Copyright © 201X Inderscience Enterprises Ltd.

Model Selection for Servo Control Systems

Mathias Tantau

Institute of Mechatronic Systems,
Leibniz University Hannover,
An der Universität 1, 30823, Garbsen, Germany
Tel: +49 511 762-17844 E-mail: mathias.tantau@imes.uni-hannover.de

Lars Perner

Lenze Automation GmbH,
Am Alten Bahnhof 11, Germany
E-mail: lars.perner@lenze.com

Mark Wielitzka

Institute of Mechatronic Systems,
Leibniz University Hannover,
An der Universität 1, 30823, Garbsen, Germany
Tel: +49 511 762-5718 E-mail: mark.wielitzka@imes.uni-hannover.de

Abstract: Physically motivated models of electromechanical motion systems are required in several applications related to control design, model-based fault detection and simply interpretation. Often, however, the high effort of modelling prohibits these model-based methods in industrial applications. Therefore, all approaches of automatic modelling / model selection are naturally appealing. An intuitive approach is to identify the parameters of several model candidates and to select the one with the best fit on unseen data. A shortcoming of this approach is that the chosen model may be one with high complexity in which some of the physically interpretable parameters are not practically identifiable and uncertain. Also, ambiguities in selecting the model structure would not be disguised resulting in false confidence in a chosen model. Designing a reasonable set of candidate models requires that distinguishability of models can be checked prior to the identification procedure.

This paper proposes a strategy for frequency domain model selection. The resulting model is tailored to ensure practical identifiability of all parameters for the given excitation. The analysis is based on local sensitivity calculated for the frequency domain cost function. Also, the paper describes distinguishability analysis of candidate models utilizing transfer function coefficients and Markov parameters. Model selection and distinguishability analysis are applied to a class of models as they are commonly used to describe servo control systems. It is shown in experiments on an industrial stacker crane that model selection works with little user interaction, except from defining normalized hyperparameters and ensuring that the resulting model is sound. Distinguishability is analysed systematically for all models that result from rearranging actuator, sensor and spring-damper elements along a chain of discrete masses. It can be proved or disproved for almost all combinations of potential models.

Keywords: model selection; structure and parameter identification; frequency domain; distinguishability analysis; equivalence of structures; multiple mass resonators; servo control system; electromechanical motion systems; transfer function approach; Markov parameter approach.

Reference to this paper should be made as follows: Tantau, M. et al. (2020) 'Model Selection for Servo Control Systems', *International Journal of Mechatronics and Automation*, Vol. x, No. x, pp.xxx–xxx.

1 Introduction

Physically motivated models of servo control systems are the basis of several applications that require knowledge of the system's internal structure. Examples are control design and auto-tracking, feed-forward, model-based fault diagnosis, Kalman-filtering, and simulations. Sometimes the inner structure of a given system is also quested for understanding and interpretation.

Because modelling is time-consuming and requires considerable expert knowledge and expertise in multi-disciplinary fields (Lakhoua et al. (2020)), the usage of model-based methods is difficult, especially in products of automation industry that are built only in small quantities, e.g. storage and retrieval systems, positioning systems. Approaches towards automatic modelling are of interest as they could leverage these problems and provide objective decisions. Automatic model selection is also called structure and parameter identification or model structure identification (Stigter and Beck (1994)).

While many previous works on servo system identification in the sense of identifying the parameters of a given model exist, e.g. Chen et al. (2002); Villwock (2007), works on model selection are mainly limited to data-driven, static models (Hoeting et al. (1999)) of other disciplines such as biology (Volinsky et al. (1996)) and finance (Draper (1995)). Here, transfer function models of servo systems are identified in frequency domain allowing an intuitive comparison with the measurements and avoiding the difficulty to automatize time domain simulations, see Tantau et al. (2019, 2020). The challenge with these models is that linearity in the physical parameters is generally not maintained. Extensions of identifiability tests to nonlinear models exist (Brun et al. (2001); Gábor et al. (2017)), but they have rarely been applied to the frequency domain.

An analysis of practical identifiability in frequency domain can be found in Bizeray et al. (2018), where contour lines of the cost function are interpreted. A limited model selection in the field of dynamic models for servo control applications is performed in Schütte et al. (1997), but the process is not fully automatic and leaves the final decision for the commissioning engineer.

As a first step in this direction we intend to search for the model structure that best describes the input-output behaviour of a system measured against unseen data, while considering exclusion criteria of practical identifiability in frequency domain. The latter ensures that only those models are included, that allow identification of model parameters with some certainty so that interpretability is not compromised by randomly interspersed extra parameters without physical meaning. This approach can be seen as a counterpart to optimizing the excitation for a given model, see for example Pronzato and Pázman (2013), which has limitations for nonlinear models in combination with limited prior knowledge. Here, the model is tailored to a given excitation.

A problem with selecting the model with the lowest residual, as described by Burnham and Anderson (1998), is that of indistinguishability / indiscriminability of structures, i.e. different models show the exactly same input-output behaviour for certain parameterizations, while the internal structure is different. No experiments exist to discriminate between them. If this problem is ignored, the model selection algorithm will output only one model although others could be equally true, leading to false interpretations, inappropriate feed-forward design, etc.

There are two approaches to handle the indistinguishability problem. One is to identify all models and then to compare the model predictions. Two models with similar predictions in a stochastic framework can be judged indistinguishable (Diard (2009); Burnham and Anderson (1998)). The second approach is to perform an a-priori analysis of the model equations in order to prove (in)distinguishability deterministically. Only the latter is followed in this paper. The advantage is a reduced number of identification runs in the structure and parameter identification. Additionally, false certainty in the model selection can be avoided.

Definitions of indistinguishability are not reviewed but can be found in Vajda (1981); Walter et al. (1984); Godfrey and Chapman (1989); Chapman and Godfrey (1989); Avdeenko and Kargin (2000); Rosa and Silvestre (2011). Here, the term structural indistinguishability is understood as follows: Two model structures $M(\cdot)$ and $\hat{M}(\cdot)$ with parameter values $\boldsymbol{p}_0 \in \Omega$ and $\hat{\boldsymbol{p}}_0 \in \hat{\Omega}$ and the same number of inputs and outputs are called *equivalent* if the corresponding system outputs are identical for all admissible input trajectories. Ω and $\hat{\Omega}$ are sets of possible parameter values, not necessarily of the same dimension. If for all $\boldsymbol{p}_0 \in \Omega$ there is at least one equivalent parameter value $\hat{\boldsymbol{p}}_0 \in \hat{\Omega}$ and vice versa, except possibly for a set of zero measure, then the two systems $M(\boldsymbol{p})$ and $\hat{M}(\hat{\boldsymbol{p}})$ are structurally indistinguishable (Vajda (1981)). For notational simplicity we call systems, which are not structurally indistinguishable, distinguishable. Note that strictly speaking systems can be nether of them (Raksanyi et al. (1985)).

