#### Methods

Tarek Kösters\*, Christopher Illg\* and Oliver Nelles

# Handling of time delays in system identification with regularized FIR models

Systemidentifikation mit regularisierten FIR-Modellen unter Berücksichtigung von Totzeiten

https://doi.org/10.1515/auto-2023-0007 Received February 2, 2023; accepted July 7, 2023

Abstract: System identification with finite impulse response (FIR) models has been recently revised by introducing specific forms of regularization in the estimation. By utilizing prior knowledge of the dynamic process under investigation different kernels for regularization can be derived. However, the dynamical processes considered are mainly restricted to time-delay free systems. Therefore, in this paper we propose two novel methods to handle the time-delay case. Both methods incorporate the time-delay estimation in the hyperparameter optimization of the regularized FIR models. The first method is built on a single kernel (SK) and the second utilizes a multiple kernel (MK) approach. Simulations on different dynamical systems show the superior behavior of both methods in comparison to standard regularization methods, whereupon the MK outperforms the SK approach.

Keywords: hyperparameter optimization; regularized FIR models; system identification; time-delay systems

FIR-Zusammenfassung: Systemidentifikation mit Modellen hat durch die Einführung von neuen Formen von Regularisierung kürzlich deutlich an Aufmerksamkeit gewonnen. Mit dieser Methodik lässt sich der hohe Varianzfehler, der typischerweise den Einsatz von FIR-Modellen beschränkt, signifikant reduzieren. Anhand von Vorwissen über den zu untersuchenden dynamischen Prozess können verschiedene Regularisierungskernel

(T. Kösters), christopher.illg@uni-siegen.de (C. Illg)

abgeleitet werden. Bisher beschränkt sich die Betrachtung allerdings auf totzeitfreie Systeme. Daher werden zwei neue Methoden entwickelt, welche auch Totzeitprozesse berücksichtigen. Beide Methoden integrieren Totzeitschätzung in die Hyperparameteroptimierung der regularisierten FIR-Modelle. Die erste Methode basiert auf einem Single-Kernel (SK), während die zweite Methode einen Multiple-Kernel (MK) Ansatz verwendet. Simulationen an verschiedenen dynamischen Totzeitsystemen zeigen das Potential beider Methoden im Vergleich zu Standard-Regularisierungsmethoden.

Schlagwörter: Totzeitsysteme; regularisierte FIR-Modelle; Systemidentifikation; Hyperparameteroptimierung

# 1 Introduction

Linear system identification solves the problem of generating dynamic models from measured input/output data [1]. One approach to address this problem is the utilization of finite impulse response (FIR) models and the identification of their parameters by least squares (LS) [2]. In recent publications specific forms of regularization were discovered as a possibility to overcome the main drawback of high variance errors that FIR models suffer from [3, 4].

Regularization in general offers the possibility to reduce the flexibility of the parameter estimation. So called kernels allow for the adaption of the regularization to problem specific requirements. This is utilized in the field of regularized FIR (RFIR) models to incorporate prior knowledge of the process properties into the estimation. Several regularization matrices have been developed implying different assumptions on the FIR model parameters. Popular methods to apply regularization use the stable-spline (SS) kernel [5], tuned-correlated (TC) kernel [6], or the impulse response preserving (IRP) [7] matrix.

These regularization matrices incorporate prior knowledge of the impulse response of dynamical systems such as exponential decay (TC, SS) or information on poles of the process (IRP). Using these properties allows to derive

പ്പ

<sup>\*</sup>Corresponding authors: Tarek Kösters and Christopher Illg, Automatic Control - Mechatronics, Institute of Mechanics and Control Engineering - Mechatronics, University of Siegen, Paul-Bonatz-Str. 9-11, 57068 Siegen, Germany, E-mail: tarek.koesters@uni-siegen.de

Oliver Nelles, Automatic Control - Mechatronics, Institute of Mechanics and Control Engineering - Mechatronics, University of Siegen, Paul-Bonatz-Str. 9-11, 57068 Siegen, Germany, E-mail: oliver.nelles@uni-siegen.de

differently structured kernels. The parameterization of the kernels which is dependent on kernel-specific hyperparameters, however, depends on the process and the available data for the estimation. The tuning of the best set of hyperparameters is called hyperparameter optimization [3]. Different optimization goals have been studied [8, 9]. They can either be applied to directly optimize the hyperparameters the kernel is parameterized with [7, 10] or can be used to optimize the regularization strength of preliminary parameterized multiple kernels [11, 12]. Compared to single kernels, multiple kernels can capture complicated dynamics in a better way or can be used for structure detection.

By using regularization it has been shown that the estimation of a wide range of dynamic processes can be improved, even for only partially true assumptions [3, 7]. However, this is not true for time-delay processes [12].

The main reason for this is that time-delay processes violate the fundamental assumption of smooth impulse responses all mentioned kernels have in common. Timedelay systems possess a finite time interval before the impulse response of the dynamic part starts. The behavior of parameters corresponding to this time interval follows different rules than the parameters of the remainder of the dynamic impulse response that follows and therefore need to be treated differently.

In the literature different approaches to deal with time delays are distinguished [13]. In general, two-step and singlestep procedures are differentiated. Two-step procedures identify the time delay first and subsequently perform a standard dynamic system identification on the time-delay compensated signals. Single-step methods, however, estimate the dynamic parameters along with the time delay.

Cross-correlation analysis of the input-output signals, as presented in [14, 15], belong to the two-step procedures. Another form of two-step procedures are called time-delay approximation methods, or area and moment methods [13], which analyze an estimated impulse or step response, respectively.

The Padé approximation is a famous single-step procedure [16]. Here, the time delay is approximated by a non-minimal phase transfer function. The drawback of this method is the possible poor performance for large time delays and the increased number of parameters [17]. Explicit time-delay parameter estimations also follow a single-step procedure by estimating the time delay along with the dynamic parameters. In [18, 19] the problem of identifying the time delay is transformed in a way, that the regression vector of the dynamic parameters is extended to simultaneously estimating the time delay along with the dynamic parameters. The resulting problem is solved by an iterative or recursive application of the LS algorithm. Another common approach is to assume high-order autoregressive models with exogenous inputs (ARX) and varying the time delay to generate a set of candidate models. With a brute-force approach the appropriate model and its corresponding time delay can be found by selecting the best performing model from this set [13]. Here, an *a priori* order selection of the ARX models is needed to estimate the dynamic part of the model.

