Vrije Universiteit Brussel

FRF estimation using multiple kernel-based regularisation

Hallemans, Noel; Pintelon, Rik; Joukovsky, Boris Joseph; Peumans, Dries; Lataire, John

Published in:
Automatica

Publication date:
2022

License:
CC BY-NC-ND

Document Version:
Accepted author manuscript

Link to publication

Citation for published version (APA):
Hallemans, N., Pintelon, R., Joukovsky, B. J., Peumans, D., & Lataire, J. (2022). FRF estimation using multiple kernel-based regularisation. *Automatica*, *136*(2), [110056]. https://doi.org/10.1016/j.automatica.2021.110056

Copyright
No part of this publication may be reproduced or transmitted in any form, without the prior written permission of the author(s) or other rights holders to whom publication rights have been transferred, unless permitted by a license attached to the publication (a Creative Commons license or other), or unless exceptions to copyright law apply.

Take down policy
If you believe that this document infringes your copyright or other rights, please contact openaccess@vub.be, with details of the nature of the infringement. We will investigate the claim and if justified, we will take the appropriate steps.

Download date: 20. Apr. 2024

Automatica 136 (2022) 110056

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

FRF estimation using multiple kernel-based regularisation ☆

Noël Hallemans *, Rik Pintelon, Boris Joukovsky, Dries Peumans, John Lataire

Vrije Universiteit Brussel, Brussels, Belgium

ARTICLE INFO

Article history:
Received 4 March 2021
Received in revised form 25 August 2021
Accepted 6 October 2021
Available online xxxx

Keywords:
Data-driven modelling
Kernel-based modelling
Local rational modelling
Gaussian process regression
Machine learning
Lightly damped systems

ABSTRACT

Nonparametric kernel-based modelling of dynamical systems offers important advantages over other nonparametric techniques; the estimate is a continuous function, the model complexity is continuously tuneable, and stability, causality and smoothness are imposed on the impulse response estimate. However, for lightly damped systems, most of the existing kernel-based approaches for estimating the impulse- or frequency response function fail because classical kernels are not appropriate for describing lowly damped resonances. Smoothness is imposed on the entire frequency axis with the diagonal correlated or stable spline kernel, with as a result that resonances cannot be captured well. By introducing the superposition of different kernels, carrying prior knowledge about the resonant poles of the system, we make the kernel-based modelling of lightly-damped systems possible with high-accuracy. We use a frequency domain local rational modelling technique as preprocessing step to determine the most dominant poles, and include these as prior knowledge in the kernels. The performance of the new kernel is demonstrated on a highly resonating simulated system and compared to the state of the art nonparametric frequency domain approaches.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Data-driven modelling builds models for dynamical systems starting from noisy data. This paper studies exclusively the class of linear time-invariant (LTI) systems, which is well documented in the literature (Kwakernaak & Sivan, 1991), and convenient to work with.

When modelling a system, distinction is made between parametric and nonparametric models. In the former, the number of parameters is fixed, while in the latter it is proportional to the amount of data. In this paper we build an estimator for a nonparametric model, which is most of the time used as a preprocessing step for determining the complexity and for validation of the parametric modelling step. Notwithstanding, nonparametric models also have their stand-alone applications. They have the key advantage that no model structure ought to be chosen, and hence, as opposed to parametric models, are able to represent infinite dimensional systems. Moreover, nonparametric models give quick insight about the behaviour of the system, and certain control applications, such as the iterative feedback tuning (Hjalmarsson, Gevers, Gunnarsson, & Lequin, 1998), are directly applicable on nonparametric models.

The nonparametric data-driven modelling can be tackled in the time- or frequency domain, see Ljung (1999) and Pintelon and Schoukens (2012b). In this paper we present a frequency domain formulation, which has its advantages, as argumented in Section 3. That is, we estimate the frequency response function (FRF), being the transfer function evaluated along the imaginary axis or the unit circle in, respectively, the continuous-time (CT) or discrete-time (DT) case. This in a user defined frequency band.

Throughout the last decade, kernel methods have gained immense popularity in the field of data-driven modelling (Pillonetto, Dinuzzo, Chen, De Nicolao, & Ljung, 2014). There is some talk of a paradigm shift from the traditional maximum likelihood and output error methods towards these kernel methods (Ljung, Chen, & Mu, 2020). Most of these methods can be interpreted as Gaussian process regression and are related to the regularised least squares (Rasmussen & Williams, 2006). It is known that regularisation introduces a bias, but in such a way that the trade-off between variance and bias decreases the mean square error (MSE). Gaussian process regression was introduced around the new millennium (MacKay, 1998), and has since been applied successfully to a broad spectrum of regression problems. From frequency-domain modelling of nonlinear (Stoddard,

☆ This research was supported in part by the Fund for Scientific Research (FWO Belgium, Ph.D. fellowship strategic basic research 1SB5721N), in part by the Vrije Universiteit Brussel (SRP-19), and in part by the Flemish Government (Methusalem Grant METH1). The material in this paper was presented at the 19th IFAC Symposium on System Identification, July 13–16 2021, Padova, Italy. This paper was recommended for publication in revised form by Associate Editor Gianluigi Pillonetto under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: noel.hallemans@vub.be (N. Hallemans), boris.joukovsky@vub.be (B. Joukovsky), dries.peumans@vub.be (D. Peumans), john.lataire@vub.be (J. Lataire).

https://doi.org/10.1016/j.automatica.2021.110056
0005-1098/© 2021 Elsevier Ltd. All rights reserved.

Birpoutsoukis, Schoukens, & Welsh, 2019), Birpoutsoukis, Marconato, Lataire, and Schoukens (2017), and time-varying systems (Hallemans, Lataire, & Pintelon, 2020) to the identification of relaxation times from battery impedance measurements (Liu & Ciucci, 2020).

Nonparametric kernel-based modelling of LTI systems has already been studied both in the time- (Chen, Ohlsson, & Ljung, 2012) and in the frequency domain Lataire and Chen (2016), where, respectively, the impulse response and the FRF of the system are estimated. Let us first draw the advantages of the kernel-based approach over other nonparametric approaches. For other nonparametric frequency-domain approaches such as the local polynomial method (LPM) (Pintelon, Schoukens, Vandersteen, & Barbé, 2010) and local rational method (LRM) (McKelvey & Guérin, 2012; Pintelon, Peumans, Vandersteen, & Lataire, 2021), a model structure is chosen, respectively, a polynomial or rational function, in each local frequency band. Then, the model order, i.e. the degree of the polynomials, must be determined using criteria such as the minimum description length (MDL) or Akaike information criterion (AIC). Note that the model order is discretely tuneable, the degrees of the polynomials are integers. In the kernel-based approach, the aforementioned questions of model structure and order selection translate into the design of the regularisation matrix, or also called kernel. This is done in two steps; (i) setting a continuous expression for the sampled kernel, imposing the desired properties on the estimate and, hence, defining the model class, and (ii) tuning the hyper parameters of the kernel, which is equivalent to the model order selection. The tuning of the hyper parameters boils down to solving a non-convex optimisation problem, maximising the marginal likelihood, which inherently determines the trade-off between data fit and model complexity. Note that this continuous "model order selection" is a very powerful tool, as it allows for much more freedom within the specified model class than for other nonparametric approaches.

As a summary, the advantages of kernel methods over other nonparametric methods are mainly threefold; (i) the model complexity is continuously tuneable, (ii) the estimate is a continuous function of time or frequency, and (iii) it allows one to impose stability, causality and smoothness on the estimate of the impulse response of the system.

The choice of the continuous expression for the kernel, however, is not trivial. Here lays the heart of kernel-based modelling. In both existing time- and frequency domain approaches (Chen et al., 2012)(Lataire & Chen, 2016), it is imposed that the impulse response is smooth, stable and causal and, hence, can be captured by the discrete- or continuous-time DC (diagonal correlated) or TC (tuned correlated) kernel, which are first order stable-spline kernels (Chen, 2018a, 2018b; Pillonetto, Chiuso, & De Nicolao, 2011). In Scandella, Mazzoleni, Formentin, and Previdi (2020), the use of higher order stable-spline kernels is elaborated, where the order of the kernel is a hyper parameter that is derived from the data. In Zorzi and Chiuso (2018), the so-called generalised power spectral density (GPSD) is introduced as a tool for designing kernels. By choosing a GPSD, which is a two-dimensional function in the decay rate-frequency space describing the statistical power of the impulse response, one can compute the corresponding kernel. It is shown that commonly used kernels such as the DC, TC and stable spline kernels can be derived from GPSDs. Moreover, also new kernels arise, such as the extended DC and TC (tuned correlated) kernels, which are basically the DC and TC kernels (Pillonetto et al., 2014), but where the statistical power can be centred around a certain frequency instead of being centred around frequency zero. Other kernels arise for distributions of poles in certain frequency bands. In Chen, Andersen, Ljung, Chiuso, and Pillonetto (2014), multiple kernels, such as the DC and TC kernels, are superposed. The premultiplying hyper parameters of the kernels are set free, while the remaining hyper parameters are fixed on a grid. It is shown that the hyper parameters can be tuned using sequential convex optimisation techniques and that this method inherently favours sparse solutions.

Also, kernel-based methods commonly assume Gaussian white measurement noise, such that the statistical properties of the noise are determined by a single hyper parameter: the variance.

However, in the case of lightly damped systems measured with low frequency resolution, the modelling becomes more challenging due to the fast changes over the frequency resolution of the FRF. The assumption of smoothness is not satisfied anymore. The LPM fails as a polynomial is not sufficient for modelling the FRF, that is, the model class is not inclusive enough. In this case, the LRM is shown to perform better due to its extended model class (Peumans, Busschots, Vandersteen, & Pintelon, 2018). Similarly, most existing kernel methods fail for lightly-damped systems. The imposed existing kernels are not inclusive enough to capture all the dynamics well.

In this paper, we improve the kernel-based frequency-domain approach of Lataire and Chen (2016) by introducing a new kernel. This new kernel, introduced shortly in Hallemans, Pintelon, Peumans, and Lataire (2021), is a superposition of several kernels; the DC-kernel, capturing the smooth contributions, and kernels carrying prior knowledge about the resonating poles. These resonating poles are estimated through an LRM preprocessing step, and, for each of them, a prior knowledge kernel is formed. An expression for the prior knowledge kernel function has been derived starting from the impulse response of a resonating pole. The performance of this analytic kernel function is compared to the extended DC and TC kernels, obtained from the GPSD approach. These kernels can centre their statistical power around a certain frequency, and hence, are applicable to resonating poles. The weighted sum of the kernels carries information about the resonances of the studied system, and allows one to model highly resonating systems with a better accuracy than the LRM, while keeping the advantages of the kernel-based approach.