Methodologically, one determines a minimal representation of the system called the *exhaustive summary* (Walter et al. (1984)) or the *structural invariant vector* (Vajda (1981)) of the model. This set of algebraic or differential equations is then investigated for the existence of solutions by the help of elimination theory and computer algebra (Raksanyi et al. (1985); Godfrey et al. (1994)). For example in Zhang et al. (1991) necessary and sufficient conditions for the existence of solution in equations are given that have recently been integrated into a web application for distinguishability studies (Davidson et al. (2017)).

Methods for generating the exhaustive summary include the transfer function or Laplace transform approach (Vajda (1981)) and the similarity transform approach (Avdeenko and Kargin (2000)) for linear systems, as well as the time-power series or Taylor series approach (Pohjanpalo (1978)), and the generating series approach (Raksanyi et al. (1985); Walter and Pronzato (1996); Motchon et al. (2017)) for nonlinear systems. For nonlinear but rational models differential algebra has been used in Meshkat et al. (2018).

In this work the Markov parameter approach for generating the exhaustive summary which has been applied to strict distinguishability investigations

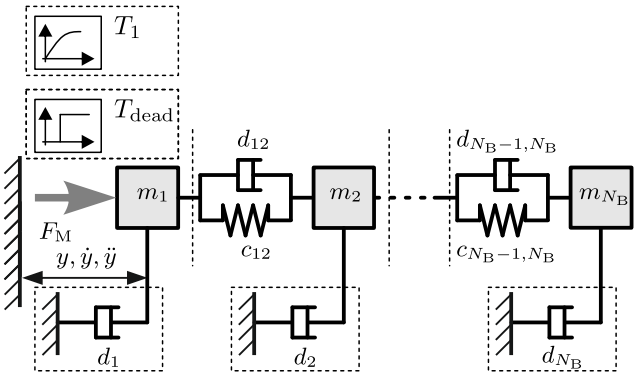


Figure 1: Class of candidate models

in Motchon et al. (2016) will be applied to the distinguishability question as defined above, utilizing the conditions from Vajda and Rabitz (1988); Zhang et al. (1991); Davidson et al. (2017), see below. Its effectiveness is compared with the transfer function approach in the application of model selection for servo systems. Existing works are mainly limited to compartmental models (Yates et al. (2009)) in biomedicine (Godfrey et al. (1994); Walter and Pronzato (1996); Evans et al. (2004)), although the question is important in many fields.

2 MODEL SELECTION AND PRACTICAL IDENTIFIABILITY

In sections 2 and 3 the objective is to carry out a model selection for a limited set of candidate models as explained next. In sections 4 and 5 the set of candidate models is broadened in order to analyse alternative possible formulations of the problem in view of distinguishability.

2.1 Candidate models

Only a limited variety of multiple-mass models is chosen as candidate models as shown in Fig. 1. In general, the candidate models are composed of submodels as indicated by the dashed lines with 0 or more estimation parameters. Only the purely translational case is shown in the figure but the purely rotary case would work equivalently. All models consist of a chain of N_B elastically coupled masses $m_1, m_2, \cdots, m_{N_B}$ of which the first element is actuated. The force of the driving motor F_M may be subject to input delay time T_{dead}. In addition, the torque control dynamics are either neglected or modelled as a P_{T1} element. As the number of masses in the chain varies the spring-damper elements between adjacent masses will also appear / disappear, but there is never a spring-damper element in front of the first mass. In order to incorporate the effect of bearing and guideway friction dampers between masses and the environment can be included for each existing mass.

What results is a single input single output (SISO) transfer function (TF) model so that in the following only the SISO case is considered. This set of candidate models could be extended easily, e.g. by springs between distant masses, but including a large set of candidate models without justification through experience and prior knowledge could be seen critically as data dredging / mining) (Burnham and Anderson (1998); Chatfield (1995)), even if distinguishability is generally given. The question of structural distinguishability will be discussed in Sec. 5.

2.2 Model selection

In the process of model selection an exhaustive search over all combinations of submodels is carried out. More elegant procedures such as genetic programming are avoided here in order to reduce the complexity and the effect of coincidence. For each model the parameters are optimized by matching the calculated and measured frequency response in an equation error formulation. The exact cost function is given in the next section. Then, the best model is chosen based on its merit as explained next.

Depending on the intended purpose of a model different ways to determine its merit can be defined. Reasonable criteria are the minimization of the Kullback-Leibler distance to the 'true model' by means of the Akaike information criterion or cross validation, a test for whiteness of the residuals, a test for cross-correlation between the inputs and the residuals, a χ^2 significance test of the cost function and others (Pintelon and Schoukens (2012)). However, the problem with all these approaches is that they are based on the stochastic nature of measurements and they seek for the true model. If the repeatability is high, which can always be achieved by averaging several measurements under similar conditions (Chatfield (1995)), very complex models would result with possibly superfluous parameters. Since physical interpretability of model parameters is the prerequisite in this work, it is reasonable to check *practical identifiability* of all model parameters. Practical identifiability means that all model parameters can be identified accurately from sparce, noisy data for a given excitation (Vu (2015)). Criteria for checking practical identifiability are given below. Only models with all parameters practically identifiable are kept in the set of candidate models, see next sections. Among the models that fulfill these criteria the one with the best fit on an independent test data set is chosen.

2.3 Sensitivity calculation

The notion of practical identifiability based on sensitivity comes from the parameter-linear model

$$\boldsymbol{S}_{\text{lin}}\boldsymbol{p} = \boldsymbol{y} \tag{1}$$

with sensitivity matrix $\boldsymbol{S}_{\text{lin}} \in \mathbb{R}^{N_y \times N_p}$, parameter vector $\boldsymbol{p} \in \mathbb{R}^{N_p \times 1}$, and system outputs $\boldsymbol{y} \in \mathbb{R}^{N_y \times 1}$. In the

least-squares sense the cost function $J_{\mathrm{lin}} = (\boldsymbol{S}_{\mathrm{lin}}\boldsymbol{p} - \boldsymbol{y}_{\mathrm{m}})^{\mathrm{T}}(\boldsymbol{S}_{\mathrm{lin}}\boldsymbol{p} - \boldsymbol{y}_{\mathrm{m}})$ is minimized in order to determine optimal parameters, where $\boldsymbol{y}_{\mathrm{m}}$ is the measured output. In the study of practical identifiability the sensitivity matrix $\boldsymbol{S}_{\mathrm{lin}}$ is checked for (multi-)collinearity and sensitivity of the parameters as explained later.