The motivation of this contribution is to overcome the drawback of an explicit order selection while preserving a single step method. RFIR models structurally can cope with any process order (order of differential equation) [3]. Therefore, no explicit model order selection is required and they are perfectly suited to solve the given problem. However, their issues to deal with time-delay systems are only rarely addressed yet. For example [12] approaches it by assuming that the parameters of the FIR model can be split into two parts, whereas each part follows an exponential decay. The calculation of this splitting point, however, is excluded from the main estimation, resulting in a two-step procedure.

Our approach is to avoid the two step procedure by including the time-delay estimation into the hyperparameter optimization of the kernel. To do so, we propose two novel methods. One is based on a single kernel by overlapping two different kernels smoothly. The second approach utilizes the multiple kernel approach to find the best fitting kernel for the given data.

The paper is structured as follows. Section 2 explains the special properties of time-delay systems. Thereupon, Section 3 introduces FIR models and the regularized estimation along with the hyperparameter optimization. Section 4 shows the incorporation of the time delay into the kernel. Subsequently, the results of the proposed methods are demonstrated on various example processes of different orders in Section 5. Finally we conclude our findings in Section 6.

# 2 Time delay systems

Time-delay systems (TDS), also called dead time systems, are a special class of dynamic systems. Their representation differs from the typical rational transfer function in the continuous-time case. For the discrete-time case, however, a rational description is possible, if the time delay corresponds to an integer multiple of the sampling time  $T_0$ :

$$T_d = d \cdot T_0 \quad d \in \mathbb{N}. \tag{1}$$

This case is the subject of the following investigation. Furthermore, we focus on linear causal discrete-time



**Figure 1:** Impulse response  $\tilde{q}(k)$  of time-delay system distinguished in time delay part • and dynamic remainder • up to the 40th time step.

systems without direct feedthrough such that the output y can be described by the convolution sum of the input *u* and the impulse response of the process g

$$y(k) = \sum_{i=1}^{\infty} g(i)u(k-i).$$
 (2)

For time-delay systems the impulse response is extended with a zero part (**o**) in front of the remainder dynamic impulse response (**o**), see Figure 1. Exemplary, the time delay is set to d = 10.

For the following, the impulse response of such a system is referred to as  $\tilde{g}$  which is defined by its two parts

$$\tilde{g}(k) = \begin{cases} 0 & 1 \le k \le d \\ g(k-d) & k > d \end{cases}$$
(3)

where g(k) is the impulse response of a dynamic system without time delay.

The handling of fractional time delays is not explicitly considered in this contribution. However, in the discrete time they can be approximated by an integer multiple of the sampling time and a slightly deviating dynamic impulse response. The maximum error of the time delay estimation resulting from this approximation is one sampling interval. By taking the deviating impulse response into account, such systems can also be treated in the methods presented below.

# 3 FIR models

As explained in Section 2 the output of a linear process y(k)without direct feedthrough can be calculated by

$$y(k) = \sum_{i=1}^{\infty} \tilde{g}(i)u(k-i) + v(k),$$
 (4)

where v(k) accounts for possibly existing measurement noise. With the assumption that the process is stable and v(k) is white Gaussian noise, an *n*th-order FIR model can be used to approximate this linear process by the linear combination of past inputs

$$\hat{y}(k) = \sum_{i=1}^{n} \theta(i) u(k-i) = \underline{x}(k)^{T} \underline{\theta},$$
(5)

where  $\underline{x}(k) = \begin{bmatrix} u(k-1) & u(k-2) & \dots & u(k-n) \end{bmatrix}^T$  with u(k) = 0, for  $k \le 0$  and  $\theta$  denoting the *n*-dimensional parameter vector. Here, the parameter vector represents the approximated finite impulse response coefficients of the infinite process impulse response including the time delay.

As can be seen in (5), the model is linear in its parameters and therefore can be estimated with LS [1]. The squared estimation error of the FIR model

$$J = \sum_{k=1}^{N} \left( y(k) - \hat{y}(k) \right)^2 = \|\underline{y} - \underline{X} \underline{\theta}\|_2^2$$
(6)

is minimized with respect to the model parameters  $\theta$ . Here, X denotes the  $(N \times n)$ -dimensional regressor matrix, y denotes the N-dimensional measured output vector with N being the number of data samples and  $\|\cdot\|_2$  denotes the Euclidean norm [2]. This property in combination with the output error configuration are main advantages of FIR models [2]. Because of the output error configuration the parameter estimation is unbiased (strictly speaking for  $n \rightarrow$  $\infty$ ) w.r.t. the noise assumption, which does not hold for e.g. ARX models. Additionally, a major benefit is that the FIR model estimation has a low sensitivity w.r.t. the model order n or preliminary unknown time delays d. Additionally, FIR models benefit from their inherent stability and therefore, they are also a preferable choice for online learning algorithms [20-22]. It only has to be ensured that the estimated impulse response is long enough to cover all coefficients of the true impulse response which significantly deviate from zero. However, this may lead to a high number of model coefficients n [1]. Therefore, especially for a small number of data samples, FIR models suffer from a high variance error which causes overfitting in many cases [2]. To overcome this, in [3] a regularization method has been introduced which decreases the variance error.

#### 3.1 Regularized FIR models

Due to its model structure, a large number of parameters is required to cover the dynamics of the given process properly. By regularization, the effective number of parameters can drastically be reduced. Therefore, a satisfying small variance error can be achieved. Thus, the major drawback of FIR models can be overcome. By introducing an additional penalty term in the estimation procedure, regularization is applied:

$$J = \sum_{k=1}^{N} \left( y(k) - \hat{y}(k) \right)^2 + \lambda \underline{\theta}^T \underline{R}(\underline{\eta}) \underline{\theta},$$
(7)

where *R* is the regularization matrix which is dependent on the hyperparameters  $\eta$ . The hyperparameter  $\lambda$  denotes the regularization strength. The second term is the regularization term which introduces a relationship between the parameters of the FIR model themselves. The linking of adjacent parameters is defined by the matrix R. Due to the fact that the parameters of an FIR model correspond to the impulse response coefficients of the modeled process, prior knowledge on the dynamic behavior can be incorporated in the matrix R. From the Bayesian perspective the model parameters  $\theta$  can be interpreted as a Gaussian process. The parameter vector  $\theta$  is assumed to be a random variable with a prior distribution of zero mean and the covariance matrix P [10], which is also known as the kernel matrix. It has to be noted that R corresponds to the inverse parameter covariance matrix  $P^{-1}$  of the prior. The hyperparameter  $\lambda$ is used to adjust the strength of the regularization. With a regularization strength of zero the prior knowledge in R has no influence on the estimated parameters. In contrast, for a high regularization strength almost only the prior is used for the parameter estimation. Furthermore, this approach preserves the analytical form of the solution from LS [2, 3]:

$$\underline{\hat{\theta}} = \left(\underline{X}^T \underline{X} + \lambda \underline{R}(\underline{\eta})\right)^{-1} \underline{X}^T \underline{y}.$$
(8)