The main contributions of this paper with respect to the method of Lataire and Chen (2016) are; (i) the ability of identifying lightly damped systems, (ii) the extension to coloured output noise, (iii) the inclusion of error bounds on the estimate, (iv) a numerically stable implementation for evaluating the marginal likelihood and (v) expressions for obtaining starting values of the hyper parameters.

The remainder of this paper is organised as follows. First, in Section 2, we recall some analytic expressions for the transfer function and impulse response which depend on the poles and residues of the system. Next, the frequency domain modelling problem is described in Section 3. Then, the kernel-based estimator conditioned on the prior knowledge of poles and residues of the system is detailed in Section 4. Further, in Section 5 we explain how the poles and residues can be identified from the LRM. Finally, we compare the performance of the new kernel method with that of the LRM on a simulation example in Section 6. Conclusions are drawn in Section 7.

2. A brief system theoretic background

Consider a causal, strictly proper, single-input single-output (SISO), linear time-invariant (LTI) system G with impulse response $g(t)$. Under zero initial conditions, the input–output relation of such a system yields

$$y_0(t) = g(t) * u(t) \tag{1a}$$

$$y(t) = y_0(t) + n_y(t) \tag{1b}$$

where $u(t)$ and $y(t)$ stand for, respectively, the input and output signals, $n_y(t)$ is filtered continuous-time band-limited white noise and $*$ denotes the convolution

$$g(t) * u(t) = \int_0^\infty g(\tau)u(t-\tau)\mathrm{d}\tau \qquad \text{(CT)} \tag{2a}$$

$$g(nT_s) * u(nT_s) = \sum_{m=0}^{\infty} g(mT_s)u((n-m)T_s) \qquad \text{(DT)}, \tag{2b}$$

where T_s is the sampling period. Note that a distinction has been made between the continuous-time (CT) and discrete-time (DT) cases. Taking, respectively, the Laplace and Z-transform of (1a) yields

$$\begin{aligned} Y_0(s) &= G(s)U(s) && \text{(CT)} \\ Y_0(z) &= G(z)U(z) && \text{(DT)}. \end{aligned} \tag{3}$$

The transfer function is defined as $G(s)$ or $G(z)$ for, respectively, the CT and DT cases.

Assuming simple poles p_n, the impulse response function can be written as

$$g(t) = \sum_{n=1}^{n_p} g_n(t) \tag{4a}$$

$$g_n(t) = e^{-\alpha_n t}\big(\zeta_{1,n}\cos(\omega_n t) + \zeta_{2,n}\sin(\omega_n t)\big), \tag{4b}$$

with n_p the number of real plus complex conjugate pole-pairs. The constants α_n, ω_n, $\zeta_{1,n}$ and $\zeta_{2,n}$ are related to the pole p_n and its residue r_n as

$$\begin{aligned} &\alpha_n = -\mathrm{re}(p_n) \quad \omega_n = \mathrm{im}(p_n) \quad \zeta_{2,n} = -2\,\mathrm{im}(r_n) \\ &\zeta_{1,n} = \begin{cases} \mathrm{re}(r_n) & \text{if } \omega_n = 0, \\ 2\,\mathrm{re}(r_n) & \text{else.} \end{cases} \end{aligned} \tag{5}$$

Based on (4), a kernel for the Gaussian process regression of the FRF from data will be derived in Section 4.2.

Comment: In this section we have assumed simple poles. The method of this paper is also applicable when the non-resonant poles have higher multiplicity.

3. Frequency domain data-driven modelling

The goal of data-driven modelling is to find a mathematical relation between the input and output of the system from data. In practice, the data, obtained from measurements, consists of sampled and windowed time-series

$$\mathbf{u} = [u(0), u(1), \ldots, u(N-1)]^T \tag{6a}$$

$$\mathbf{y} = [y(0), y(1), \ldots, y(N-1)]^T, \tag{6b}$$

where in the continuous-time case $x(n)$ is a shorthand notation for $x(nT_s)$, $x = u, y$, N is the number of samples, $T_s = 1/f_s$ is the sampling period, f_s is the sampling frequency and $T = NT_s$ is the length of the measurement. For dynamical systems, the model is mostly chosen as the impulse response (time domain approach) or transfer function (frequency domain approach). In this paper we follow a frequency domain approach. The argumentation for choosing the frequency domain approach over the time domain one is threefold;

(i) In the time domain approach, the convolution (1a) should be solved for $g(t)$. In the CT case, this convolution is an integral, see (2), which can only be approximated discretely. In the frequency domain, however, it boils down to a simple product (3), and no approximation is needed. Hence, both CT and DT cases can be solved in the frequency domain without approximations.

(ii) For finite impulse response (FIR) approaches (Chen et al., 2012; The MathWorks Inc., 2021), as much data samples should be thrown away as the order of the FIR. This is due to the initial conditions. Hence, from N samples, at most $N/2$ impulse response coefficients can be estimated by the classical algorithms. Such a restriction is not present in the frequency domain approach (Lataire & Chen, 2016).

(iii) The frequency domain approach allows one to estimate the model in a local frequency band $[f_{\min}, f_{\max}]$, which is useable for many real-life applications, and which shortens the computation time.

The data (6) is now transformed into the frequency domain by using the discrete Fourier transform (DFT).

Definition 1. The scaled discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) are, respectively, defined as

$$X(k) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)e^{-j2\pi\frac{kn}{N}} \tag{7a}$$

$$x(n) = \sum_{k=0}^{N-1} X(k)e^{j2\pi\frac{kn}{N}}. \tag{7b}$$

The frequency domain input–output relation now yields,

$$Y(k) = G(\Omega_k)U(k) + L_G(\Omega_k) + N_y(k), \tag{8}$$

where $L_G(\Omega_k)$ stands for the system's transient, a rational function of Ω that has the same poles as $G(\Omega)$,

$$\Omega_k = \begin{cases} j\omega_k & \text{(CT)} \\ e^{j\omega_k/f_s} & \text{(DT)} \end{cases} \quad \text{with } \omega_k = \frac{2\pi k}{T} \tag{9}$$

and $N_y(k)$ represents the output noise, satisfying Assumption 1.

Assumption 1. The disturbing noise signal $n_y(t)$ is assumed to be filtered continuous-time band-limited white noise. Hence its DFT, (7a), can be written as

$$N_y(k) = \underbrace{H(\Omega_k)E(k)}_{=V(k)} + L_H(\Omega_k), \tag{10}$$

where $H(\Omega)$ is a rational function of Ω representing the filtering operation, $E(k)$ is circular complex Gaussian distributed and uncorrelated over the frequency, and $L_H(\Omega)$ a rational function of Ω representing the noise transient. The noise variance is denoted as $\sigma_V^2(k) = \mathbb{E}\{|V(k)|^2\}$.

Eq. (8) can now be rewritten as

$$Y(k) = G(\Omega_k)U(k) + L(\Omega_k) + V(k), \tag{11}$$

where the total transient $L(\Omega)$ groups the system and noise transient,

$$L(\Omega) = L_G(\Omega) + L_H(\Omega). \tag{12}$$

Note that $L(\Omega)$ has the same poles as $G(\Omega)$ under the assumption that the output noise originates from an unobserved input. Both, the FRF $G(\Omega)$ and the total transient $L(\Omega)$ are smooth functions of the frequency. Hence, in order to discriminate between these, the excitation signal $U(k)$ should have a random nature over k, such that $G(\Omega_k)U(k)$ is not correlated over frequency anymore. This translates into the following assumption on the input signal (Pintelon et al., 2021).

Assumption 2. The spectral difference at the excited frequencies $f_k = kf_s/N$ of the input signal $u(t)$,

$$|U(k+1) - U(k)| = \mathcal{O}(N^0), \tag{13}$$

does not converge to zero when $N \to \infty$.

Examples of signals $u(t)$ satisfying Assumption 2 include;

(i) filtered white noise,
(ii) periodic filtered white noise,
(iii) pseudo-random binary sequences (PRBS) and
(iv) random phase multisines.

In many applications, the user is only interested in the FRF over a certain frequency band $[f_{\min}, f_{\max}]$. Accordingly, this frequency band of interest is now set by choosing the begin and end samples in the frequency domain

$$F_1 = \text{floor}(f_{\min}T) \text{ and } F_2 = \text{ceil}(f_{\max}T) \in \mathbb{N}. \tag{14}$$

The frequency domain data in this frequency band of interest is then defined as

$$\mathbf{U} = \begin{bmatrix} U(F_1) & U(F_1+1) & \dots & U(F_2) \end{bmatrix}^{\mathrm{T}} \tag{15a}$$

$$\mathbf{Y} = \begin{bmatrix} Y(F_1) & Y(F_1+1) & \dots & Y(F_2) \end{bmatrix}^{\mathrm{T}}. \tag{15b}$$

Note that these two vectors are complex-valued and have dimensions $F \times 1$ with $F = F_2 - F_1 + 1$. Eq. (11) can now be rewritten in vector form for the frequency band of interest,

$$\mathbf{Y} = \text{diag}(\mathbf{U})\mathbf{G} + \mathbf{L} + \mathbf{V}, \tag{16}$$

with the unknowns,

$$\mathbf{G} = \begin{bmatrix} G(\Omega_{F_1}) & G(\Omega_{F_1+1}) & \dots & G(\Omega_{F_2}) \end{bmatrix}^{\mathrm{T}} \tag{17a}$$

$$\mathbf{L} = \begin{bmatrix} L(\Omega_{F_1}) & L(\Omega_{F_1+1}) & \dots & L(\Omega_{F_2}) \end{bmatrix}^{\mathrm{T}} \tag{17b}$$

$$\mathbf{V} = \begin{bmatrix} V(F_1) & V(F_1+1) & \dots & V(F_2) \end{bmatrix}^{\mathrm{T}}. \tag{17c}$$

Comments:

(1) Assumption 1 of band-limited noise is not restrictive, in the sense that practical measurements of CT systems should always be done using a band-limited acquisition setup: the analog signals are lowpass (anti-alias) filtered before sampling. This makes the observed noise band-limited (Pintelon, Schoukens, & Guillaume, 2006). In DT, noise is always band-limited; from zero to Nyquist.
(2) In case of non-ideal actuators and/or feedback (Pintelon & Schoukens, 2012a), the input of the system cannot be imposed. Hence one cannot choose a signal satisfying Assumption 2. Still, by following the reasoning of Pintelon and Schoukens (2012a), one can retrieve the FRF of the considered system by modelling from reference $r(t)$ to input $u(t)$ and output $y(t)$ simultaneously and, next, dividing both FRFs. This so long as the reference signal $r(t)$ satisfies Assumption 2.
(3) The frequency band of interest can be chosen as $[0, f_s/2]$, and, hence, using all the information of the measurement.