Transfer function models are not in general linear in the parameters. In the nonlinear case $\boldsymbol{g} : \mathbb{R}^{N_{\mathrm{P}}} \to \mathbb{R}^{N_y}$

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{p}) \tag{2}$$

the cost function $J \in \mathbb{R}$ is of the form:

$$J = (\boldsymbol{y} - \boldsymbol{y}_{\mathrm{m}})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{y}_{\mathrm{m}}) = ||\boldsymbol{\Lambda}(\boldsymbol{y} - \boldsymbol{y}_{\mathrm{m}})||_2^2 . \tag{3}$$

$\boldsymbol{W} \in \mathbb{R}^{N_y \times N_y}$ is a weighting matrix, often chosen as the inverse covariance matrix of the measurements assuming zero mean, Gaussian noise, maximum the likelihood. The decomposition $\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda} = \boldsymbol{W}$ exists if $\boldsymbol{W}$ is positive definite (Pintelon and Schoukens (2012)). Motivated by the Taylor series expansion (Brun et al. (2001))

$$\boldsymbol{g}(\boldsymbol{p}) = \boldsymbol{g}(\boldsymbol{p}_0) + \left.\frac{\partial \boldsymbol{g}(\boldsymbol{p})}{\partial \boldsymbol{p}^T}\right|_{\boldsymbol{p}=\boldsymbol{p}_0} (\boldsymbol{p} - \boldsymbol{p}_0) + \cdots \tag{4}$$

a substitute for $\boldsymbol{S}_{\mathrm{lin}}$ can be found:

$$\boldsymbol{S}_{\mathrm{nl}} = \boldsymbol{\Lambda}\left.\frac{\partial \boldsymbol{g}(\boldsymbol{p})}{\partial \boldsymbol{p}^T}\right|_{\boldsymbol{p}=\boldsymbol{p}_0} . \tag{5}$$

The weighting $\boldsymbol{\Lambda}$ should be considered in the analysis of practical identifiability in agreement with the penalty function (3) as the weights also influence the result. $\boldsymbol{S}_{\mathrm{nl}} \in \mathbb{R}^{N_y \times N_{\mathrm{P}}}$ characterizes the effect of changing a parameter locally.

In frequency domain identification of dynamical systems the output vector elements are associated to the set of measured frequency components and possibly also to several elements in the transfer function matrix of a MIMO system. Here, only SISO systems are considered. The observations and the noise are complex, which gives several options to formulate the penalty function. In the sequel, different exemplary formulations are given together with the implications for calculating the sensitivity matrix.

Case 1: The cost function is the distance in the complex plane between model $\boldsymbol{G}(\boldsymbol{p}) \in \mathbb{R}^{N_{\mathrm{f}} \times 1}$ and measurement $\boldsymbol{G}_{\mathrm{M}} \in \mathbb{R}^{N_{\mathrm{f}} \times 1}$ for each of the N_{f} frequency components (Pintelon and Schoukens (2012); Galarza et al. (1995)). This can be written in complex notation:

$$J = (\boldsymbol{G}(\boldsymbol{p}) - \boldsymbol{G}_{\mathrm{M}})^{\mathrm{H}}\,\boldsymbol{W}\,(\boldsymbol{G}(\boldsymbol{p}) - \boldsymbol{G}_{\mathrm{M}}), \tag{6}$$

with $\boldsymbol{W} \in \mathbb{R}^{N_y \times N_y}$ or in real notation:

$$J = \left|\left|\boldsymbol{\Lambda}_{\mathrm{r/i}}\begin{bmatrix}\mathrm{Re}\,\{\boldsymbol{G}(\boldsymbol{p}) - \boldsymbol{G}_{\mathrm{M}}\}\\ \mathrm{Im}\,\{\boldsymbol{G}(\boldsymbol{p}) - \boldsymbol{G}_{\mathrm{M}}\}\end{bmatrix}\right|\right|_2^2 . \tag{7}$$

$\boldsymbol{\Lambda}_{\mathrm{r/i}} \in \mathbb{R}^{2N_{\mathrm{f}} \times 2N_{\mathrm{f}}}$ can be defined by the square root of the inverse sample covariance matrix of the real measurement vector or if not available as the unity matrix. Assuming that real and imaginary part at each spectral line of the observation are uncorrelated having equal variances, matrix $\boldsymbol{\Lambda}_{\mathrm{r/i}}$ has the form

$$\begin{pmatrix}\boldsymbol{\Lambda}_{\mathrm{r}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Lambda}_{\mathrm{i}}\end{pmatrix} \quad \text{with} \quad \boldsymbol{\Lambda}_{\mathrm{r}} = \boldsymbol{\Lambda}_{\mathrm{i}} \in \mathbb{R}^{N_{\mathrm{f}} \times N_{\mathrm{f}}} \tag{8}$$

and matrix $\boldsymbol{W}$ is given by $2\boldsymbol{\Lambda}_{\mathrm{r}}^{\mathrm{T}}\boldsymbol{\Lambda}_{\mathrm{r}}$ due to variance summation of uncorrelated variables (Pintelon and Schoukens (2012)).

The latter cost function formulation is favoured here because the resulting sensitivity matrix $\boldsymbol{S}_1 \in \mathbb{R}^{2N_{\mathrm{f}} \times N_p}$ is real and all criteria of practical identifiability can readily be applied:

$$\boldsymbol{S}_1 = \boldsymbol{\Lambda}_{\mathrm{r/i}}\begin{bmatrix}\boldsymbol{S}_{\mathrm{r}}\\ \boldsymbol{S}_{\mathrm{i}}\end{bmatrix}, \tag{9}$$

$$\boldsymbol{S}_{\mathrm{r}} = \frac{\mathrm{dRe}\,\{\boldsymbol{G}(\boldsymbol{p})\}}{\mathrm{d}\boldsymbol{p}} = \mathrm{Re}\left\{\frac{\mathrm{d}\boldsymbol{G}(\boldsymbol{p})}{\mathrm{d}\boldsymbol{p}}\right\}, \tag{10}$$

$$\boldsymbol{S}_{\mathrm{i}} = \frac{\mathrm{dIm}\,\{\boldsymbol{G}(\boldsymbol{p})\}}{\mathrm{d}\boldsymbol{p}} = \mathrm{Im}\left\{\frac{\mathrm{d}\boldsymbol{G}(\boldsymbol{p})}{\mathrm{d}\boldsymbol{p}}\right\}. \tag{11}$$

It can easily be verified that exchanging the decomposition into real and imaginary part, and the derivative operator in (10,11) is possible, because the complex unit can be treated as a prefactor in the derivative. The right side of (10,11) allows to calculate the real and imaginary part of the sensitivity numerically after the complex derivative has been calculated symbolically by the help of computer algebra, which simplifies the calculations. Still, for very complex models even $\mathrm{d}\boldsymbol{G}/\mathrm{d}\boldsymbol{p}$ cannot be calculated symbolically.

