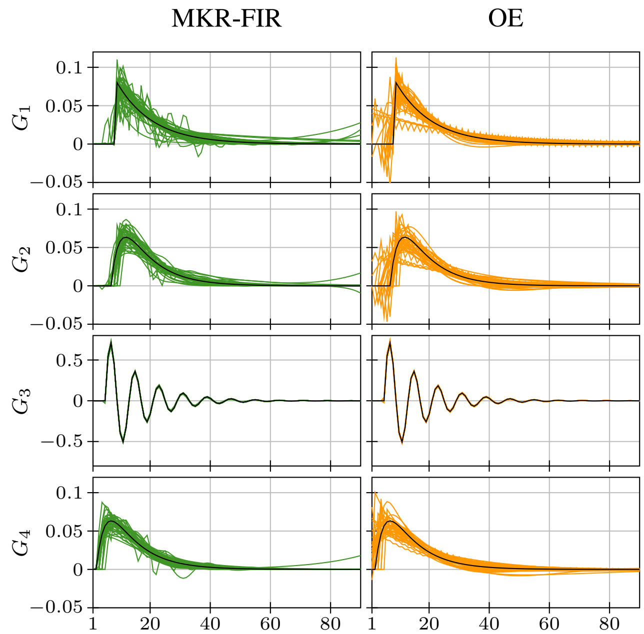


Fig. 5: 40 impulse responses of the estimation methods MKR-FIR and OE.

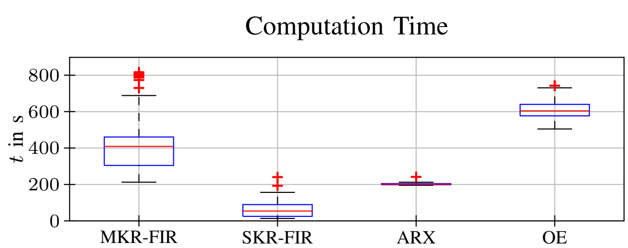


Fig. 6: Box plot of the required computation time of the respective estimation method.

A part of the test signals' output along with the estimated outputs of each median model is illustrated in Fig. 7. It can be seen that the ARX model is significantly worse than the other three model types. This is caused by the equation error configuration and the bad time-delay estimation. The other model types are comparable, whereas the model output of the MKR-FIR is most similar to the process output y.

It has to be noted, by considering even more inputs, the number of ARX and OE models to be investigated grows exponentially. In contrast, the number of hyperparameters for the regularized FIR models increases only linear.

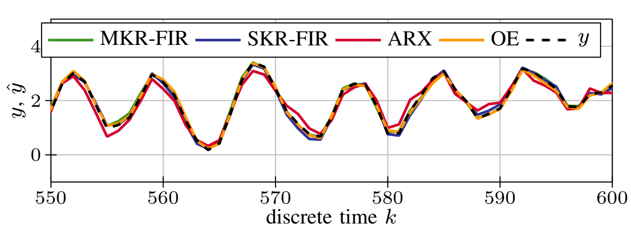


Fig. 7: The output of the median models on the test data set.

V. Conclusion

We extended novel regularization schemes for estimating time-delay systems with regularized FIR models for the multiple input single output (MISO) case. Two methods are proposed: Single kernel regularized FIR models (SKR-FIR) and multiple kernel regularized FIR models (MKR-FIR). Both methods require the novel impulse response and time-delay preserving (IRDP) matrix to estimate time-delay system. These methods are compared with common autoregressive with exogenous input (ARX) and output error (OE) models, for which a combinatorial search for the time delays is performed. It is demonstrated that the hyperparameter optimization in regularized FIR models leads to superior model estimations (although the number of hyperparameters is high). OE models also yield good results, however it comes with the drawback of long computation times as well as a high variance in the time-delay estimation.

Further research will done by applying the multiple kernel regularization scheme to nonlinear time-delay processes. This will be realized by linear local model networks.

References

[1] L. Ljung, *System Identification: Theory for the User*. Prentice Hall PTR, 1999, no. 2.

[2] O. Nelles, *Nonlinear System Identification from Classical Approaches to Neural Networks, Fuzzy Systems, and Gaussian Processes*, 2nd ed. Springer International Publishing AG, 2021.

[3] T. Münker and O. Nelles, "Local Model Network with Regularized MISO Finite Impulse Response Models," in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, Jul. 2016.

[4] G. Pillonetto, F. Dinuzzo, T. Chen, G. D. Nicolao, and L. Ljung, "Kernel Methods in System Identification, Machine Learning and Function Estimation: A Survey," *Automatica*, vol. 50, no. 3, pp. 657–682, Mar. 2014.

[5] S. Björklund, "A Survey and Comparison of Time-delay Estimation Methods in Linear Systems," Master's thesis, University Linköping, 2003.

[6] G. Pillonetto and G. D. Nicolao, "A New Kernel-based Approach for Linear System Identification," *Automatica*, vol. 46, no. 1, pp. 81–93, Jan. 2010.

[7] T. Münker, T. J. Peter, and O. Nelles, "Gray-box Identification with Regularized FIR Models," *at - Automatisierungstechnik*, vol. 66, no. 9, pp. 704–713, Sep. 2018.

[8] T. Kösters, C. Illg, and O. Nelles, "Handling of Time Delays in SystemIdentification with Regularized FIR Models," *at - Automatisierungstechnik*, 2023, (submitted).

[9] T. Chen, M. S. Andersen, L. Ljung, A. Chiuso, and G. Pillonetto, "System Identification Via Sparse Multiple Kernel-based Regularization Using Sequential Convex Optimization Techniques," *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 2933–2945, Nov. 2014.

[10] X. Chen, Z. Mao, and R. Jia, "A New Multiple Kernel-based Regularization Method for Identification of Delay Linear Dynamic Systems," *Chemometrics and Intelligent Laboratory Systems*, vol. 199, p. 103971, Apr. 2020.

[11] A. Marconato, M. Schoukens, and J. Schoukens, "Filter-based Regularisation for Impulse Response Modelling," *IET Control Theory & Applications*, 2016.

[12] G. H. Golub, M. Heath, and G. Wahba, "Generalized Cross-validation As a Method for Choosing a Good Ridge Parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, May 1979.

[13] S. Björklund, "Experimental Evaluation of some Cross Correlation Methods for Time-Delay Estimation in Linear Systems," Division of Automatic Control, Linköping University, Tech. Rep. available, 4 2003.