Problem formulation: The frequency-domain nonparametric data-driven modelling boils down to solving (16) for $\mathbf{G}$, given the data $\mathcal{D} = \{\mathbf{U}, \mathbf{Y}\}$. Note that, as there are twice as many unknowns as equations, this problem is unsolvable without introducing a local modelling technique or regularisation. Here, we tackle this regression problem by a kernel-based estimator, allowing a beneficial trade-off between bias and variance, imposing stability, causality and smoothness, and interpolating $G(\Omega)$ between the DFT frequencies.

4. Kernel-based estimator

In this section, we show how a continuous estimate of the FRF $G(\Omega)$ can be found by solving (16) using a kernel-based estimator. It is based on the approach of Lataire and Chen (2016), yet with an extended kernel (Hallemans et al., 2021), suited for lightly-damped systems, considering coloured noise and enclosing uncertainty bounds on the estimate.

4.1. Maximum a posteriori estimate

Consider the real-valued impulse response function $g(t)$ (4), assumed to be distributed as a Gaussian process (Rasmussen & Williams, 2006) with covariance function $k_g(t,\tau) = \mathbb{E}\{g(t)g(\tau)\}$,

$$g(t) \sim \mathcal{GP}(0, k_g(t,\tau)). \tag{18}$$

Using a frequency domain approach, it is possible to compute the *maximum a posteriori* (MAP) estimate for the FRF, that is, the expected value of $G(\Omega)$ conditioned on the data $\mathcal{D} = \{\mathbf{U}, \mathbf{Y}\}$, and its uncertainty

$$\hat{G}(\Omega) = \mathbb{E}\{G(\Omega)|\mathcal{D}\} \tag{19a}$$

$$\hat{\sigma}_G^2(\Omega) = \mathbb{E}\{|\hat{G}(\Omega) - \mathbb{E}\{\hat{G}(\Omega)\}|^2\} \tag{19b}$$

$$\hat{\rho}_G^2(\Omega) = \mathbb{E}\{(\hat{G}(\Omega) - \mathbb{E}\{\hat{G}(\Omega)\})^2\}. \tag{19c}$$

Note that $\hat{G}(\Omega)$ is not circular complex distributed, because $g(t)$ (18) is a non-stationary Gaussian process. Therefore, (19c) is nonzero. These can be obtained under closed form, as studied in Lataire and Chen (2016) and Hallemans et al. (2021). Here, we only summarise the results. The *a posteriori* estimates depend on the data $\mathcal{D}$, the noise variances $\sigma_V^2(k)$, and the frequency domain kernels $k_G(\Omega, \tilde{\Omega})$ and $k_L(\Omega, \tilde{\Omega})$. The kernel for the system term is obtained by taking the two-dimensional Fourier transform of $k_g(t,\tau)$

$$k_G(\Omega, \tilde{\Omega}) = \mathbb{E}\{G(\Omega)\overline{G(\tilde{\Omega})}\} = \begin{cases} \int_0^\infty \int_0^\infty k_g(t,\tau) e^{-\Omega t} e^{\tilde{\Omega}\tau} \mathrm{d}t\mathrm{d}\tau & \text{(CT)} \\ \sum_{n=0}^\infty \sum_{m=0}^\infty k_g(nT_s, mT_s) \Omega^{-nT_s} \tilde{\Omega}^{mT_s} & \text{(DT)} \end{cases} \tag{20}$$

and the kernel of the transient term is assumed to be proportional to (20)

$$k_L(\Omega, \tilde{\Omega}) = \mathbb{E}\{L(\Omega)\overline{L(\tilde{\Omega})}\} \propto k_G(\Omega, \tilde{\Omega}). \tag{21}$$

The latter is justified because $G(\Omega)$ and $L(\Omega)$ have the same poles (except for the noise transient). Moreover, since $\hat{G}(\Omega)$ is not circular complex distributed, the relation function is also needed, which is easily computed from the covariance function for $X = G, L$

$$\mathbb{E}\{X(\Omega)X(\tilde{\Omega})\} = \mathbb{E}\{X(\Omega)\overline{X(\tilde{\Omega})}\} = k_X(\Omega, \overline{\tilde{\Omega}}). \tag{22}$$

The maximum a posteriori estimate and its variance are given by

$$\hat{G}(\Omega) = \hat{\mathbf{G}}_{\text{re}[1]}(\Omega) + j\hat{\mathbf{G}}_{\text{re}[2]}(\Omega), \tag{23a}$$

$$\hat{\sigma}_G^2(\Omega) = \text{cov}\{\hat{\mathbf{G}}_{\text{re}}\}_{[1,1]}(\Omega) + \text{cov}\{\hat{\mathbf{G}}_{\text{re}}\}_{[2,2]}(\Omega) \tag{23b}$$

$$\hat{\rho}_G^2(\Omega) = \text{cov}\{\hat{\mathbf{G}}_{\text{re}}\}_{[1,1]}(\Omega) - \text{cov}\{\hat{\mathbf{G}}_{\text{re}}\}_{[2,2]}(\Omega) + 2j\text{cov}\{\hat{\mathbf{G}}_{\text{re}}\}_{[1,2]}(\Omega), \tag{23c}$$

with

$$\hat{\mathbf{G}}_{\text{re}}(\Omega) = \mathbf{K}_{G_{\text{re}},Y_{\text{re}}}(\Omega)\mathbf{K}_{Y_{\text{re}}}^{-1}\mathbf{Y}_{\text{re}}$$

$$\text{cov}\{\hat{\mathbf{G}}_{\text{re}}\}(\Omega) = \mathbf{K}_{G_{\text{re}},Y_{\text{re}}}(\Omega)\mathbf{K}_{Y_{\text{re}}}^{-1}\mathbf{K}_{V_{\text{re}}}\mathbf{K}_{Y_{\text{re}}}^{-T}\mathbf{K}_{G_{\text{re}},Y_{\text{re}}}^{T}(\Omega)$$

$$\mathbf{X}_{\text{re}} = \begin{bmatrix} \text{re}(\mathbf{X}) & \text{im}(\mathbf{X}) \end{bmatrix}^T \text{ with } \mathbf{X} = \mathbf{Y},\ \mathbf{G}$$

$$\mathbf{K}_{G_{\text{re}},Y_{\text{re}}}(\Omega) = \frac{1}{2}\begin{bmatrix} \text{re}\big(\mathbf{K}_{G,Y} + \mathbf{C}_{G,Y}\big) & -\text{im}\big(\mathbf{K}_{G,Y} - \mathbf{C}_{G,Y}\big) \\ \text{im}\big(\mathbf{K}_{G,Y} + \mathbf{C}_{G,Y}\big) & \text{re}\big(\mathbf{K}_{G,Y} - \mathbf{C}_{G,Y}\big) \end{bmatrix}$$

$$\mathbf{K}_{Y_{\text{re}}} = \frac{1}{2}\begin{bmatrix} \text{re}\big(\mathbf{K}_{Y} + \mathbf{C}_{Y}\big) & -\text{im}\big(\mathbf{K}_{Y} - \mathbf{C}_{Y}\big) \\ \text{im}\big(\mathbf{K}_{Y} + \mathbf{C}_{Y}\big) & \text{re}\big(\mathbf{K}_{Y} - \mathbf{C}_{Y}\big) \end{bmatrix}$$

$$\mathbf{K}_{V_{\text{re}}} = 0.5\ I_2 \otimes \mathbf{K}_V$$

$$\mathbf{K}_{G,Y}(\Omega) = \begin{bmatrix} k_G(\Omega, \Omega_{F_1}) & \dots k_G(\Omega, \Omega_{F_2}) \end{bmatrix} \text{ diag}(\overline{\mathbf{U}})$$

$$\mathbf{C}_{G,Y}(\Omega) = \begin{bmatrix} k_G(\Omega, \overline{\Omega_{F_1}}) & \dots k_G(\Omega, \overline{\Omega_{F_2}}) \end{bmatrix} \text{ diag}(\mathbf{U})$$

$$\mathbf{K}_Y = \mathbf{U}\mathbf{U}^H \odot \mathbf{K}_G + \mathbf{K}_L + \mathbf{K}_V \tag{24}$$
$$\mathbf{C}_Y = \mathbf{U}\mathbf{U}^T \odot \mathbf{C}_G + \mathbf{C}_L$$
$$\mathbf{K}_{X[l,m]} = k_X(\Omega_{l+F_1-1}, \Omega_{m+F_1-1}) \quad X = G, L$$
$$\mathbf{C}_{X[l,m]} = k_X(\Omega_{l+F_1-1}, \overline{\Omega_{m+F_1-1}}) \quad X = G, L$$
$$\mathbf{K}_{V[l,m]} = \sigma_V^2(l + F_1 - 1)\delta_{l,m}.$$

The symbols $\otimes$ and $\odot$ denote, respectively, the Kronecker and the elementwise product. The noise variances $\sigma_V^2(k)$ are *a priori* unknown, but, they will be estimated in an LRM preprocessing step (see Section 5).

Comments:

(1) Note that (19) are continuous estimates defined over the entire complex axis or unit circle. However, as the model is only trained for data inside a certain frequency band, the extrapolation outside this band will not be accurate. More precisely, these equations do not impose that the estimate at DC is real-valued, which should. The methodology for imposing this is detailed in Lataire and Chen (2016), and is omitted here for the sake of clarity. Also, the DC-value of the FRF can easily be obtained by exciting the system with a constant input.
(2) There is a relation between the Gaussian process regression and the regularised linear weighted least squares, see Hallemans and Pintelon (2021).
(3) The estimator is based on the prior knowledge incorporated in the kernel functions $k_G(\Omega, \tilde{\Omega})$, $k_L(\Omega, \tilde{\Omega})$ and the noise covariance matrix $\mathbf{K}_V$. These define the model class and are *a priori* unknown. In the next section we derive analytic expressions for the kernel functions imposing stability, causality and smoothness, while carrying prior knowledge. In Section 5 we show how to obtain this prior knowledge from an LRM preprocessing step.