Case 2: In Schütte (2003) the following cost function is used for multiple mass systems because of its superior robustness:

$$J = \left|\left|\boldsymbol{\Lambda}_{\mathrm{A/P}}\begin{bmatrix}|\boldsymbol{G}(\boldsymbol{p})| - |\boldsymbol{G}_{\mathrm{M}}|\\ \angle\{\boldsymbol{G}(\boldsymbol{p})\} - \angle\{\boldsymbol{G}_{\mathrm{M}}\}\end{bmatrix}\right|\right|_2^2 . \tag{12}$$

The problem is the arbitrary weighting of amplitude and phase if the inverse covariance is not used for weighting.

The sensitivity matrix is given by:

$$\boldsymbol{S}_2 = \boldsymbol{\Lambda}_{\mathrm{A/P}}\begin{bmatrix}\boldsymbol{S}_{\mathrm{A2}}\\ \boldsymbol{S}_{\mathrm{P2}}\end{bmatrix}, \tag{13}$$

$$\boldsymbol{S}_{\mathrm{A2},kj} = \frac{1}{|\boldsymbol{G}_k|}\left[\mathrm{Im}\,\{\boldsymbol{G}_k\}\,\boldsymbol{S}_{\mathrm{i},kj} + \mathrm{Re}\,\{\boldsymbol{G}_k\}\,\boldsymbol{S}_{\mathrm{r},kj}\right],$$

$$\boldsymbol{S}_{\mathrm{P2},kj} = \frac{1}{|\boldsymbol{G}_k|^2}\left[\mathrm{Re}\,\{\boldsymbol{G}_k\}\,\boldsymbol{S}_{\mathrm{i},kj} - \mathrm{Im}\,\{\boldsymbol{G}_k\}\,\boldsymbol{S}_{\mathrm{r},kj}\right].$$

$\boldsymbol{G}_k$ denotes the k-th spectral component of the modelled TF. As it turns out, it is possible to write this Jacobian matrix in terms of $\boldsymbol{S}_{\mathrm{r}}, \boldsymbol{S}_{\mathrm{i}}$. So, whenever these two can be calculated from analytic expressions in case 1, numeric differentiation is not necessary for $\boldsymbol{S}_2$ either.

Case 3: In the third case to consider the logarithm of the amplitude is evaluated instead of the amplitude itself:

$$J = \left|\left|\boldsymbol{\Lambda}_{\mathrm{A/P}}\begin{bmatrix}\log_{10}\{|\boldsymbol{G}(\boldsymbol{p})|\} - \log_{10}\{|\boldsymbol{G}_{\mathrm{M}}|\}\\ \angle\{\boldsymbol{G}(\boldsymbol{p})\} - \angle\{\boldsymbol{G}_{\mathrm{M}}\}\end{bmatrix}\right|\right|_2^2 . \tag{14}$$

The advantage is that in the log scale a reasonable weighting of amplitude and phase can be defined. The penalty depends less on the absolute size of the amplitudes, which eliminates the dependence on testbed properties. For example π and 1 can be chosen as weightings for amplitude and phase, respectively. Then 20 dB difference in the amplitude has an equivalent effect as 180° in the phase.

The sensitivity matrix can be calculated symbolically, if analytic expressions exist for $\boldsymbol{S}_\mathrm{r}, \boldsymbol{S}_\mathrm{i}$:

$$\boldsymbol{S}_3 = \boldsymbol{\Lambda}_{\mathrm{A/P}} \begin{bmatrix} \boldsymbol{S}_{\mathrm{A3}} \\ \boldsymbol{S}_{\mathrm{P3}} \end{bmatrix}, \tag{15}$$

$$\boldsymbol{S}_{\mathrm{A3},kj} = \frac{1/\ln(10)}{|\boldsymbol{G}_k|^2} \left[\mathrm{Im}\left\{\boldsymbol{G}_k\right\} \boldsymbol{S}_{\mathrm{i},kj} + \mathrm{Re}\left\{\boldsymbol{G}_k\right\} \boldsymbol{S}_{\mathrm{r},kj} \right],$$

$$\boldsymbol{S}_{\mathrm{P3},kj} = \frac{1}{|\boldsymbol{G}_k|^2} \left[\mathrm{Re}\left\{\boldsymbol{G}_k\right\} \boldsymbol{S}_{\mathrm{i},kj} - \mathrm{Im}\left\{\boldsymbol{G}_k\right\} \boldsymbol{S}_{\mathrm{r},kj} \right].$$

As said before, cases 2 and 3 tend to find the correct model parameters for multiple mass models easier and weighting is more robust in the third case. Therefore, in the following only the third case, (14) is used. A potential problem is, however, that the measured and simulated phase must be interpolated continuous beyond $\pm 2\pi$.

2.4 Assessing practical identifiability

Criteria for assessing practical identifiability are not reviewed here but can be found in Farrar and Glauber (1967); Marquardt (1970); Stewart et al. (1987); Belsley (1991); Burnham and Anderson (1998); Brun et al. (2001); Kovács et al. (2005); Nienałtowski et al. (2015); Gábor et al. (2017). Only the criteria actually used to exclude models form the model selection are introduced briefly.

For a model structure to be valid all parameters must have a certain minimal importance according to the **msqr parameter importance index (PII)** given by Brun et al. (2001):

$$\delta_j^{\mathrm{msqr}} = \left|\left| \boldsymbol{S}_j \Delta p_j \right|\right|_2, \tag{16}$$

with $\boldsymbol{S}_j$ the column of the sensitivity matrix corresponding to parameter j. If this criterion is violated, at least one of the parameters is of little importance and cannot be identified precisely. Also, the model is not parsimonious. The normalization constant Δp_j is supposed to be in the unit of parameter p_j, for example the nominal value or a quarter of the range of reasonable values (Brun et al. (2001)). Here, the previously identified parameter value is chosen for Δp_j. Defining the threshold for δ_j^{msqr} is somewhat arbitrary. Gábor et al. (2017) set it to four orders of magnitude below the maximum PII.

Furthermore, the **collinearity index** γ_k:

$$\gamma_k = \frac{1}{\sqrt{\lambda_k}}, \tag{17}$$

reveals (multi-)collinearity among parameters. λ_k is the smallest singular value of $\tilde{\boldsymbol{S}}$, that is the matrix $\boldsymbol{S}$ with all columns normalized to unit length. This normalization ensures a clear separation of collinearity and sensitivity, measured by the PIIs. Collinearity is critical if γ_k exceeds 5...20 (Brun et al. (2001)).