#### 3.1.1 Choice of the regularization matrix

The choice of the regularization matrix <u>R</u> can strongly influence the parameter estimation of the FIR model. The most common regularization method is named ridge regression, where <u>R</u> = <u>I</u> [2]. Another regularization method contributed in [3] correlates the impulse response coefficients. Hereby, smoothness of the impulse response and knowledge about the exponential decay of the process is imposed. By decomposing the regularization term as

$$\underline{\theta}^{T}\underline{R}\,\underline{\theta} = \underline{\theta}^{T}\underline{F}^{T}\underline{F}\,\underline{\theta} = \|\underline{F}\,\underline{\theta}\|_{2}^{2},\tag{9}$$

the regularization matrix can be interpreted as a filter  $\underline{F}$  of the parameter vector [23]. It enables the incorporation of the prior knowledge directly in the filter matrix  $\underline{F}$ .

In [7] a method is contributed, which assumes a transfer behavior of the process under investigation. Typically, a linear  $n_{\rm IRP}$ th-order transfer function is used as prior from which only the denominator polynomial is needed. For the sake of convenience here  $n_{\rm IRP} = 2$  is chosen as an example

$$G(z) = \frac{B(z)}{a_0 + a_1 z^1 + z^2}.$$
 (10)

However, the extension to lower- and higher-order transfer functions is straightforward. With the coefficients  $\underline{a} = \begin{bmatrix} a_0 & a_1 \end{bmatrix}^T$ , the second-order impulse response preserving  $(\text{IRP}_2)^1$  filter matrix  $\underline{F}_{\text{IRP}}(\underline{a}) \in \mathbb{R}^{(n-2) \times n}$  is given for processes without direct feedthrough by [7]

$$\underline{F}_{\text{IRP}}(\underline{a}) = \begin{bmatrix} a_0 & a_1 & 1 & 0 & \dots & 0 \\ 0 & a_0 & a_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_0 & a_1 & 1 \end{bmatrix}.$$
 (11)

The approach to use only one regularization matrix, which corresponds from the Bayesian perspective to one kernel matrix, is also called single kernel method. In contrast, the authors of [11] propose a multiple kernels method. Here,  $n_{\rm MK}$  kernels  $\underline{R}_i$ ,  $i = 1, 2, ..., n_{\rm MK}$  are superimposed each with its own regularization strength  $\lambda_i$ .

#### 3.1.2 Hyperparameter optimization

To determine the hyperparameters, e.g. for the IRP kernel  $\underline{a}$  and  $\lambda$ , a hyperparameter optimization can be performed. Various criteria have been proposed to determine the model quality for a certain hyperparameter combination. Typically, these criteria are nonlinear and non-convex in  $\underline{a}$  and  $\lambda$ . In [8] the generalized cross-validation (GCV) error

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

with  $\underline{S} = \underline{X} (\underline{X}^T \underline{X} + \lambda \underline{R})^{-1} \underline{X}^T$  denoting the smoothing matrix, is contributed. The objective  $J_{GCV}$  has to be minimized to determine the optimal hyperparameters. From a Bayesian perspective, usually, the marginal likelihood (ML) function is used. For an easier calculation and better numerics, the log-likelihood

$$J_{\rm ML} = \underline{y}^T \underline{\Sigma}^{-1} \underline{y} + \text{logdet}(\underline{\Sigma}), \qquad (13)$$

with  $\underline{\Sigma} = \lambda \underline{X} \underline{R}^{-1} \underline{X}^T + \sigma^2 \underline{I}$  and  $\sigma^2$  denoting the variance of the process noise, is minimized [9].

# 3.2 Identification of time-delay systems with regularized FIR models

A major benefit of FIR models in general is the property that they are robust w.r.t. unknown time delays, because

**1** The index stands for the IRP order  $n_{\text{IRP}}$  corresponding to the order of the transfer function which is used as prior to build the filter matrix.



**Figure 2:** Model parameters  $\theta_j$  from the estimation of a second-order process with time delay with an unregularized (\_\_\_\_) and a regularized FIR model (\_\_\_\_).

each parameter can be independently set to zero. By regularization neighboring parameters are coupled and this advantage is weakened or lost. This coupling leads to undesired behavior, because only after the time delay, the model parameters should deviate from zero. Thereby, regularization causes a blurring of the parameter vector, which normally changes suddenly after the time delay has passed. This behavior can be seen in Figure 2. It causes that the regularization strength can no longer be chosen high, because the regularization penalty term does not fit to the process anymore. To overcome this issue, the regularization matrix (11) has to be modified to also allow the incorporation of time delays.

# 4 Incorporation of time delay in regularization matrix

In this chapter a modified regularization matrix is proposed which allows the incorporation of time delays. Additionally, two methods – a single kernel and a multiple kernel method – are developed applying this kernel on the identification of processes with unknown time delays. These methods were developed for the single input single output (SISO) case. However, it is also possible to extend them to the multi input single output (MISO) case.

#### 4.1 Time-delay kernel

From the Bayesian perspective the inverse of the regularization matrix  $\underline{P} = \underline{R}^{-1}$  can be interpreted as the covariance matrix of a Gaussian prior with zero mean. Partitioning the prior of the parameters into two parts ('d' and 'dyn') leads to

$$p(\underline{\theta}) = \mathcal{N}\left(\underbrace{\underbrace{\begin{bmatrix}\underline{0}\\\underline{0}\end{bmatrix}}}_{\underline{\mu}}, \underbrace{\begin{bmatrix}\underline{P}_{d} & \underline{0}\\\underline{0} & \underline{P}_{dyn}\end{bmatrix}}_{\underline{P}}\right).$$
(14)

The zero off-diagonal elements account for the fact that parameters corresponding to the time delay and parameters corresponding to the dynamic behavior are not correlated with each other. Additionally, two different covariance matrices  $\underline{P}_{d}$  and  $\underline{P}_{dyn}$  represent the two parameter groups.

For the second part ( $\underline{P}_{dyn}$ ) it can clearly be argued that all assumptions on the undelayed system, such as exponential decay and smoothness in the TC kernel case or impulse response preservation for the IRP kernel, still hold. For the first part ( $\underline{P}_d$ ), however, these assumptions do not hold. So the questions arises, how to build the covariance matrix  $\underline{P}_d$ for the time-delay parameters.