4.2. Design of the kernel functions

In this section, we derive analytic expressions for the kernel functions $k_G(\Omega, \tilde{\Omega})$ (20) and $k_L(\Omega, \tilde{\Omega})$ (21), based on the system theoretic background of Section 2.

It is shown in Lataire and Chen (2016), that the causality, stability and smoothness of the impulse response is well described by a Gaussian process with the DC-kernel, that is $k_g(t, \tau) = \gamma_g k_{\mathrm{dc}}(t, \tau)$ with

$$k_{\mathrm{dc}}(t, \tau) = e^{-\alpha(t+\tau)} e^{-\beta|t-\tau|} \quad \alpha, \beta, t, \tau \geq 0. \tag{25}$$

This time-domain kernel imposes that the impulse response decays exponentially over time at a rate α and that the correlation of the impulse response function between times t and τ decreases exponentially at a rate β. The hyper parameters γ_g, α and β are *a priori* unknown, but can be estimated using the marginal likelihood, see Section 4.3.

Unfortunately, for lightly-damped systems measured with insufficient frequency resolution, the sampled FRF is not smooth across the DFT lines. Hence, the DC-kernel introduces large estimation errors for lightly-damped systems.

The main idea of this article is to modify the kernel function $k_g(t, \tau)$ by adding prior knowledge (pk). This is done by superposing on the DC-kernel a certain number of kernels $k_{\mathrm{pk},n}(t, \tau)$, carrying information about the resonant poles of the system:

$$k_g(t, \tau) = \gamma_g k_{\mathrm{dc}}(t, \tau) + \sum_{n|p_n \in \mathbb{P}_\mathrm{r}} k_{\mathrm{pk},n}(t, \tau). \tag{26}$$

The set of resonating poles $\mathbb{P}_\mathrm{r}$, with cardinality n_r, is *a priori* unknown, but, as we show in Section 5.1, it can be estimated from the data using an LRM preprocessing step. The LRM will decide whether a pole should be considered as a resonating one, depending on the SNR and frequency resolution of the measurement. The non-resonant poles are captured by the DC-kernel, while the resonant poles each have their own kernel. Note that it has been assumed that the $g_n(t)$ in (4) are independent Gaussian processes.

Different analytical expressions for the kernel $k_{\mathrm{pk},n}(t, \tau)$, carrying prior knowledge, are considered.

(1) The so-called prior knowledge kernel function is defined in Hallemans et al. (2021). It is assumed that for a complex pole p_n, $\zeta_{1,n}$ and $\zeta_{2,n}$ in (4b) are independent zero-mean Gaussian random variables. The corresponding term $g_n(t)$ (4b) of the impulse response function is then a Gaussian process with kernel function

$$k_{\mathrm{pk},n}(t, \tau) = e^{-\alpha_n(t+\tau)}\big(\gamma_{1,n}\cos(\omega_n(t-\tau)) + \gamma_{2,n}\cos(\omega_n(t+\tau))\big) \quad t, \tau \geq 0, \tag{27}$$

where

$$\gamma_{1,n} = \frac{\sigma^2_{\zeta_{1,n}} + \sigma^2_{\zeta_{2,n}}}{2}, \quad \gamma_{2,n} = \frac{\sigma^2_{\zeta_{1,n}} - \sigma^2_{\zeta_{2,n}}}{2} \tag{28}$$

and with $\sigma^2_{\zeta_{i,n}}$ the variance of the variable $\zeta_{i,n}$, $i = 1, 2$. Note that the hyper parameters α_n, ω_n, $\gamma_{1,n}$ and $\gamma_{2,n}$ are directly linked to the information about the poles and residues of the system through (5).

(2) The extended DC-kernel, from the GPSD approach (Zorzi & Chiuso, 2018), is given by

$$\tilde{k}_{\mathrm{pk},n}(t, \tau) = \gamma_n e^{-\alpha_n(t+\tau)} e^{-\beta_n|t-\tau|}\cos(\omega_n(t-\tau)). \tag{29}$$

(3) The extended TC-kernel, also from the GPSD approach, is a special case of the extended DC-kernel, obtained by choosing $\beta_n = \alpha_n$ in (29)

$$\check{k}_{\mathrm{pk},n}(t, \tau) = \gamma_n e^{-2\alpha_n \max(t,\tau)}\cos(\omega_n(t-\tau)). \tag{30}$$

A comparison between the kernel functions $k_{\mathrm{pk},n}(t, \tau)$, $\tilde{k}_{\mathrm{pk},n}(t, \tau)$ and $\check{k}_{\mathrm{pk},n}(t, \tau)$ is made in the simulation section. The first one performs best, while the last one performs worst on the proposed simulation.

The frequency domain kernels are obtained using the two-dimensional Fourier transform (20). Accordingly, one obtains

$$k_G(\Omega, \tilde{\Omega}) = \gamma_g k_{\mathrm{DC}}(\Omega, \tilde{\Omega}) + \sum_{n|p_n \in \mathbb{P}_\mathrm{r}} k_{\mathrm{PK},n}(\Omega, \tilde{\Omega}), \tag{31}$$

with the frequency domain kernel functions $k_{\mathrm{DC}}(\Omega, \tilde{\Omega})$, $k_{\mathrm{PK},n}(\Omega, \tilde{\Omega})$, $\tilde{k}_{\mathrm{PK},n}(\Omega, \tilde{\Omega})$ and $\check{k}_{\mathrm{PK},n}(\Omega, \tilde{\Omega})$ defined in Appendices A and B for, respectively, the CT and DT cases.

As $L(\Omega)$ shares the poles of $G(\Omega)$, it is natural to use the same covariance function for $k_L(\Omega, \tilde{\Omega})$. However, with different premultiplying factors γ, as the zeros are not shared. In the remainder of this text, premultiplying factors for the system and transient term are, respectively, denoted with subscripts G and L.

Comments:

(1) The kernels carrying prior knowledge, $k_{\mathrm{pk},n}(t, \tau)$, $\tilde{k}_{\mathrm{pk},n}(t, \tau)$ and $\check{k}_{\mathrm{pk},n}(t, \tau)$, impose causality, stability and smoothness by setting $\alpha_n > 0$.
(2) For $\omega_n = 0$, the kernels $\tilde{k}_{\mathrm{pk},n}(t, \tau)$ and $\check{k}_{\mathrm{pk},n}(t, \tau)$, respectively, yield the DC and TC kernels (Pillonetto et al., 2014). The extension of these kernels makes it possible to centre the statistical power around ω_n, hence its applicability for capturing resonating poles.

(3) One single kernel of the type (27), (29) or (30) is unable to capture different strong resonances located in different frequency bands. Hence, one has to use the superposition of these kernels with different hyper parameters. As a drawback, the number of hyper parameters increases linearly with the number of resonating poles. In Section 5.3, we show how the number of hyper parameters can be reduced by fixing some of them to values obtained by an LRM preprocessing step.
(4) $k_{\mathrm{dc}}(t,\tau)$, $\tilde{k}_{\mathrm{pk},n}(t,\tau)$ and $\check{k}_{\mathrm{pk},n}(t,\tau)$ are non-degenerate kernels, that is, they have infinite rank. $k_{\mathrm{pk},n}(t,\tau)$ is a degenerate one, that is, it has rank two. See Rasmussen and Williams (2006), Definition 4.1.
(5) The poles of the noise transient term $L_H(\Omega)$ that are different of the poles of $G(\Omega)$ are captured by the DC-kernel.

4.3. Tuning of the kernel functions

The previously derived covariance functions depend on so called hyper parameters. These hyper parameters are listed in a vector θ, which is estimated by maximising the marginal likelihood (Rasmussen & Williams, 2006);

$$\hat{\theta} = \arg\min_{\theta} -2\log p(\mathbf{Y}_{\mathrm{re}} \mid \theta), \tag{32a}$$

with

$$-2\log p(\mathbf{Y}_{\mathrm{re}} \mid \theta) = \mathbf{Y}_{\mathrm{re}}^T \mathbf{K}_{Y_{\mathrm{re}}}^{-1}(\theta)\mathbf{Y}_{\mathrm{re}} + \log\det \mathbf{K}_{Y_{\mathrm{re}}}(\theta) + 2F\log(2\pi). \tag{32b}$$

Comments:

(1) Note that the above cost function is non-convex and, hence, nonlinear optimisation is needed. Moreover, the larger the number of resonating poles n_{r}, the larger the number of hyper parameters to be estimated. Therefore, the starting values θ_0 are extremely important.
(2) The matrix $\mathbf{K}_{Y_{\mathrm{re}}}$ has dimension $2F \times 2F$, hence the allocated memory for this matrix increases with $4F^2$, possibly making the computation time prohibitively long when a high number of frequencies are considered.
(3) On top of that $\mathbf{K}_{Y_{\mathrm{re}}}$ becomes ill-conditioned for high signal-to-noise ratios of the measurements, which makes the computation of the inverse and the determinant very inaccurate. This can be seen from (24), where the only full rank matrix is $\mathbf{K}_V$ which has small values compared with the other terms for high SNR. Thereupon, the tuning of the kernel functions is the bottleneck of this kernel-method.

In the following sections we propose solutions for the problems (1) and (3) mentioned above. Section 4.4 proposes a numerically stable evaluation of the marginal likelihood and Section 5 explains how starting values of the hyper parameters θ_0 can be found from an LRM preprocessing step and how some of them can be fixed.

4.4. Numerically stable evaluation of the cost function

In this section we derive numerically stable expressions for the inverse and the determinant of $\mathbf{K}_{Y_{\mathrm{re}}}$, leading to a numerically stable evaluation of the marginal likelihood (32).

Theorem 1. *Define $A = I_2 \otimes \mathbf{K}_V$ and $B = 2\mathbf{K}_{Y_{re}} - A$. The inverse of the real-valued positive definite symmetric matrix $\mathbf{K}_{Y_{re}} = \frac{1}{2}(A+B)$ can be computed as*

$$\mathbf{K}_{Y_{\mathrm{re}}}^{-1} = 2MM^T, \tag{33}$$

with

$$M = A^{-1/2}WS^{-1/2}, \quad S = I_{2F} + \Sigma \tag{34}$$

and where $W\Sigma W^T$ is the singular value decomposition of the symmetric positive semi-definite matrix $A^{-1/2}BA^{-1/2}$ with singular value matrix $\Sigma = \mathrm{diag}([\sigma_1, \sigma_2, \ldots, \sigma_{2F}])$.