Alternatively, collinearity can be measured by the **scaled condition indices** by Belsley (1991):

$$\tilde{\eta}_k = \frac{\mu_{\max}}{\mu_k}. \tag{18}$$

μ_k and $\mu_{\max}$ are the k-th and the maximal singular value of $\tilde{\boldsymbol{S}}$, respectively. According to Belsley (1991) large values are critical, above 10...30.

In the experiments of the next section a combination of these three exclusion criteria is used for model selection. The exact thresholds will be given together with the results.

3 Experimental results on model selection

Two testbeds serve the experimental validation of the structure and parameter identification. Since these results are highly dependent on the chosen thresholds, the subsequent section focusses on these dependencies.

3.1 Structure and parameter optimization

The structure optimization is applied to the two testbeds shown in Fig. 2. Testbed 1 is a stacker crane with 5.6 m mast height and 5 m length of the horizontal axis (x) along the shelf. All experiments are carried out on x with the vertical axis in a position of 2 m. Testbed 2 has only one axis which is driven in direct drive. In Figs. 3 and 4 the FRFs of training and test are shown. They have both been recorded with stepped sine excitation, but different amplitudes.

Parameter ranges have been chosen 0...1000 for stiffnesses, and 0...0.1 for damping constants and moments of inertias (physical units). If more prior knowledge is available, the ranges can be narrowed. Parameters are identified with particle swarm optimization which is parameterized with a number of particles that equals 200 times the number of parameters of the current model. In the structure identification a maximum of 4 masses are considered (120 different models) for which the calculation takes approx. seven hours (for a maximum of 3 masses it takes only 35 min), implemented in Matlab on an i7 4-core computer, running at 3.7 GHz with 16 GB DDR4 RAM. The overproportionally long calculation time for 4-mass models can be explained by the fact that they often require numerical sensitivity matrix calculation due to TF complexity.

In Figs. 3 and 4 the best model is also shown, including asymptotes. The chosen thresholds for the three criteria are given in the first row of each section in Tab. 1. For testbed 1 a 3-mass system with additional damper to the base at mass 2 and delay time results, 9 estimation parameters. For testbed 2 it is a 4-mass

(a) 1: Stacker crane

(b) 2: Linear positioning system

Figure 2: Experimental testbeds

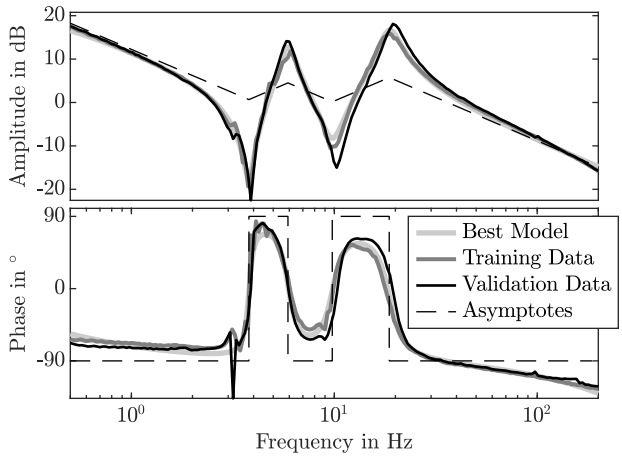


Figure 3: Frequency response functions of testbed 1, measurement and best model

system with additional damper at mass 1 and delay time, 12 estimation parameters.

The effect of changing the thresholds has been studied for two alternative settings, see rows two and three in Tab. 1. Testbed 1 is always assigned the same model, while for testbed 2 different models result (3-mass system with additional damper at mass 3 and delay time, 9 estimation parameters, for the second set of thresholds and a 2-mass system with additional damper at mass 2 and delay time, 6 estimation parameters, for the third set). The corresponding FRFs are not shown. So especially for testbed 2, which cannot be categorized visually very clearly, the algorithmic result also depends strongly on the thresholds.

3.2 Threshold dependencies

To further investigate the dependence on thresholds of the identifiability criteria each of the three criteria is applied once exclusively in Figs. 5 and 6. Its threshold is varied in steps from 1 to 10000. For each step the model with the lowest cost J is selected among all models that fulfil this criterion. The number of estimation parameters of the best model for this threshold is plotted. For

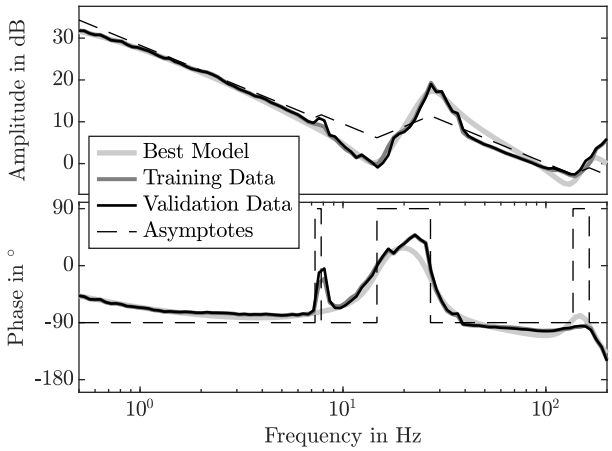


Figure 4: Frequency response functions of testbed 2, measurement and best model

Table 1 Three chosen settings for the identifiability criteria and corresponding best model for both testbeds. The first row in each section corresponds to the first set of settings and so on.

Criterion	Threshold	Testbed 1	Testbed 2
Max. scaled	30	6.4	11.3
condition	10	6.4	10.3
index	10	6.4	2.6
Max. PII / Min. PII	1000	10.2	12.0
	1000	10.2	7.7
	1000	10.2	7.5
Max.	20	4.0	7.0
collinearity	10	4.0	7.0
index	5	4.0	1.9

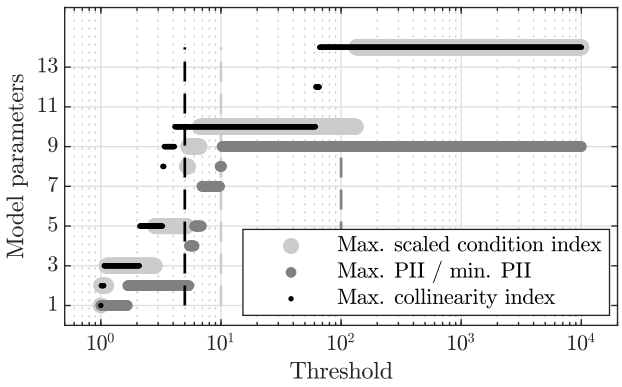


Figure 5: Number of estimation parameters of the best model for a given threshold of the one current criterion, testbed 1

reference, the thresholds of the first setting in Tab. 1 are indicated by dashed lines.