A natural choice can be made from the fact, that all time-delay parameters should be exactly zero. Due to the absolute certainty of this condition, all variances of the parameter prior, which populate the diagonal elements of the covariance matrix, can be set to zero. Additionally, no covariance between the parameters exist which results in zero entries at the off-diagonal elements in  $\underline{P}_d$  as well.

The goal of the identification from the Bayesian perspective is to find the parameter distribution conditioned on the measured outputs. The mean of this distribution corresponds to the regularized LS solution [3]

$$\underline{\hat{\theta}} = \underline{P} \underline{X}^{T} (\underline{X} \underline{P} \underline{X}^{T} + \sigma^{2} \underline{I}_{N})^{-1} \underline{y}.$$
(15)

By splitting the parameter vector

$$\underline{\theta} = \begin{bmatrix} \underline{\theta}_{\rm d} \\ \underline{\theta}_{\rm dyn} \end{bmatrix}$$
(16)

as well as the regressor matrix

$$\underline{X} = \begin{bmatrix} \underline{X}_{d} & \underline{X}_{dyn} \end{bmatrix}$$
(17)

into a time-delay and a dynamic part and including this into the mean of the parameter distribution we get

$$\frac{\hat{\theta}}{\underline{\theta}} = \begin{bmatrix} \underline{P}_{d} \underline{X}_{d}^{T} \\ \underline{P}_{dyn} \underline{X}_{dyn}^{T} \end{bmatrix} \left( \underline{X}_{d} \underline{P}_{d} \underline{X}_{d}^{T} + \underline{X}_{dyn} \underline{P}_{dyn} \underline{X}_{dyn}^{T} + \sigma^{2} \underline{I}_{N} \right)^{-1} \underline{y}.$$
(18)

With the previous choice of  $\underline{P}_{d} = \underline{0}$ , from (18) it can be reasoned that  $\underline{\theta}_{d} = \underline{0}$  since  $\underline{P}_{d} \underline{X}_{d}^{T} = \underline{0}$ . Additionally, due to  $\underline{X}_{d} \underline{P}_{d} \underline{X}_{d}^{T} = \underline{0}$  the regressor part  $\underline{X}_{d}$  is not influencing the estimation of the dynamic parameters  $\underline{\theta}_{dyn}$ . Which is intended by the choice of  $\underline{P}_{d}$ .

From the implementation perspective as well as the filter perspective however, a choice of  $\underline{P}_{d} = \underline{0}$  is questionable. For the filter interpretation the inverse of the covariance matrix

$$\underline{R} = \underline{P}^{-1} = \begin{bmatrix} \underline{P}_{d}^{-1} & \underline{0} \\ \underline{0} & \underline{P}_{dyn}^{-1} \end{bmatrix}$$
(19)

is needed. This inversion is not possible with a zeros matrix for  $\underline{P}_d$  since it leads to a rank-deficient matrix  $\underline{P}$ . Also for small values of  $\sigma^2$  the inversion in (18) is poorly conditioned. This problem is also known from the IRP kernel and can be overcome executing a SVD, for details see [24].

A second possibility is exploiting an approximation, that preserves the desired property of forcing all parameters corresponding to the time delay to zero. It is given by

$$\underline{P}_{d}(\beta) = \beta \underline{I}_{n}.$$
(20)

With  $\beta$  the variance of the prior of all time-delay parameters can be adjusted. Compared to the previously explained special case  $\beta = 0$ , increasing  $\beta$  results in higher variance of the parameter prior. Still, with small values for  $\beta$  similar results can be achieved as with  $\beta = 0$ . This choice coincidences with the well-known ridge regression but only for the parameters describing the time delay.

For the filter perspective of the kernel the inverse of  $\underline{P}_d$  is needed which can be calculated by

$$\underline{R}_{\mathrm{d}}(\beta) = \underline{P}_{\mathrm{d}}^{-1}(\beta) = \frac{1}{\beta} I_{\mathrm{n}_{\mathrm{d}}}.$$
(21)

This results in an additional penalty term in the LS problem for  $\hat{\theta} = \arg \min_{\theta} J$  with

$$J = \|\underline{y} - \underline{X} \underline{\theta}\|_{2}^{2} + \|\lambda \frac{1}{\sqrt{\beta}} \underline{\theta}_{d}\|_{2}^{2} + \|\lambda \underline{F}_{dyn} \underline{\theta}_{dyn}\|_{2}^{2}$$
$$= \|\underline{y} - \underline{X} \underline{\theta}\|_{2}^{2} + \|\lambda \begin{bmatrix} \frac{1}{\sqrt{\beta}} \underline{I}_{n_{d}} & \underline{0} \\ \underline{0} & \underline{F}_{dyn} \end{bmatrix} \begin{bmatrix} \underline{\theta}_{d} \\ \underline{\theta}_{dyn} \end{bmatrix} \|_{2}^{2}.$$
(22)

As explained in Section 3.1, we intend using the IRP kernel for the dynamic parameters. With the time-delay extension we propose the novel kernel named *impulse response and time-delay preserving* (IRDP) kernel which leads to the following regularization matrix

$$\underline{R}_{\mathrm{IRDP}}(n_{\mathrm{d}}, \underline{a}, \beta) = \begin{bmatrix} \underline{R}_{\mathrm{d}}(\beta) & \underline{0} \\ \underline{0} & \underline{R}_{\mathrm{IRP}}(\underline{a}) \end{bmatrix}$$
(23)

with  $\underline{R}_{d} \in \mathbb{R}^{n_{d} \times n_{d}}$  and  $\underline{R}_{IRDP} \in \mathbb{R}^{n \times n}$ .

While deriving the time-delay kernel, the assumption of exactly known time delay is made. This, however, in most practical situations does not hold. Therefore, a mechanism has to be developed identifying the time delay along with the FIR parameters, to properly build the time-delay kernel. This problem is addressed in the following sections and solved in two different ways.

#### 4.2 Single kernel method

The first approach for identifying RFIR models with unknown time delays is based on a single kernel method. As explained in Section 4.1 the incorporation of the time delay into the regularization matrix is done by strict distinction between the time-delay and the dynamics part. Thus, choosing the point of separation  $n_d$  in the regularization matrix is an integer optimization problem. Solving it along with the other real hyperparameters ( $\underline{a}$ ,  $\lambda$ ) results in a mixed-integer optimization problem which usually is not easy to solve.