Proof. see Appendix C.

Theorem 2. *The logarithm of the determinant of the real-valued positive definite symmetric matrix $\mathbf{K}_{Y_{\mathrm{re}}}$ can be computed as*

$$\log\det \mathbf{K}_{Y_{\mathrm{re}}} = 2\sum_{k=1}^{F}\log\sigma_V^2(k) + \sum_{k=1}^{2F}\log(1+\sigma_k) - 2F\log 2, \tag{35}$$

where σ_k is defined in Theorem 1.

Proof. see Appendix D.

Comments: Note that the matrix $A^{-1/2}$ can be computed in a stable way as it is diagonal, and that the entries of S^{-1} cannot explode as $\sigma_k \geq 0$

$$\sup_{\sigma_k} \frac{1}{1+\sigma_k} = 1. \tag{36}$$

Also, direct computation of $\det \mathbf{K}_{Y_{\mathrm{re}}}$ can rapidly lead to a numerical under- or overflow, and this is solved via (35).

5. Preprocessing step with the LRM

In order to maximise the marginal likelihood, which is a non-convex cost function, one needs good initial values for the hyper parameters. In this section we explain how these can be obtained from an LRM preprocessing step. Moreover, this LRM preprocessing step makes it possible to fix some hyper parameters, simplifying the tuning of the remaining ones.

5.1. Obtaining estimates of the resonating poles

We use the LRM implementation of Pintelon et al. (2021) which solves (11) for $G(\Omega_k)$ in a local frequency band of $2n+1$ DFT frequencies centred around k,

$$\text{local band} = [k-n, \ldots, k, \ldots, k+n]. \tag{37}$$

In this band, the FRF $G(\Omega)$ and transient term $L(\Omega)$ are modelled by rational forms in $j\omega$, both in CT and DT, with complex coefficients, i.e.

$$G(j\omega_{k+r}) = \frac{B_k(j\omega_r)}{A_k(j\omega_r)}, \qquad L(j\omega_{k+r}) = \frac{I_k(j\omega_r)}{A_k(j\omega_r)}, \tag{38}$$

with $r = -n, \ldots, 0, \ldots, n$ and

$$X_k(j\omega_r) = \sum_{l=0}^{n_{x,k}} x_{l,k}(j\omega_r)^l \quad \text{with } X = A, B, I. \tag{39}$$

Note that the denominators of both G and L are chosen to be equal as the transfer function and transient term share the same poles. The noise transient is captured by the choice $n_{i,k} > n_{a,k}$. Moreover, the LRM will decide in each local band on the degrees of the polynomials $n_{a,k}$, $n_{b,k}$ and $n_{i,k}$. From (11), the local model yields

$$Y(k+r) = \frac{B_k(j\omega_r)}{A_k(j\omega_r)}U(k+r) + \frac{I_k(j\omega_r)}{A_k(j\omega_r)} + V(k+r). \tag{40}$$

From this equation, the complex coefficients $x_{l,k}$ are solved using the output error method, that is, minimising the cost function,

$$V_{OE} = \sum_{r=-n}^{n} \left| Y(k+r) - \frac{B_k(j\omega_r)}{A_k(j\omega_r)} U(k+r) - \frac{I_k(j\omega_r)}{A_k(j\omega_r)} \right|^2, \tag{41}$$

with respect to the parameters $b_{l,k}$, $a_{l,k}$ and $i_{l,k}$ for each local band. Complex coefficients are used because, compared with the real valued case, only half the model order is needed. The residual impact of the poles at the negative frequencies is captured by the choice that the numerator order is always larger than the denominator order. Note that the output error yields the maximum likelihood estimator when assuming that the noise is white in the local frequency band. This at the cost of a nonlinear minimisation problem that is initialised with the linear least squares solution (Levy, 1959). An estimate of the FRF at DFT frequency k is then given by

$$\hat{G}(\Omega_k) = \frac{\hat{b}_{0,k}}{\hat{a}_{0,k}}. \tag{42}$$

From the residuals, an estimate of the noise variance $\hat{\sigma}_V^2(k)$ is also computed. Moreover, the polynomial $\hat{A}_k(s)$ estimated by the LRM also provides information about the poles of the system. Define the set, with cardinality $n_{a,k}$, of the estimated poles for the local band centred around harmonic k,

$$\mathbb{P}_k = \{p + j\omega_k \mid \hat{A}_k(p) = 0\}. \tag{43}$$

For each of these sets we select the pole laying within the local band that resonates the most, i.e. the one with the smallest absolute angle with respect to the imaginary axis,

$$\begin{cases} p_k = \arg\min_{p\in\mathbb{P}_k} |\angle p| \\ \frac{\pi}{2} < |\angle p_k| < \frac{3\pi}{4} \\ \frac{k-n}{T} \le \frac{\mathrm{im}(p_k)}{2\pi} \le \frac{k+n}{T}. \end{cases} \tag{44}$$

For each of the poles p_k, the corresponding residue r_k and variance of the pole $\sigma_{p_k}^2$ can also be computed from the LRM (Pintelon, Guillaume, & Schoukens, 2007). Moreover, it can be shown that the estimated poles are circular complex normally distributed. It is proven in Pintelon et al. (2021) that the coefficients of the rational functions are asymptotically circular complex distributed. This is also true for the poles as they are an analytical function of these coefficients.

Note that for some indices k, the LRM will only choose a polynomial model, i.e. $n_{a,k} = 0$, and hence no poles can be estimated. The LRM decides whether a pole is significant to be modelled by a rational form instead of a polynomial. Remark that this decision depends on the frequency resolution and the SNR of the measurements, the higher the SNR and/or the lower the frequency resolution, the more complex the local rational model. We assume that the true poles of the system that are not estimated by LRM, because the local spectrum could simply be modelled by a polynomial, correspond to smooth functions and hence can be captured by the DC-kernel.

Of course, the poles p_k estimated by the LRM will not all coincide with the true poles of the system and the number of poles p_k will be higher than the true number of resonating poles. In order to obtain from these values p_k the set $\mathbb{P}_\mathrm{r}$ of resonating poles, one needs to solve the problem of clustering the poles in an unknown number n_r of clusters, and defining for each cluster the resonating pole. This is done by using a density-based algorithm for discovering clusters, Density Based Spatial Clustering of Applications with Noise (DBSCAN) (Sander, Ester, Kriegel, & Xu, 1998; Schubert, Sander, Ester, Kriegel, & Xu, 2017). The advantage of this algorithm is that not all points are attributed to a cluster, some points are considered as outliers. The DBSCAN algorithm works with two input variables; a search radius `epsilon`, chosen as the frequency resolution f_s/N, and a minimal number of points per cluster `minpts`, which is chosen to be n. We use the Euclidean distance for the two dimensional data $(\mathrm{re}(p_k), \mathrm{im}(p_k))$. This makes sense because the poles are circular complex normally distributed. The DBSCAN algorithm attributes to each pole a cluster number $n \in \{1, 2, \ldots, n_\mathrm{r}\}$ or the value -1 when the pole is regarded as an outlier. Results are shown in Fig. 2. Finally, within each cluster, the pole with the smallest variance is selected, i.e.

$$\hat{p}_n = \arg \min_{p_k \in \text{cluster } n} \sigma_{p_k}^2 \qquad \text{for } n = 1, 2, \ldots, n_\mathrm{r}, \tag{45}$$

with corresponding residues $\hat{r}_n$. The sets of resonating poles and residues are then formed as

$$\mathbb{P}_\mathrm{r} = \{\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_{n_\mathrm{r}}\} \text{ and } \mathbb{R}_\mathrm{r} = \{\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_{n_\mathrm{r}}\}. \tag{46}$$

5.2. Initialisation of the hyper parameters

In this section we derive some heuristics for obtaining starting values for the hyper parameters.

5.2.1. DC-kernel

Starting values for the hyper parameters γ_g, α and β of the DC-kernel (25) are obtained as follows. First, remove the influence of the resonating poles in the LRM estimate $\hat{G}(j\omega_k)$ (42), giving

$$\hat{G}_\mathrm{smooth}(j\omega_k) = \hat{G}(j\omega_k) - \sum_{n=1}^{n_\mathrm{r}} \frac{\hat{r}_n}{j\omega_k - \hat{p}_n}. \tag{47}$$

From this smooth FRF we compute the IDFT (7b) $\hat{g}_\mathrm{smooth}(t)$ which we assume to be distributed as

$$\hat{g}_\mathrm{smooth}(t) \sim \mathcal{GP}(0, \gamma_g k_\mathrm{dc}(t, \tau)), \tag{48}$$

with $k_\mathrm{dc}(t)$ the DC-kernel given by (25). For β we take minus the slope of the linear regression of $\log|R(m)|$ as starting value, where $R(m)$ is the sample correlation of $g_\mathrm{smooth}(t)$,

$$R(m) = \frac{1}{\tilde{N} - m} \sum_{t=0}^{\tilde{N}-m-1} \hat{g}_\mathrm{smooth}(t)\hat{g}_\mathrm{smooth}(t+m). \tag{49}$$

Finally, α is initialised as minus the slope of the linear regression of $\log|g_\mathrm{smooth}(t)|$, and $\sqrt{\gamma_g}$ as the intercept parameter. The rationale for these choices are that γ_g sets the variance at time $t = 0$, β sets the decay rate of the correlation and α sets the exponential damping of the impulse response, as it can be seen in (25). For both linear regression problems the values of m and t should run from 0 till $\tilde{N} - 1$, where $\tilde{N}$ is chosen such that there is no time leakage present in the iDFT (7b) (Rabiner & Gold, 1975).