Clearly, a strong dependence on the thresholds exists and mostly the number of parameters increases as the threshold increases. Exceptions exist where models with fewer parameters are more critical regarding one of the criteria than models with one or two more parameters.

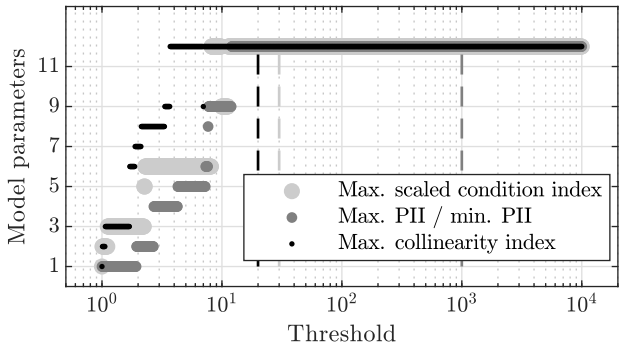


Figure 6: Number of estimation parameters of the best model for a given threshold of the one current criterion, testbed 2

What is disguised by this figure is the interplay between the three criteria. E.g. a model may be highly collinear, while all parameters have approx. the same sensitivity or vice versa. Thus, it can happen that the resulting model is very simple although all criteria applied exclusively allow for more complex models. The combined result would be hard to visualize in a 2D plane. Also, it is not known how many models are ruled out by a certain criterion and a given threshold. In order to investigate this Figs. 7 and 8 show the number of models that fulfil a certain criterion depending on the chosen threshold out of all 120 models.

The graphs of testbed 1 show a clear edge at approx. 60 models. Only if the thresholds are chosen below that, the number of valid models diminishes steeply. As expectable from Tab. 1, testbed 2 shows a more gradual slope, making the choice of a threshold more critical.

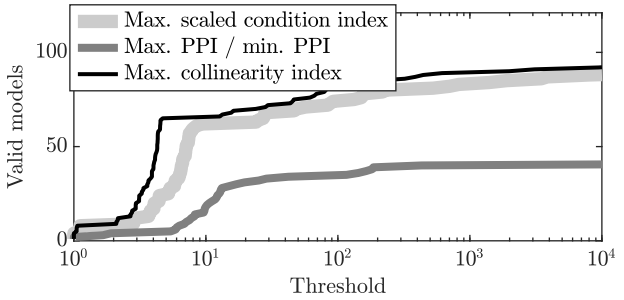


Figure 7: Number of models that suffice the considered criterion depending on the threshold, testbed 1

4 Distinguishability analysis

The second part of this work deals with distinguishability analysis in order to investigate if the set of candidate models is chosen well so that a unique structure identification is possible and if it could even be extended to a larger set of models (Sec. 5). In this section the general methodology is explained.

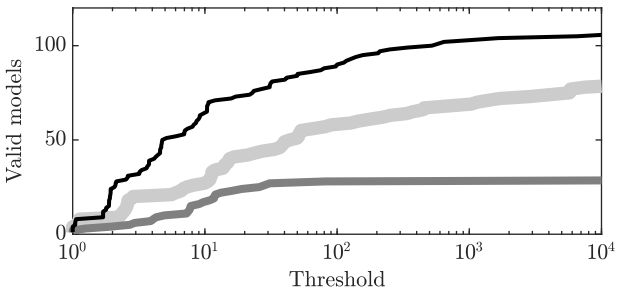


Figure 8: Number of models that suffice the considered criterion depending on the threshold, testbed 2, legend as in Fig. 7

Starting point is the SISO state space form:

$$\dot{\boldsymbol{x}} = \boldsymbol{A}_\mathrm{S}\boldsymbol{x} + \boldsymbol{B}_\mathrm{S}u, \qquad y = \boldsymbol{C}_\mathrm{S}\boldsymbol{x} + D_\mathrm{S}u. \tag{19}$$

Here, u is the input, $\boldsymbol{x} \in \mathbb{R}^{N_\mathrm{S}}$ are the states and y is the output. $\boldsymbol{A}_\mathrm{S}$, $\boldsymbol{B}_\mathrm{S}$, $\boldsymbol{C}_\mathrm{S}$ and D_S are analytic functions in the parameters $\boldsymbol{p} \in \mathbb{R}^{N_\mathrm{P}}$. Initial conditions are assumed to be zero so that models cannot be discriminated from these.

The state space form for an N_B mass system can be obtained from the mass matrix $\boldsymbol{M}$, the damping matrix $\boldsymbol{D}$ and the stiffness matrix $\boldsymbol{C}$:

$$\boldsymbol{A}_\mathrm{S} = \begin{pmatrix} \mathbf{0}_{N_\mathrm{B}\times N_\mathrm{B}} & \mathbf{I}_{N_\mathrm{B}\times N_\mathrm{B}} \\ -\boldsymbol{M}^{-1}\boldsymbol{C} & -\boldsymbol{M}^{-1}\boldsymbol{D} \end{pmatrix}, \tag{20}$$

$$\boldsymbol{B}_\mathrm{S}' = \begin{pmatrix} \mathbf{0}_{N_\mathrm{B}\times N_\mathrm{B}} \\ \boldsymbol{M}^{-1} \end{pmatrix}, \tag{21}$$

$$\boldsymbol{C}_S' = \begin{pmatrix} \mathbf{I}_{2N_\mathrm{B}\times 2N_\mathrm{B}} \\ (\mathbf{0}_{N_\mathrm{B}\times N_\mathrm{B}}\ \mathbf{I}_{N_\mathrm{B}\times N_\mathrm{B}})\boldsymbol{A}_\mathrm{S}' \end{pmatrix}, \tag{22}$$

$$\boldsymbol{D}_\mathrm{S}' = \begin{pmatrix} \mathbf{0}_{2N_\mathrm{B}\times N_\mathrm{B}} \\ (\mathbf{0}_{N_\mathrm{B}\times N_\mathrm{B}}\ \mathbf{I}_{N_\mathrm{B}\times N_\mathrm{B}})\boldsymbol{B}_\mathrm{S}' \end{pmatrix}. \tag{23}$$

The symbol $\mathbf{0}$ stands for the zero matrix. So far, the input of the resulting MIMO system consists of N_B forces. The output vector contains N_B positions, N_B velocities and N_B accelerations. The actual actuator location, resp. sensor type (position, velocity, acceleration) and location are defined by input and output multiplication, leading to the desired matrices without prime $'$ in (19). Here, only SISO system are considered.

For distinguishability analysis of these models it seems that the transfer function approach is the most straightforward method to generate the exhaustive summary, because it natively eliminates irrelevant information related to the non-unique representation via states. Their arbitrary assignment can lead to a large number of equivalent models (Zhang et al. (1991)). The similarity transform approach could become infeasible for models with many states (Godfrey and Chapman (1989)). Furthermore, the Markov parameter approach can be applied as outlined below.