Therefore, we modify the kernel by fuzzifying the separation. First, we build two separate regularization matrices of size  $n \times n$ . The first purely consists of the time-delay kernel regularization matrix. The second is the regular dynamic kernel regularization matrix, which can be decomposed into the filter matrix as shown in (9). Then both kernels are superimposed in a weighted manner. A sigmoid function and its remainder to one are used for this purpose. The combined regularization matrix is given by

$$\underline{R}_{SK} = \operatorname{diag}(\underline{\mu})\underline{R}_{d} + \underline{F}_{dyn}^{T}\operatorname{diag}(\underline{1} - \underline{\mu})\underline{F}_{dyn}, \qquad (24)$$

with  $\underline{R}_{d} \in \mathbb{R}^{n \times n}$ . It has to be noted that the rows of  $\underline{F}_{dyn}$  are weighted with diag( $\sqrt{1-\mu}$ ) and therefore, the model parameters in the corresponding rows with zero weighting are not penalized. The values for  $\underline{\mu}$  are calculated by evaluating the sigmoid function

$$sig(x, y, z) = \frac{1}{1 + exp((x - y)z)}$$
 (25)

at different query points. Here *y* defines the inflection point of the sigmoid and *z* the slope. The query points are chosen corresponding to the number of FIR parameters:

$$\underline{\mu} = \left[\operatorname{sig}(1, \hat{d}, \sigma_{\mathrm{d}}), \operatorname{sig}(2, \hat{d}, \sigma_{\mathrm{d}}), \dots, \operatorname{sig}(n, \hat{d}, \sigma_{\mathrm{d}})\right]^{T}.$$
 (26)

Figure 3 visualizes this approach. With it, the mixedinteger optimization problem of the time-delay selection is changed into a continuous optimization problem w.r.t. the hyperparameter  $\hat{d}$ . This may result in a slightly different solution, but it is much easier to solve. Additionally, the fuzzification strength can be adjusted by the second hyperparameter  $\sigma_d$ . By increasing  $\sigma_d$  the overlapping region of both regularization matrices is reduced. It is worth noticing that the case of strict distinction between both regularization matrices explained in Section 4.1 is included in this optimization problem for  $\sigma_d \rightarrow \infty$ .

Now, the hyperparameter optimization from Section 3.1 can be utilized to optimize  $\hat{d}$  and  $\sigma_d$  along with the other kernel specific hyperparameters. An intermediate version of this is approach is achieved by presetting  $\sigma_d$  and only



Figure 3: Sigmoid functions for fuzzification of kernel superim-position.

optimizing  $\hat{d}$  in the hyperparameter optimization. To distinguish between these two versions, in the following we call the full optimization SK<sub>2</sub> and the intermediate optimization SK<sub>1</sub>.

#### 4.3 Multiple kernel method

The second method to incorporate time delay into the procedure of identifying RFIR models is based on a multiple kernel method. Here, multiple time-delay kernels are used for the identification of time-delay processes. Thereby, the mixed-integer optimization problem discussed in Section 4.2 can be overcome. The time-delay kernel  $\underline{R}_{\text{IRDP}}$  in Section 4.1 with different time delays  $\underline{\delta} = [d_{\min}, d_{\min} + 1, \dots, d_{\max}]$  is used. The number of  $n_{\text{MK}} = d_{\max} - d_{\min} + 1$  regularization matrices  $\underline{R}_{\text{IRDP}}$  are each multiplied with a separate  $\lambda_i$  (all gathered in  $\underline{\lambda}$ ) and summed up

$$\underline{R}_{\mathrm{MK}}(\underline{\lambda},\underline{a}) = \sum_{i=1}^{n_{\mathrm{MK}}} \lambda_i \underline{R}_{\mathrm{IRDP}}(\delta_i,\underline{a},\beta).$$
(27)

The considered time delays and therefore also the number of multiple kernels  $n_{\rm MK}$  as well as the parameter  $\beta$  have to be set *a priori*. The model parameters can be calculated with

$$\underline{\hat{\theta}} = \left(\underline{X}^T \underline{X} + \underline{R}_{\mathrm{MK}}(\underline{\lambda}, \underline{a})\right)^{-1} \underline{X}^T \underline{y}.$$
(28)

In comparison to (8), a vector of  $\underline{\lambda}$  is used instead of a scalar  $\lambda$ . This leads to more hyperparameters. To determine them, a hyperparameter optimization as described in Section 3.1.2 is performed. For the calculation of the scalar estimated time delay  $\hat{d}$  the time delays of the different kernels are weighted with the corresponding regularization strengths  $\lambda_i$ 

$$\hat{d} = \frac{\sum_{i=1}^{n_{\text{MK}}} \lambda_i \delta_i}{\sum_{i=1}^{n_{\text{MK}}} \lambda_i}.$$
(29)

#### 4.4 Comparison of the methods

In this section a kernel, which incorporates time delay in the structural assumptions of the regularization matrix,

**Table 1:** Comparison of the methods MK,  $SK_1$ , and  $SK_2$  regarding the used regularization matrix as well as the to be optimized hyperparameter.

Method	МК	SK <sub>1</sub>	SK <sub>2</sub>
Regularization matrix	<u></u> Я <sub>МК</sub> (Еq. (27))	<u>R<sub>sk</sub></u> (Eq. (24))	<u> Я</u> <sub>sк</sub> (Eq. (24))
Hyperparameter	<u>λ</u> , <u>a</u>	λ, <u>a</u> , <i>â</i>	$\lambda, \underline{a}, \hat{d}, \sigma_d$

was proposed. To include an unknown time delay in the RFIR estimation and avoiding a mixed-integer optimization problem, two methods were developed. The complexity of the utilized hyperparameter optimization increases by the proposed methods. Compared to the estimation with the regular IRP kernel, method SK1 requires one additional hyperparameter (inflection point) and SK<sub>2</sub> requires two additional hyperparameters (inflection point and slope). The number of hyperparameters required for the MK method depends on the number of kernels which has to be chosen a priori. This requires the selection of a time interval the time delay is expected in. The higher the uncertainty of the unknown time delay, the more hyperparameters have to be optimized. Therefore, the MK method requires  $n_{\rm MK} - 1$ more hyperparameters than the estimation with the regular IRP kernel. The kernels as well as the hyperparameters used in the different methods are summarized in Table 1.

### **5** Results

To investigate the methods proposed in Section 4, first, a parameter study for  $\beta$  is performed and the influence of the two hyperparameter optimization criteria on the model quality is evaluated. Finally, a simulation study on a wide range of different processes is performed to compare the proposed approaches.

#### 5.1 Simulation setup

To investigate the different methods regarding robustness of the time-delay estimation and the model quality, different Monte Carlo studies with 100 runs are performed. In Table 2 three different process types are defined. The pole(s) of the different processes are chosen uniformly distributed in the given range for each run. The range is determined to cover typically present poles with a well chosen sample time.