5.2.2. PK-kernel

Starting values for the hyper parameters of the prior knowledge kernel (27) are obtained as follows,

$$\alpha_n = -\mathrm{re}(\hat{p}_n) \qquad \omega_n = \mathrm{im}(\hat{p}_n) \tag{50a}$$

$$\gamma_{G,1,n} = 2|\hat{r}_n|^2 \qquad \gamma_{G,2,n} = 2\mathrm{re}(\hat{r}_n^2), \tag{50b}$$

where $\gamma_{G,1,n}$ and $\gamma_{G,2,n}$ are set to the best estimates we can have of (28) from the estimated residues, i.e. by setting $\hat{\sigma}_{\zeta_{i,n}}^2 = \zeta_{i,n}^2$, where $\zeta_{1,n}$ and $\zeta_{2,n}$ are given by (5);

$$\gamma_{G,1,n} = \frac{\zeta_{1,n}^2 + \zeta_{2,n}^2}{2} \quad \text{and} \quad \gamma_{G,2,n} = \frac{\zeta_{1,n}^2 - \zeta_{2,n}^2}{2}. \tag{51}$$

Table 1
Number of hyper parameters to be tuned.

Kernel carrying pk	n_θ	n_θ after fixing
$k_{\text{pk},n}(t,\tau)$	$4+6n_\text{r}$	$4+2n_\text{r}$
$\tilde{k}_{\text{pk},n}(t,\tau)$	$4+5n_\text{r}$	$4+3n_\text{r}$
$\check{k}_{\text{pk},n}(t,\tau)$	$4+4n_\text{r}$	$4+2n_\text{r}$

5.2.3. Extended DC and TC-kernel

Starting values for the hyper parameters of the extended DC-kernel (29) are obtained as follows,

$$\alpha_n = -\text{re}(\hat{p}_n) \qquad \omega_n = \text{im}(\hat{p}_n) \tag{52a}$$

$$\beta_n = -\text{re}(\hat{p}_n) \qquad \gamma_{G,n} = 4\ \text{re}(\hat{r}_n)^2. \tag{52b}$$

Note for all kernels: initial values for the premultiplying factors for the transient are chosen as $\gamma_L = \gamma_G/10$.

5.3. Reducing the number of hyper parameters

Depending on the choice of the kernel carrying prior knowledge, a different number of hyper parameters n_θ should be tuned. These are given in Table 1. Note again that we have chosen the scale factors γ differently for $k_G(\Omega,\tilde{\Omega})$ and $k_L(\Omega,\tilde{\Omega})$, while the other hyper parameters α, β, α_n, β_n and ω_n are chosen to be equal for both.

In order to keep the amount of hyper parameters θ reasonable for the tuning of the kernel functions, we propose to keep the hyper parameters obtained from the resonating poles and residues (46) fixed. Still, for the extended DC and TC kernels, simulations indicate that all premultiplying factor should be tuned. That is, we keep the following parameters fixed

$$\text{PK} - \text{kernel}: \alpha_n, \omega_n, \gamma_{G,1,n}, \gamma_{G,2,n}$$

$$\text{Extended DC and TC-kernel}: \alpha_n, \omega_n \tag{53}$$

with $n = 1, 2, \ldots, n_\text{r}$. The remaining number of hyper parameters to be tuned is also given in Table 1. Note that the remaining hyper parameters to be tuned are premultiplying factors (except for the α and β hyper parameters of the DC-kernel). In Chen et al. (2014), it is shown that the premultiplying hyper parameters can be tuned using sequential convex optimisation techniques. This could be applied here too, repeatedly tuning the premultiplying hyper parameters while keeping α and β fixed and, next, tuning α and β while keeping the premultiplying factors fixed.

The price to be paid for keeping these hyper parameters fixed is that the accuracy of the kernel model around the resonating frequencies is limited by the accuracy of the estimated resonating poles and residues, hence, limited by the performance of the LRM. However, this improves the robustness of the kernel-method. Being able to fix some hyper parameters in the kernel is a strong advantage.

Setting all hyper parameters free would mean a high number of parameters, and, hence, increasing the risk of divergence while minimising the negative marginal likelihood. Note that all premultiplying factors γ can separately be tuned by a sequential convex optimisation algorithm (Chen et al., 2014).

The complete modelling procedure is summarised by the flow chart in Fig. 1.

6. Simulation example

In this section we study the performance of the improved nonparametric method described in this paper on a simulation example. The three different alternatives for kernels carrying prior knowledge, see Section 4.2, are compared with the existing state of the art nonparametric frequency domain methods (Lataire & Chen, 2016) and Pintelon et al. (2021).

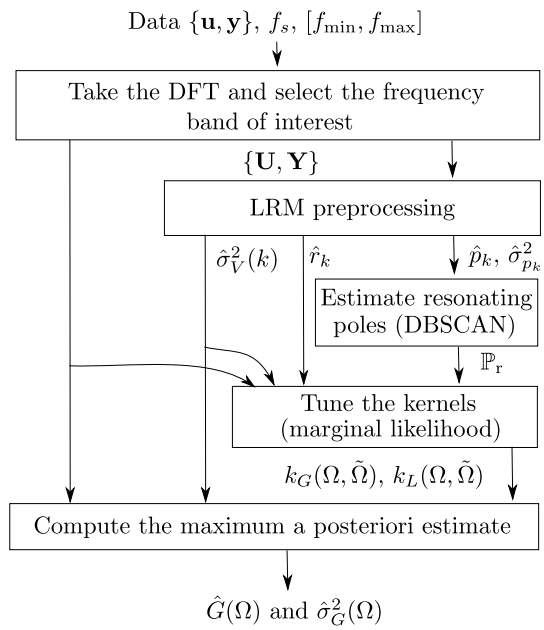

Fig. 1. Flow chart of the complete modelling procedure.

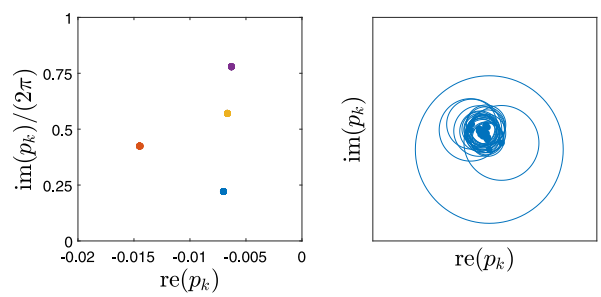

Fig. 2. Clustering of the estimated poles p_k from LRM for an SNR of 60 dB. The colours indicate the clusters. The right plot is a zoomed version of the blue area. The dots are the estimated poles and the circles represent 95% uncertainty bounds for these poles. x-limits: $-0.0070 \pm 5\ 10^{-5}$ rad/s, y-limits: $1.3882 \pm 5\ 10^{-5}$ rad/s. According to (45) we select from this set of blue dots the pole with the smallest circle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6.1. Simulated system and data

The simulated CT system G is given by the transfer function

$$G(s) = \frac{\Pi_{m=1}^{9}(s - z_m)}{\Pi_{n=1}^{13}(s - p_n)}, \tag{54}$$

with model order $n_a = 13$, and the pole and zero pairs shown in Table 2. Fig. 3(a) and (b) shows the frequency response function is shown in the band [0, 1] Hz. The input signal $u(t)$ is a random phase multisine with period $T = 1024$ s. The output signal is computed as (1), and corrupted by an additive noise signal $v(t)$, satisfying Assumption 1, with

$$H(s) = \frac{10.12}{s^2 + s + 10.12}. \tag{55}$$

In order to create strong transients, the signals are not measured for an entire period, but only for $t \in [24, 1024)$ s, resulting in a new measurement length of $T = 1000$ s. The input and output signals are sampled at a rate of $f_s = 10$ Hz, resulting in the data (6), with $N = 10^4$. We are only interested in the

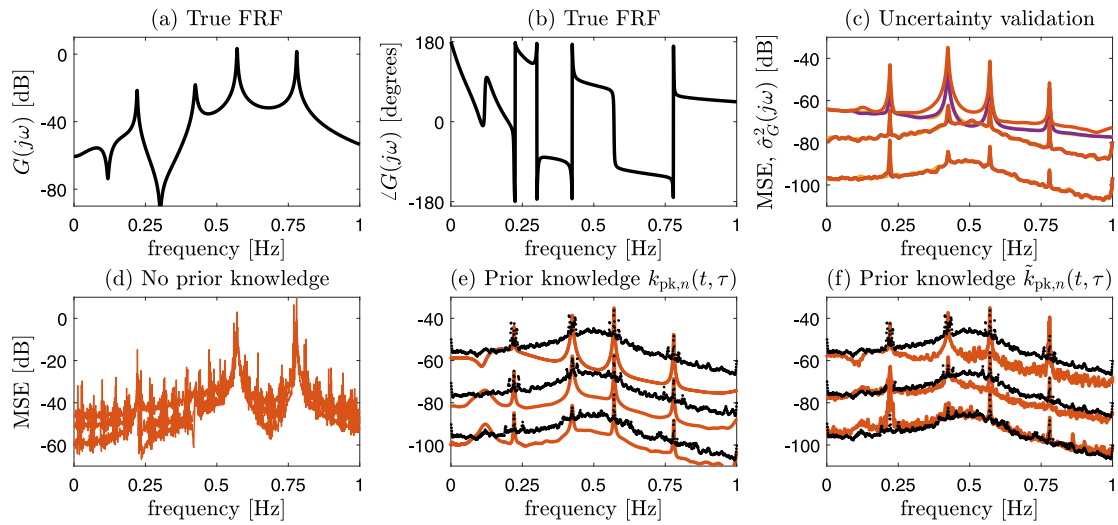

Fig. 3. Study of the performance of the improved kernel method for different SNR levels [20, 40 and 60 dB top to bottom in (c)–(f)]. The true FRF $G(j\omega)$ (54) is shown in (a) and (b). The estimated variance $\hat{\sigma}_G^2(j\omega)$ of the improved kernel method is validated in (c). Here we show results over 100 Monte-Carlo runs with independent noise realisations, where the hyper parameters are tuned in the first run and kept constant for the remaining 99 runs. MSE (red), sample variance (purple) and estimated variance (yellow, which coincides everywhere with purple). In (d)–(f), the improved kernel method is studied for the different kernels carrying prior knowledge: no prior knowledge, $k_{\mathrm{pk},n}(t,\tau)$ (27) and $\tilde{k}_{\mathrm{pk},n}(t,\tau)$ (29). 100 Monte-Carlo runs with independent noise realisations, where the hyper parameters are tuned for each run are performed. The MSE of the kernel method (red) is compared to the one of the LRM (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Poles p_n, their damping ratios, and zeros z_m of (54).

p_n	Damping	z_m
-2.8445	1	0.2647
$-0.2076 \pm 0.7393j$	0.2703	$-0.0139 \pm 0.7453j$
$-0.7302 \pm 0.9092j$	0.6262	$-0.0158 \pm 1.8977j$
$-0.0070 \pm 1.3883j$	0.0050	0.4359 ± 1.9303
$-0.0145 \pm 2.6667j$	0.0054	$-0.4178 \pm 2.0403j$
$-0.0067 \pm 3.5868j$	0.0019	
$-0.0063 \pm 4.8967j$	0.0013	

system's behaviour in the frequency band $[10^{-3}, 1]$ Hz, resulting in the frequency domain spectra $\mathbf{U}$ and $\mathbf{Y}$ (15) with $F_1 = 1$ and $F = 1000$.