In this section necessary and sufficient conditions for distinguishability analysis of linear SISO state space models based on the TF are reviewed. Additional criteria

and considerations for MIMO systems can be found in Zhang et al. (1991); Davidson et al. (2017).

4.1 Criteria for structural indistinguishability based on the TF

For the transfer function approach (TFA) the state space representation of each system is reduced to a TF assuming zero initial conditions ($\mathbf{I}$ is the unity matrix):

$$G(s) = \boldsymbol{C}_\mathrm{S}(\mathbf{I}s - \boldsymbol{A}_\mathrm{S})^{-1}\boldsymbol{B}_\mathrm{S} + D_\mathrm{S}. \tag{24}$$

It has the general form:

$$G(s) = \frac{b_0 + b_1 s^1 + \cdots + b_n s^n}{a_0 + a_1 s^1 + \cdots + a_{m-1}s^{m-1} + s^m}. \tag{25}$$

The coefficients $a_0, a_1, \ldots, a_{m-1}$ are analytic functions of the model parameters $\boldsymbol{p}$. All common factors have been cancelled and one of the coefficients is fixed, for example the denominator coefficient corresponding to the highest power of s can be normalized to 1. Then the coefficients define the finite set of moment invariants $\boldsymbol{\Phi}$ for the given system and can be used as structural invariant vector (Zhang et al. (1991)). The constant coefficient can optionally be included into the moment invariants, but it has no use. So, the number of moment invariants N_MI is either $n + m + 1$ or $n + m + 2$.

Distinguishability of two models could be checked by equating the complete set of moment invariants of both systems. The resulting set of equations can be complex to solve, but the following necessary and sufficient conditions for structural indistinguishability supercede these calculations in many cases. Conditions 1 to 5 are necessary, i.e. whenever one of them is violated, distinguishability is proved and the remaining conditions can be skipped. The conditions should be evaluated in the given order because they are sorted in increasing computational complexity and because dependencies exist:

- Cond. 1: The same number of moment invariants N_MI exist.
- Cond. 2: The same number of coefficients exist in numerator n and denominator m in both models.
- Cond. 3: Both sets of moment invariants have the same symbolic form. The symbolic form is a representation of the moment invariants with zeros for constant elements and ones for non-constant (depending on $\boldsymbol{p}$, resp. $\hat{\boldsymbol{p}}$) elements (Vajda (1984)).
- Cond. 4: The rank N_R of the Jacobian matrix $\boldsymbol{J} = \partial\boldsymbol{\Phi}/\partial\boldsymbol{p}$ of the moment invariants with respect to the parameters is the same for both models (Zhang et al. (1991); Vajda and Rabitz (1988)).
- Cond. 5: The same *linear dependencies* among moment invariants exist (Zhang et al. (1991)). These are determined by creating all possible combinations of N_R rows of $\boldsymbol{J}$, leading to as many as $\binom{N_\mathrm{MI}}{N_\mathrm{R}}$ submatrices. Those combinations of rows with a full row rank of the resulting reduced matrix in symbolic form are noted as largest sets of linearly independent rows and must be identical for the compared systems.

As a note to condition 5: The largest sets of linearly independent rows define the linear dependencies among moment invariants fully. Searching for linear dependencies among smaller subsets of $\boldsymbol{J}$ with less than N_R rows would not reveal any new information: If a smaller subset is included in a larger set of linearly independent rows, it will be linearly independent because eliminating rows will maintain independence of the remaining rows. If a smaller subset is not contained in any of the known larger sets, it will be linearly dependent. This is because the set of N_R rows contains all combinations, except those with linear dependencies. Consequently, if a subset is not included, it must be linearly dependent. Simplifications can be made by excluding constant moment invariants (which correspond to zero rows).

For two model that satisfy the above necessary conditions, the following sufficient conditions for structural indistinguishability can be checked.

- Cond. 6: A renaming and permutation scheme for the parameters of the first model exists so that the moment invariants are identical to those of the second model in symbolic form and vice versa.
- Cond. 7: The two models have as many determinable parameters as non-constant moment invariants and the constant moment invariants are identical across the two models. The number of determinable parameters equals the generic rank of the Jacobian $\boldsymbol{J}$ (Vajda (1984)).

Strictly speaking, these conditions are valid only for certain open sets of the parameter space. Possible parameter values are not explicitly considered in this paper. Examples are non-negative parameters or prior knowledge on possible ranges of physical parameters.

4.2 The applicability of (in)distinguishability criteria depends on the system's representation

If none of the above conditions can be applied, no statement can be made in the given representation of the model, but there is a chance that transforming the model into a different representation helps to answer the question of distinguishability. This is shown in the following minimal example with complex number models $z_1, z_2 \in \mathbb{C}$:

$$z_1 = a_1 b_1 + 0 \cdot b_1 i, \quad z_2 = a_2 b_2 (1 + 1i) \tag{26}$$

with the model parameters $a_1, b_1, a_2, b_2 \in \mathbb{R}^+$ and the complex unit i. Clearly, these two models are

distinguishable. Magnitude and phase are now chosen for the moment invariants, as they describe a complex number uniquely:

$$\begin{array}{ll} \boldsymbol{\Phi}_1: & |z_1| = a_1 b_1, \quad \angle z_1 = 0, \\ \boldsymbol{\Phi}_2: & |z_2| = \sqrt{2} a_2 b_2, \quad \angle z_2 = \pi/4. \end{array} \tag{27}$$

Cond 1, 3, 4, 5, 6, 7 are applied to the example (cond. 2 works only with TFs). Both models have a total of $N_{\mathrm{MI}} = 2$ moment invariants (cond. 1) with the symbolic form $\begin{bmatrix} 1\,0 \end{bmatrix}$ (cond 3). The Jacobian matrices are:

$$\boldsymbol{J}_1 = \begin{bmatrix} b_1\, a_1 \\ 0 \;\; 0 \end{bmatrix}, \boldsymbol{J}_2 = \begin{bmatrix} \sqrt{2} b_2\, \sqrt{2} a_2 \\ 0 \qquad 0 \end{bmatrix}. \tag{28}$$

The rank of each Jacobian matrix is $N_{\mathrm{R1}} = N_{\mathrm{R2}} = 1$ (cond. 4). So, for cond. 5 sets of linearly independent rows of size one row must be found. In both models this is the first row only. Cond. 6, 7 cannot be applied. As a result, distinguishability could not be proved.