White Gaussian noise with N = 1000 data samples is used as excitation signal. In each run other noise realizations are used for the excitation. White Gaussian noise is added to the output data such that a signal-to-noise-ratio

 $G_{3}(z^{-1})$ 

 $10^{-3}/10^{-1}$ 

10-3/10-0

10^3/10^0

 $G_2(z^{-1})$ 

 $10^{-3}/10^{-1}$ 

10-3/10-0 10^3/10^0

$G_1(z^{-1}) =$	$G_2(z^{-1}) = \frac{1}{z^{-1}z^{-d}}$	$G_3(z^{-1}) = \frac{1}{z^{-1}z^{-d}}$	best $eta$
$\frac{z-z}{1-az^{-1}}$	$\frac{z}{(1-az^{-1})(1-bz^{-1})}$	$\frac{2 - 2}{(1 - az^{-1})(1 - bz^{-1})(1 - cz^{-1})}$	IRDP <sub>1</sub>
$a \in [0.7, 0.92]$ $d \in [5, 20]$	$a, b \in [0.7, 0.92]$ $d \in [5, 20]$	$a, b, c \in [0.7, 0.92]$ $d \in [5, 20]$	IRDP <sub>2</sub> IRDP <sub>3</sub>

Table 2: Three different processes types: a first-order, a second-order, and a third-order process.

**Table 3:** Choice of  $\beta$  corresponding to the best fit for the methods  $SK_2/MK$  with the  $IRDP_{n_{IRP}}$  kernel.

 $G_1(z^{-1})$ 

 $10^{-1}/10^{-1}$ 

10<sup>-1</sup>/10<sup>-0</sup>

 $10^{-1}/10^{-1}$ 

(SNR) of 10 dB results. The FIR model order <i>n</i> is set to 100 in order to capture all significant impulse response coefficients. For the MK method $n_{\rm MK} = 21$ kernels with $\underline{\delta} = [5, 6, \dots, 25]$ are used. To evaluate the model quality, the fit between the true impulse response coefficients $\underline{\theta}^*$ , which we assume to be the first <i>n</i> coefficients of $\underline{\tilde{g}}$ , and the esti-	Exemplary, in t fit calculated with realizations for sec choices on $\beta$ and t of the fit values sho given process is rol
we assume to be the first <i>n</i> coefficients of $\underline{\tilde{g}}$ , and the esti-	given process is rol
mated model parameters $\hat{ heta}$ is calculated by using	estimation becomes

$$\operatorname{fit}\left(\underline{\theta}^{*},\underline{\hat{\theta}}\right) = 100 \,\% \left(1 - \frac{\|\underline{\theta}^{*} - \underline{\hat{\theta}}\|_{2}}{\|\underline{\theta}^{*}\|_{2}}\right). \tag{30}$$

#### 5.2 Kernel specific hyperparameter choice

The time-delay kernel explained in Section 4.1 contains the hyperparameter  $\beta$  which accounts for numerical stability in the calculation of the parameters  $\hat{ heta}.$  To make a reasonable choice on this parameter, a Monte Carlo study as described in Section 5.1 for each process order from Table 2 is performed. The parameter  $\beta$  is varied in the heuristically chosen range of

$$\beta = 10^{-i} \quad \forall \quad i \in \{0, 1, 2, \dots, 7\}.$$
(31)

The order of the IRDP kernel, which is part of the timedelay kernel, is varied from first to third order.

For the SK<sub>1</sub> method the results show, that the mean model fit is not sensitive w.r.t. the choice of  $\beta$ . The same holds for the estimation accuracy of the time-delay. However, it could be observed that the smaller  $\beta$  is chosen the more outlier are present.

Table 3 summarizes the results of the investigation on the SK<sub>2</sub> and MK method w.r.t. the choice of  $\beta$ . It can be seen that the optimal choice of  $\beta$  for the SK<sub>2</sub> method is dependent on the process order. For first-order processes  $G_1(z^{-1})$  higher values for β yield the best results both, for fit and accuracy of the time-delay estimation. For higher-order processes such as  $G_2(z^{-1})$  and  $G_3(z^{-1})$  lower values for  $\beta$  show better results. Against that, the order of the IRDP kernel does not influence the best choice of  $\beta$  which can be seen from identical best values for  $\beta$  in each column.

he upper part of Figure 4 the parameter (30) is shown for 100 input and noise cond-order processes, for the different the SK<sub>2</sub> IRDP<sub>2</sub> kernel. The distribution ows that the estimation method for the bust for  $\beta \ge 10^{-4}$ . For lower values the less robust which can be substantiated by the increasing number of outlier models. Similar results are also observed for the other process orders.

The MK method performs almost independently from the process order and the order of the IRDP kernel and yields the best results with high values for  $\beta$  such as  $10^{-1}$ or 10°, respectively. From the lower part of Figure 4 it can be reasoned that the robustness and performance is significantly reduced for  $\beta \leq 10^{-2}$ . This behavior can also be observed for the other process and IRDP kernel order combinations.

To summarize the results, we advise to choose  $\beta =$ 10° for the SK<sub>1</sub> method. A choice of  $\beta$  dependent on the



**Figure 4:** Boxplot of parameter fit for different choices of  $\beta$  for second-order processes and the methods SK<sub>2</sub> and MK with the IRDP<sub>2</sub> kernel for 100 realizations.

process is advised for the SK<sub>2</sub> method. If this knowledge is not available,  $\beta = 10^{-1}$  achieves a good trade-off between robustness and performance of the estimation. For the MK method choosing  $\beta = 10^{-1}$  generally yields good estimation performance and robustness simultaneously and is recommended.

#### 5.3 Criteria for hyperparameter optimization

The choice of the criterion for the hyperparameter optimization is crucial, since the model quality strongly depends on the hyperparameter. Therefore, the marginal likelihood (ML) function and generalized cross-validation (GCV) error explained in Section 3.1.2 are investigated with a Monte Carlo study described in Section 5.1 and the resultant parameter fit as well as the time-delay estimation are compared. For this purpose only the results of the second-order process  $G_2(z^{-1})$  from Table 2 are shown, since the results of the other process orders are similar.

For the investigation of the different hyperparameter optimization criteria an interior-point optimization algorithm is utilized. It is chosen based on a comparison of different optimization algorithms, in which it outperformed the competitive algorithms.