6.2. Results

The results of the LRM preprocessing step and the estimated set $\mathbb{P}_{\mathrm{r}}$ of resonating poles are shown in Fig. 2. Four clusters are chosen by the DBSCAN algorithm, resulting in the estimated set of poles $\mathbb{P}_{\mathrm{r}}$.

The validation of the uncertainty $\hat{\sigma}_G^2(j\omega)$ is studied by means of a Monte-Carlo simulation over 100 runs with independent noise realisations. The hyper parameters are tuned in the first run and kept constant for the remaining 99 runs. Fig. 3(c) shows that the estimated variance does indeed coincide with the sample variance. Moreover, from the discrepancies between the MSE and the sample variance, one notices that the kernel method introduces a bias (Pillonetto et al., 2014). For the LRM, however, there is no bias (Pintelon et al., 2021).

In order to check the performance of the improved kernel method, it is compared to the kernel-based estimator (Lataire & Chen, 2016) and the LRM (Pintelon et al., 2021). Note that in Lataire and Chen (2016), the advantages of the frequency domain approach over the time domain Chen et al. (2012) have already been studied with simulations. 100 Monte-Carlo runs with independent noise realisations, and where the hyper parameters are tuned in each run, are performed. The results are shown in Fig. 3(d)–(f). Regarding the kernel method of Lataire and Chen (2016), see (d), one can see that the method fails, as the MSE is much higher than the one of LRM. The improved kernel method with prior knowledge kernel $k_{\mathrm{pk},n}(t,\tau)$ (27), performs better than the LRM for all SNR levels, see (e). The improved kernel method with extended DC-kernel $\tilde{k}_{\mathrm{pk},n}(t,\tau)$ (29), see (f), performs less good than the kernel $k_{\mathrm{pk},n}(t,\tau)$, still, for low SNRs, it performs better than the LRM and for high SNRs as good as the LRM. The extended TC-kernel $\check{k}_{\mathrm{pk},n}(t,\tau)$ (30) is not shown as it performs less good than the LRM for all SNRs. The kernel $k_{\mathrm{pk},n}(t,\tau)$ performs best as it also contains the information about the residues of the poles.

Moreover, note that the prior knowledge has a higher impact when the SNR is low. Nonetheless, at the resonance frequencies, the accuracy of the LRM and the improved kernel method are equal due to the fixed values of some hyper parameters. And, for high SNRs, one notices at the first transmission zero an increase of the MSE for the kernel methods. Even with the stable evaluation of the marginal likelihood (see Section 4.4), the kernel method is only suitable when the SNR is not too large. For this example, the ill-conditioning of $\mathbf{K}_{Y_{\mathrm{re}}}$ leads to a deteriorated estimate for SNRs higher than approximately 80 dB. To conclude, the kernel-method offers more accurate results than the LRM for the highly resonating simulated system while the simulations contained very strong transients.

7. Conclusions

In this paper, an improved nonparametric method has been developed for modelling LTI systems from noisy data. A kernel-based frequency domain approach is followed, offering important advantages, as detailed in the introduction and Section 3.

The heart of the kernel-based approach is the design of the kernel. For this method, the idea is introduced to superpose different kernels, each carrying prior knowledge about a particular resonating pole of the system. Different alternatives for kernels carrying prior knowledge have been studied. For keeping the amount of hyper parameters reasonable, prior knowledge

is obtained from an LRM preprocessing step: (i) the number of kernels needed, (ii) fixed values for some hyper parameters of the kernels and (iii) an estimate of the noise level as a function of the frequency. Note that the LRM decides whether a pole should be considered as resonating or not, depending on the SNR and frequency resolution of the measurement. As an application, lightly-damped systems can be identified with high accuracy from measurements disturbed by coloured noise. It is shown on simulations that this improved nonparametric method outperforms both the state of the art nonparametric frequency domain approaches (Pintelon et al., 2021) and (Lataire & Chen, 2016). Finally, we have proposed solutions for the most important bottlenecks of kernel-based methods: obtaining initial values for the hyper parameters before the tuning of the kernel functions and the bad numerical conditioning of the kernels.

In memory of Rik

Professor Rik Pintelon sadly passed away during the last review round of this article. Still, every word, except this part, has carefully been proofread by him. With this note the authors would like to thank him for all his motivated input to this work, and for his contributions to the field of system identification in general.

The idea that we could always turn to Rik with any matter, and that he could probably solve it too, was enormously comforting. Rik had the far-reaching competence of diagnosing problems and pointing out what was missing in our work. He was also known for his scientific integrity and for attaching a lot of importance to rigorous mathematical formalisms. For this article specifically, he was very keen on providing uncertainty expressions for the estimates. Let us stay true to these foundations.

We are so grateful for everything Rik has taught us, both as an academic and as a friend, may he rest easy.

We zullen je missen, Rik.

Appendix A. Continuous-time frequency domain kernels

$$k_{\mathrm{DC}}(j\omega, j\tilde{\omega}) = \frac{2\alpha + 2\beta + j(\omega - \tilde{\omega})}{(2\alpha + j(\omega - \tilde{\omega}))(\alpha + \beta + j\omega)(\alpha + \beta - j\tilde{\omega})}. \tag{A.1}$$

$$k_{\mathrm{PK},n}(j\omega, j\tilde{\omega}) = \frac{1}{2}\Big(\gamma_{1,n}\big(Q(\omega - \omega_n, \tilde{\omega} - \omega_n) + Q(\omega + \omega_n, \tilde{\omega} + \omega_n)\big) + \gamma_{2,n}\big(Q(\omega - \omega_n, \tilde{\omega} + \omega_n) + Q(\omega + \omega_n, \tilde{\omega} - \omega_n)\big)\Big) \tag{A.2}$$

$$\tilde{k}_{\mathrm{PK},n}(j\omega, j\tilde{\omega}) = \frac{\gamma_n}{2}\big(R(\omega_n) + R(-\omega_n)\big) \tag{A.3}$$

with

$$Q(x, y) = \frac{1}{(\alpha_n + jx)(\alpha_n - jy)} \tag{A.4}$$

$$R(\omega_n) = \frac{2\alpha_n + 2\beta_n + j(\omega - \tilde{\omega})}{(2\alpha_n + j(\omega - \tilde{\omega}))(\alpha_n + \beta_n + j(\omega - \omega_n))} \times \frac{1}{\alpha_n + \beta_n - j(\tilde{\omega} - \omega_n)}. \tag{A.5}$$

$\check{k}_{\mathrm{PK},n}(j\omega, j\tilde{\omega})$ is obtained by setting $\beta_n = \alpha_n$ in (A.3).

Appendix B. Discrete-time frequency domain kernels

$$k_{\mathrm{DC}}(e^{\frac{j\omega}{f_s}}, e^{\frac{j\tilde{\omega}}{f_s}}) = \frac{1 - e^{-(2\alpha + 2\beta + j(\omega - \tilde{\omega}))T_s}}{(1 - e^{-(2\alpha + j(\omega - \tilde{\omega}))T_s})(1 - e^{-(\alpha + \beta + j\omega)T_s})} \times \frac{1}{1 - e^{-(\alpha + \beta - j\tilde{\omega})T_s}}. \tag{B.6}$$

$k_{\mathrm{PK},n}(e^{\frac{j\omega}{f_s}}, e^{\frac{j\tilde{\omega}}{f_s}})$ is given by (A.2), but with

$$Q(x, y) = \frac{1}{(1 - e^{-(\alpha_n + jx)T_s})(1 - e^{-(\alpha_n - jy)T_s})}. \tag{B.7}$$

$\tilde{k}_{\mathrm{PK},n}(e^{\frac{j\omega}{f_s}}, e^{\frac{j\tilde{\omega}}{f_s}})$ is given by (A.3), but with

$$R(\omega_n) = \frac{1 - e^{-(2\alpha_n + 2\beta_n + j(\omega - \Omega))T_s}}{(1 - e^{-(2\alpha_n + j(\omega - \Omega))T_s})(1 - e^{-(\alpha_n + \beta_n + j(\omega - \omega_n))T_s})} \times \frac{1}{1 - e^{-(\alpha_n + \beta_n - j(\Omega - \omega_n))T_s}}. \tag{B.8}$$

$\check{k}_{\mathrm{PK},n}(e^{j\omega}, e^{j\tilde{\omega}})$ is found by setting $\beta_n = \alpha_n$ in $\tilde{k}_{\mathrm{PK},n}(e^{j\omega}, e^{j\tilde{\omega}})$.

Appendix C. Proof of Theorem 1

Proof. We separate the matrix $\mathbf{K}_{Y_{\mathrm{re}}}$ into two matrices A and B, such that A is a diagonal matrix with the noise variances and B comprises the remaining:

$$\mathbf{K}_{Y_{\mathrm{re}}} = \frac{1}{2}(A + B). \tag{C.9}$$

The sum of A and B can then be written as

$$\begin{aligned} A + B &= A^{1/2}(I_{2F} + W\Sigma W^T)A^{1/2} \\ &= A^{1/2}W(I_{2F} + \Sigma)W^T A^{1/2}. \end{aligned} \tag{C.10}$$

where W is an orthogonal matrix. The inverse of the sum yields

$$(A + B)^{-1} = (A^{-1/2}WS^{-1/2})(A^{-1/2}WS^{-1/2})^T \tag{C.11}$$

with the diagonal matrix $S = I_{2F} + \Sigma$ and where we used that $W^{-1} = W^T$. Finally we compute

$$\mathbf{K}_{Y_{\mathrm{re}}}^{-1} = 2MM^T \quad \text{with} \quad M = A^{-1/2}WS^{-1/2}. \quad \square \tag{C.12}$$

Appendix D. Proof of Theorem 2

Proof. We separate the matrix $\mathbf{K}_{Y_{\mathrm{re}}}$ as in the previous theorem and compute,

$$\begin{aligned} \det \mathbf{K}_{Y_{\mathrm{re}}} &= \frac{1}{2^{2F}}\det(A + B) \\ &= \frac{1}{2^{2F}}\det(A^{1/2}WSW^TA^{1/2}) \\ &= \frac{1}{2^{2F}}\det{}^2(A^{1/2})\det S \\ &= \frac{1}{2^{2F}}\Big(\prod_{k=1}^{F}\sigma_V^2(k)\Big)^2\prod_{k=1}^{2F}(1 + \sigma_k), \end{aligned} \tag{D.13}$$

where the third equality uses $\det(W) = 1$. Taking the logarithm of (D.13) gives (35). $\square$

References

Birpoutsoukis, G., Marconato, A., Lataire, J., & Schoukens, J. (2017). Regularized nonparametric Volterra kernel estimation. *Automatica*, *82*, 324–327.