In Figure 5 a boxplot of the parameter fit and the timedelay estimation for the three different methods (SK<sub>1</sub>, SK<sub>2</sub>, and MK) with the IRDP<sub>2</sub> kernel is shown. The parameter  $\beta$ is chosen according Table 3. Here, both the mean and the variance of the parameter fit are significantly better using the GCV error instead of ML for optimization. Similar to the fit, the GCV error outperforms the ML and consequently is the preferred choice for the hyperparameter optimization objective. For the multiple kernel (MK) approach with ML, it can be seen that there is a mean error of -0.5 for the time-delay estimation. This means that by using ML in the optimization not only one kernel is weighted highly, but two consecutive kernels with a similar regularization strength are chosen. This undesirable choice also leads to a worse parameter fit.

In summary, for all methods the models with GCV error optimized hyperparameters show better and more robust performance. Therefore, in the following only the GCV error is used for hyperparameter optimization. This result deviates from previous investigations on the TC or IRP kernel for linear dynamic system without time-delay. In that case the ML objective usually outperforms the GCV objective [25, 26].

#### 5.4 Comparison of all proposed methods

In this section, all proposed methods for identifying timedelay processes (SK<sub>1</sub>, SK<sub>2</sub>, MK) are compared. To show the improvement of the novel methods developed for timedelay processes, also unregularized FIR models and RFIR models with the regular IRP kernel are compared. The order of the IRP and IRDP kernel, respectively are chosen identically to the order of the process. The hyperparameter optimization is performed with the GCV error and  $\beta$  is set according to Table 3. All three process orders of Table 2 are investigated. A Monte Carlo simulation with the setup of Section 5.1 is performed and the parameter fit as well as the deviation of the estimated time delay  $\hat{d}$  to the true time delay d is compared. It has to be noted, that for unregularized FIR models as well as for RFIR models with the regular IRP kernel no specific time-delay estimation procedure is carried out and therefore no explicit information can be gained.



**Figure 5:** Boxplot of parameter fit and time-delay estimation for the three different methods  $SK_1$ ,  $SK_2$ , and MK with the  $IRDP_2$  kernel for 100 realizations of  $G_2(z^{-1})$ . Here, the two criteria – generalized cross-validation (GCV) error and the marginal likelihood (ML) function are compared.

shown. It can be seen that unregularized FIR models and

In Figure 6 the investigation of first-order processes is



**Figure 6:** Boxplot of parameter fit and time-delay estimation for the five different methods FIR, IRP, SK<sub>1</sub>, SK<sub>2</sub>, and MK with the IRP<sub>1</sub> and IRDP<sub>1</sub> kernel for 100 realizations of  $G_1(z^{-1})$ . Note: There is no time-delay estimation for the methods FIR and IRP.

RFIR models with IRP kernel show a similar performance. Due to the inappropriate regularization matrix, the regularization strength has been chosen low. The specifically developed methods for time-delay processes show a significant improvement of the estimation. Especially, the multiple kernel (MK) method outperforms the other methods in the parameter fit as well as in the accuracy of the timedelay estimation. For the time-delay estimation  $SK_2$  shows a smaller 75 % quantile, but yields more outlier models in comparison to  $SK_1$ .

The results for a second-order process are shown in Figure 7. It again shows the superiority of the MK method. For second-order processes the performance of the IRP method is superior in comparison to the unregularized FIR models. This can be explained by the smoothness assumption of the IRP kernel since the impulse response of a first-order process is not continuous. However, the impulse response of a second-order process is continuous and therefore the smoothness assumption is not violated. Consequently, this allows to choose the regularization strength  $\lambda$  higher. For this process order, the method with variable slope SK<sub>2</sub> obtains better results than with a fixed slope. Here, all deviations of the time-delay estimation of the MK method lie in a range of 0– and –1.

For third-order processes the results of the parameter fit as well as the time-delay estimation are similar to the second-order process. Only the time-delay estimation of method  $SK_1$  is slightly inferior and is not as robust as for the second-order process.

In Figure 8 estimated impulse responses of a secondorder process  $G_2(z^{-1})$  with the fixed poles a = 0.9, b = 0.7, and the time delay d = 10 are shown. The process is excited and disturbed as explained in Section 5.1. The parameter variance within the different sets of FIR models corresponding to the chosen estimation methods show the superiority



**Figure 7:** Boxplot of parameter fit and time-delay estimation for the five different methods FIR, IRP, SK<sub>1</sub>, SK<sub>2</sub>, and MK with the IRP<sub>2</sub> and IRDP<sub>2</sub> kernel for 100 realizations of  $G_2(z^{-1})$ .



**Figure 8:** Impulse responses of a second-order process  $G_2(z^{-1})$  with poles a = 0.9, b = 0.7, and time delay d = 10 for 50 different noise realizations estimated with the methods FIR, IRP, SK<sub>2</sub>, and MK with the IRP<sub>2</sub> and IRDP<sub>2</sub> kernel.

of the novel methods. Especially in the time-delay part of the impulse response almost every deviation from zero can be suppressed. Additionally, improvements in the smoothness are also visible in the dynamics part. Since the results of the methods  $SK_1$  and  $SK_2$  are similar, only  $SK_2$  is shown for improved visualization. Moreover, the MK method outperforms the  $SK_2$ .

In summary, the multiple kernel (MK) method outperforms all other methods in both, robustness and performance of the estimation. However, the number of hyperparameters increases with the number of time delays assumed a priori. Nevertheless, the hyperparameter optimization achieves a sparse configuration of  $\underline{\lambda}$  even with  $n_{\rm MK} = 21$ , i.e., most  $\lambda_i = 0$ , typically only one  $\lambda_i \neq 0$ , for  $i = 1, 2, ..., n_{MK}$ . Moreover, the incorporation of the time delay in the structure of the kernel significantly improves the model estimation of time-delay processes in comparison to the standard techniques. The methods SK<sub>1</sub> and SK<sub>2</sub> show similar model qualities, although for second- or third-order processes optimizing the slope of the sigmoids improves the estimation performance. In contrast to regular regularization methods like the TC or IRP kernel, for the proposed methods the GCV error should be used as criteria for the hyperparameter optimization instead of ML to get the best model performance.

## 6 Conclusions

We proposed novel kernel techniques which incorporate knowledge on the time delay to estimate regularized FIR models. Two methods have been developed to estimate the unknown time delay by using a single kernel as well as a multiple kernel approach. Two studies were performed to obtain the best parameterization and optimization criteria for the hyperparameter optimization. It is shown that as criteria for the hyperparameter optimization generalized cross-validation (GCV) error should be used instead of the marginal likelihood (ML) function. The multiple kernel method is the most robust method and achieves the best model quality. However, the number of hyperparameters increases with the assumed range the time delay possibly lie in. Also the single kernel methods significantly improve the model estimation of time-delay processes, while the number of hyperparameters and thus the complexity of the optimization remains low.

Further research will be done on extending this concept to nonlinear time-delay processes by using linear local model networks.

**Author contributions:** All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

**Conflict of interest statement:** The authors declare no conflicts of interest regarding this article.

## References

- L. Ljung, System Identification: Theory for the User, 2nd ed. New Jersey, Prentice Hall PTR, 1999.
- [2] O. Nelles, Nonlinear System Identification from Classical Approaches to Neural Networks, Fuzzy Systems, and Gaussian Processes. From Classical Approaches to Neural Networks, Fuzzy Systems, and Gaussian Processes, 2nd ed. Cham, Springer International Publishing AG, 2020.
- [3] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: a survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.
- [4] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, *Regularized System Identification. Learning Dynamic Models from Data*, Cham, Springer International Publishing AG, 2022.
- [5] G. Pillonetto, A. Chiuso, and G. De Nicolao, "Regularized estimation of sums of exponentials in spaces generated by stable spline kernels," in 2010 American Control Conference, IEEE, 2010.
- [6] T. Chen, H. Ohlsson, G. Goodwin, and L. Ljung, "Kernel selection in linear system identification Part Ii: a classical perspective," in 2011 50th IEEE Conference on Decision and Control and European Control Conference, 2011.
- [7] T. Münker, T. J. Peter, and O. Nelles, "Gray-box identification with regularized FIR models," *Automatisierungstechnik*, vol. 66, no. 9, pp. 704–713, 2018.
- [8] 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, 1979.
- [9] G. Pillonetto and A. Chiuso, "Tuning complexity in regularized kernel-based regression and linear system identification:

the robustness of the marginal likelihood estimator," *Automatica*, vol. 58, pp. 106–117, 2015.

- [10] G. Pillonetto and G. De Nicolao, "A new kernel-based approach for linear system identification," *Automatica*, vol. 46, no. 1, pp. 81–93, 2010.
- [11] 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 Trans. Autom. Control*, vol. 59, no. 11, pp. 2933–2945, 2014.
- [12] X. Chen, Z. Mao, and R. Jia, "A new multiple kernel-based regularization method for identification of delay linear dynamic systems," *Chemom. Intell. Lab. Syst.*, vol. 199, p. 103971, 2020.
- [13] S. Björklund, "A survey and comparison of time-delay estimation methods in linear systems," Linköping Studies in Science and Technology. Thesis No. 1061, Sweden, Linköping University, 2003.
- [14] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 525–533, 1993.
- [15] W.-X. Zheng and C.-B. Feng, "Identification of stochastic time lag systems in the presence of colored noise," *Automatica*, vol. 26, no. 4, pp. 769–779, 1990.
- [16] M. Agarwal and C. Canudas, "On-line estimation of time delay and continuous-time process parameters," in *1986 American Control Conference*, 1986, pp. 728–733.
- [17] Q.-G. Wang and Y. Zhang, "Robust identification of continuous systems with dead-time from step responses," *Automatica*, vol. 37, no. 3, pp. 377–390, 2001.
- [18] S. Ahmed, B. Huang, and S. L. Shah, "Parameter and delay estimation of continuous-time models using a linear filter," *J. Process Control*, vol. 16, no. 4, pp. 323–331, 2006.
- [19] S. Bedoui, M. Ltaief, and K. Abderrahim, "New results on discrete-time delay systems identification"," *Int. J. Autom. Comput.*, vol. 9, no. 6, pp. 570–577, 2012.
- [20] C. Illg and O. Nelles, "Adaptive model predictive control with regularized finite impulse response models," in *ATHENA Research Book*, vol. 1, 2022, pp. 327–331.
- [21] C. Illg, T. Decker, J. Thielmann, and O. Nelles, "Adaptive model predictive control with finite impulse response models," in *Proceedings – 31th Workshop Computational Intelligence*, 2021, pp. 149–168.
- [22] C. Illg, T. Kösters, S. Kiroriwal, and O. Nelles, "Adaptive system identification with regularized FIR models," *IFAC-PapersOnLine*, vol. 55, no. 12, pp. 1–6, 2022.
- [23] A. Marconato, M. Schoukens, and J. Schoukens, "Filter-based regularisation for impulse response modelling," *IET Control. Theory Appl.*, vol. 11, no. 2, pp. 194–204, 2017.
- [24] T. Chen and L. Ljung, "Implementation of algorithms for tuning parameters in regularized least squares problems in system identification," *Automatica*, vol. 49, no. 7, pp. 2213–2220, 2013.
- [25] B. Mu, T. Chen, and L. Ljung, "Asymptotic properties of generalized cross validation estimators for regularized system identification," *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 203–208, 2018.
- [26] B. Mu, T. Chen, and L. Ljung, "On asymptotic properties of hyperparameter estimators for kernel-based regularization methods," *Automatica*, vol. 94, pp. 381–395, 2018.

# **Bionotes**



#### Tarek Kösters

Automatic Control – Mechatronics, Institute of Mechanics and Control Engineering – Mechatronics, University of Siegen, Paul-Bonatz-Str. 9-11, 57068 Siegen, Germany tarek.koesters@uni-siegen.de

Tarek Kösters has been a research assistant with the working group Automatic Control — Mechatronics of Prof. Nelles since November 2020. In his research work he focuses on linear and nonlinear system identification in the offline and online case. He graduated with a Master of Science degree from Otto-von-Guericke Universität Magdeburg in 2020.



#### **Christopher Illg**

Automatic Control – Mechatronics, Institute of Mechanics and Control Engineering – Mechatronics, University of Siegen, Paul-Bonatz-Str. 9-11, 57068 Siegen, Germany **christopher.illg@uni-siegen.de** 

Christopher Illg graduated with a Master of Science degree from Karlsruhe Institute of Technology in 2020. He has joined the working group Automatic Control— Mechatronics of Prof. Nelles as a research assistant. His research topics focus on linear and nonlinear system identification, adaptive system identification and model predictive control.



#### **Oliver Nelles**

Automatic Control – Mechatronics, Institute of Mechanics and Control Engineering – Mechatronics, University of Siegen, Paul-Bonatz-Str. 9-11, 57068 Siegen, Germany oliver.nelles@uni-siegen.de

Oliver Nelles is Professor at the University of Siegen in the Department of Mechanical Engineering and chair of Automatic Control — Mechatronics. He received his doctor's degree in 1999 at the Technical University of Darmstadt. His key research topics are nonlinear system identification, dynamics representations, design of experiments, metamodeling and local model networks.