Chen, T. (2018a). Continuous-time DC kernel—a stable generalized first-order spline kernel. *IEEE Transactions on Automatic Control*, *63*(12), 4442–4447.

Chen, T. (2018b). On kernel design for regularized LTI system identification. *Automatica*, *90*, 109–122.

Chen, T., Andersen, M. S, Ljung, L., Chiuso, A., & Pillonetto, G. (2014). System identification via sparse multiple kernel-based regularization using sequential convex optimization techniques. *IEEE Transactions on Automatic Control*, *59*(11), 2933–2945.

Chen, T., Ohlsson, H., & Ljung, L. (2012). On the estimation of transfer functions, regularizations and Gaussian processes—Revisited. *Automatica*, *48*(8), 1525–1535.
Hallemans, N., Lataire, J., & Pintelon, R. (2020). Nonparametric identification of linear time-varying systems using Gaussian process regression. *IFAC-PapersOnLine*, *53*(2), 1001–1006.
Hallemans, N., & Pintelon, R. (2021). Relation between Gaussian process regression and the regularised linear weighted least squares. Internal note.
Hallemans, N., Pintelon, R., Peumans, D., & Lataire, J. (2021). Improved frequency response function estimation by Gaussian process regression with prior knowledge. *IFAC-PapersOnLine*, *54*(7), 559–564.
Hjalmarsson, H., Gevers, M., Gunnarsson, S., & Lequin, O. (1998). Iterative feedback tuning: theory and applications. *IEEE Control Systems Magazine*, *18*(4), 26–41.
Kwakernaak, H., & Sivan, R. (1991). *Modern signals and systems*. Prentice-Hall.
Lataire, J., & Chen, T. (2016). Transfer function and transient estimation by Gaussian process regression in the frequency domain. *Automatica*, *72*, 217–229.
Levy, E. (1959). Complex-curve fitting. *IRE Transactions on Automatic Control*, (1), 37–43.
Liu, J., & Ciucci, F. (2020). The Gaussian process distribution of relaxation times: A machine learning tool for the analysis and prediction of electrochemical impedance spectroscopy data. *Electrochimica Acta*, *331*, 1–21.
Ljung, L. (1999). *System identification: Theory for the user*. Prentice-Hall.
Ljung, L., Chen, T., & Mu, B. (2020). A shift in paradigm for system identification. *International Journal of Control*, *93*(2), 173–180.
MacKay, D. J. C. (1998). *NATO ASI series F computer and systems sciences*: *vol. 168*, *Introduction to Gaussian processes* (pp. 133–166). Citeseer.
McKelvey, T., & Guérin, G. (2012). Non-parametric frequency response estimation using a local rational model. *IFAC-PapersOnLine*, *45*(16), 49–54.
Peumans, D., Busschots, C., Vandersteen, G., & Pintelon, R. (2018). Improved FRF measurements of lightly damped systems using local rational models. *IEEE Transactions on Instrumentation and Measurement*, *67*(7), 1749–1759.
Pillonetto, G., Chiuso, A., & De Nicolao, G. (2011). Prediction error identification of linear systems: a nonparametric Gaussian regression approach. *47*(2), 291–305.
Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., & Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, *50*(3), 657–682.
Pintelon, R., Guillaume, P., & Schoukens, J. (2007). Uncertainty calculation in (operational) modal analysis. *Mechanical Systems and Signal Processing*, *21*(6), 2359–2373.
Pintelon, R., Peumans, D., Vandersteen, G., & Lataire, J. (2021). Frequency response function measurements via local rational modeling, revisited. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–16.
Pintelon, R., & Schoukens, J. (2012a). FRF measurement of nonlinear systems operating in closed loop. *IEEE Transactions on Instrumentation and Measurement*, *62*(5), 1334–1345.
Pintelon, R., & Schoukens, J. (2012b). *System identification: A frequency domain approach*. John Wiley & Sons.
Pintelon, R., Schoukens, J., & Guillaume, P. (2006). Continuous-time noise modeling from sampled data. *IEEE Transactions on Instrumentation and Measurement*, *55*(6), 2253–2258.
Pintelon, R., Schoukens, J., Vandersteen, G., & Barbé, K. (2010). Estimation of nonparametric noise and FRF models for multivariable systems-part I: Theory. *Mechanical Systems and Signal Processing*, *24*(3), 573–595.
Rabiner, L. R., & Gold, B. (1975). *Theory and application of digital signal processing*. Prentice-Hall.
Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT press Cambridge.
Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (1998). Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Mining and Knowledge Discovery*, *2*(2), 169–194.
Scandella, M., Mazzoleni, M., Formentin, S., & Previdi, F. (2020). Kernel-based identification of asymptotically stable continuous-time linear dynamical systems. *International Journal of Control*, 1–29.
Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*, *42*(3), 1–21.
Stoddard, J. G., Birpoutsoukis, G., Schoukens, J., & Welsh, J. S (2019). Gaussian process regression for the estimation of generalized frequency response functions. *Automatica*, *106*, 161–167.
The MathWorks Inc. (2021). System identification toolbox. Natick, Massachusetts, United State.
Zorzi, M., & Chiuso, A. (2018). The harmonic analysis of kernel functions. *Automatica*, *94*, 125–137.

Noël Hallemans was born in Brussels, Belgium, in 1996. He obtained his master's degree in electrical engineering in 2019 from the Vrije Universiteit Brussel (VUB) and Université Libre de Bruxelles (ULB). Since September 2019 he has been working as a researcher with the ELEC department, VUB, Brussels. His research focuses on data-driven modelling, including linear approximations and kernel-based regression.

Rik Pintelon (M'90-SM'96-F'98) was born in Gent, Belgium, on December 4, 1959. He received a master's degree in electrical engineering in 1982, a doctorate (Ph.D) in engineering in 1988, and the qualification to teach at university level (geaggregeerde voor het hoger onderwijs) in 1994, all from the Vrije Universiteit Brussel (VUB), Brussels, Belgium. In 2014 he received the degree of Doctor of Science (DSc) from the University of Warwick (UK) for publications with the collective title "Frequency Domain System Identification: A Mature Modeling Tool". From 1982 to 1984 and 1986 to 2000, Dr. Pintelon was a researcher with the Belgian National Fund for Scientific Research (FWO-Vlaanderen) at the Electrical Engineering (ELEC) Department of the VUB. From 1984 to 1986 he did his military service overseas in Tunesia at the Institut National Agronomique de Tunis. From 1991 to 2000 he was a part-time lecturer at the department ELEC of the VUB, and since 2000 he is a full-time professor in electrical engineering at the same department. From 2009 to 2018 he was a visiting professor at the department of Computer Sciences of the Katholieke Universiteit Leuven, and from 2013 to 2018 he was a honorary professor in the School of Engineering of the University of Warwick. His main research interests include system identification, signal processing, and measurement techniques. Dr. Pintelon is the coauthor of 4 books on System Identification and the coauthor of more than 260 articles in refereed international journals. He has been a Fellow of IEEE since 1998. Dr. Pintelon was the recipient of the 2012 IEEE Joseph F. Keithley Award in Instrumentation and Measurement (IEEE Technical Field Award). He received the 2008 IOP outstanding paper award (best paper in Measurement Science & Technology), the 2014 Martin Black prize (best paper in Physiological Measurement) and the 2014 Andy Chi award (best paper in IEEE Trans. on Instrumentation and Measurement).

Boris Joukovsky received a Master degree in Electrical Engineering in 2019 from the Université Libre de Bruxelles (ULB) and the Vrije Universiteit Brussel (VUB), Belgium. Since 2019, he is a Ph.D. student with the Department of Electronics and Informatics (ETRO), VUB, under the supervision of Prof. Nikos Deligiannis. In 2020, he obtained a Ph.D. Strategic Basic Research Fellowship from the Research Foundation Flanders (FWO), Belgium. His research focuses on interpretable deep learning with applications in image and video processing, as well as theoretical machine learning concepts and convex optimisation.

Dries Peumans (1992, Belgium) received the M.Sc. degree and the Ph.D. degree in Electrical Engineering (Electronics and Information Technology) from the Vrije Universiteit Brussel (VUB), Belgium, in 2015 and 2020, respectively. His current research interests include the analysis, modelling and measurement of complex (non)linear systems in the frequency domain. His current research interests include the analysis, modelling, and measurement of complex (non)linear systems in the frequency domain.

John Lataire (S'06-M'11) was born in Brussels, Belgium, in 1983. He received the Electrical Engineer degree in electronics and information processing and the Ph.D. degree in engineering sciences (Doctor in de Ingenieurswetenschappen) from the Vrije Universiteit Brussel, Brussels, in 2006 and 2011, respectively. From October 2007 to October 2011, he was on a Ph.D. fellowship from the Research Foundation—Flanders (FWO). Since August 2006, he has been working as a Researcher with the Department ELEC-VUB, Brussels. Dr. Lataire is the coauthor of more than 30 articles in refereed international journals. He received the 2008 IOP outstanding paper award (best paper

in Measurement Science & Technology), the Best Junior Presentation Award 2010 at the 29th Benelux Meeting on Systems and Control, was the co-recipient of the 2014 Andy Chi award (best paper in IEEE Trans. on Instrumentation and Measurement), and was the recipient of the 2016 J. Barry Oakes Advancement Award (from the IEEE Instrumentation and Measurement society). His main interests include the frequency domain formulation of the identification of dynamic systems, with a specific focus on the identification of time-varying systems, and the use of kernel-based regression in system identification.