Next the representation in real and imaginary part is tested. With this set of moment invariants cond. 3 is not satisfied, because symbolic form $\begin{bmatrix} 1\,0 \end{bmatrix}$ for model z_1 differs from $\begin{bmatrix} 1\,1 \end{bmatrix}$ for model z_2. The Jacobian matrices are:

$$\boldsymbol{J}_1 = \begin{bmatrix} b_1\, a_1 \\ 0 \;\; 0 \end{bmatrix}, \boldsymbol{J}_2 = \begin{bmatrix} b_2\, a_2 \\ b_2\, a_2 \end{bmatrix}. \tag{29}$$

It follows that evaluating cond. 5 for z_1 will result in a set of rows consisting only of row 1, while for z_2 both row 1 and 2 have a rank of 1. So, the models can be distinguished by this rule.

The fact that it depends on the chosen representation if distinguishability can be shown with cond. 1 to 7 motivates the analysis of other formulations than the TF for dynamic systems. In the next section the Markov parameter approach (MPA) is introduced and in Sec. 5 its benefit is investigated.

4.3 Markov parameter approach

For system (19) an alternative representation is based on an expansion of the TF as a complex power series in s with infinitely many terms. This is a structural invariant vector of infinite length (Vajda (1981)):

$$G(s) = G_0 + G_1 s^{-1} + G_2 s^{-2} + \dots. \tag{30}$$

The set of $\{G_i\}_{i=0,1,2,\dots}$ can be referred to the step response $h(t)$, resp. to the impulse response of the system $g(t)$ and its time derivatives (Hatakeyama et al. (1999)):

$$G_0 = h(t)|_{t=0^+}, G_i = \left.\frac{\mathrm{d}^{i-1}}{\mathrm{d}t^{i-1}} g(t)\right|_{t=0^+}, i = 1, 2, \dots. \tag{31}$$

$\{G_i\}_{i=1,2,3\dots}$ are called Markov parameters. From the model (19) these terms can be calculated leading to analytic functions in $\boldsymbol{p}, \hat{\boldsymbol{p}}$ (Hatakeyama et al. (1999)):

$$G_0 = D_{\mathrm{S}}, \quad G_i = \boldsymbol{C}_{\mathrm{S}} \boldsymbol{A}_{\mathrm{S}}^{i-1} \boldsymbol{B}_{\mathrm{S}}, \quad i = 1, 2, \dots. \tag{32}$$

For distinguishability analysis of two models of the kind (19) with N_{S1}, resp. N_{S2} states it is sufficient to consider the first $N_{\mathrm{S1}} + N_{\mathrm{S2}}$ Markov parameters plus the term G_0 as moment invariants. If they are identical, then the higher-degree Markov parameters must also be identical. This can be explained as follows.

An augmented system is defined (Motchon et al. (2016)):

$$\dot{\boldsymbol{x}}_{\mathrm{a}} = \boldsymbol{A}_{\mathrm{a}} \boldsymbol{x}_{\mathrm{a}} + \boldsymbol{B}_{\mathrm{a}} u, \qquad y = \boldsymbol{C}_{\mathrm{a}} \boldsymbol{x}_{\mathrm{a}} + D_{\mathrm{a}} u \tag{33}$$

with

$$\boldsymbol{x}_{\mathrm{a}} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \quad \boldsymbol{A}_{\mathrm{a}} = \begin{bmatrix} \boldsymbol{A}_{\mathrm{S1}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathrm{S2}} \end{bmatrix}, \quad \boldsymbol{B}_{\mathrm{a}} = \begin{bmatrix} \boldsymbol{B}_{\mathrm{S1}} \\ \boldsymbol{B}_{\mathrm{S2}} \end{bmatrix},$$

$$\boldsymbol{C}_{\mathrm{a}} = \begin{bmatrix} \boldsymbol{C}_{\mathrm{S1}} - \boldsymbol{C}_{\mathrm{S2}} \end{bmatrix}, \quad D_{\mathrm{a}} = D_{\mathrm{S1}} - D_{\mathrm{S2}}.$$

As a change of notation indices 1 and 2 stand for the two systems to compare with N_{S1}, resp. N_{S2} states. Note that matrix $\boldsymbol{A}_{\mathrm{a}}$ is of dimension $N_{\mathrm{S1}} + N_{\mathrm{S2}} \times N_{\mathrm{S1}} + N_{\mathrm{S2}}$. Clearly, the output of the augmented system is the difference between the two single systems and the same holds for the structural invariants, as defined in (32):

$$G_{\mathrm{a},k} = G_{1,k} - G_{2,k} \quad \forall k \geq 0. \tag{34}$$

According to the Cayley-Hamilton theorem for any square matrix $\boldsymbol{X}$ of size $m \times m$, $k \in \mathbb{N}$ it holds that

$$\boldsymbol{X}^m + \lambda_{m-1} \boldsymbol{X}^{m-1} + \dots + \lambda_1 \boldsymbol{X} + \lambda_0 \mathbf{I} = \boldsymbol{0}, \tag{35}$$

in which λ_i are the coefficients of the characteristic polynomial of $\boldsymbol{X}$. This can be applied to the system matrix, while also premultiplying $\boldsymbol{C}_{\mathrm{a}}$ and postmultiplying $\boldsymbol{B}_{\mathrm{a}}$ (Phan et al. (1998)):

$$\begin{aligned} \boldsymbol{C}_{\mathrm{a}} \boldsymbol{A}_{\mathrm{a}}^n \boldsymbol{B}_{\mathrm{a}} + \lambda_{n-1} \boldsymbol{C}_{\mathrm{a}} \boldsymbol{A}_{\mathrm{a}}^{n-1} \boldsymbol{B}_{\mathrm{a}} + \dots \\ + \lambda_1 \boldsymbol{C}_{\mathrm{a}} \boldsymbol{A} \boldsymbol{B}_{\mathrm{a}} + \lambda_0 \boldsymbol{C}_{\mathrm{a}} \boldsymbol{B}_{\mathrm{a}} = \boldsymbol{0}. \end{aligned}$$

Accordingly, $\boldsymbol{C}_{\mathrm{a}} \boldsymbol{A}_{\mathrm{a}}^n \boldsymbol{B}_{\mathrm{a}}$ and higher orders can be written as a linear combination of lower-order Markov parameters. This means that $\boldsymbol{C}_{\mathrm{a}} \boldsymbol{A}_{\mathrm{a}}^k \boldsymbol{B}_{\mathrm{a}} = 0 \;\; \forall \;\; k \in [0, N_{\mathrm{S1}} + N_{\mathrm{S2}} - 1]$ implies $\boldsymbol{C}_{\mathrm{a}} \boldsymbol{A}_{\mathrm{a}}^k \boldsymbol{B}_{\mathrm{a}} = 0 \;\; \forall \;\; k \geq N_{\mathrm{S1}} + N_{\mathrm{S2}}$ and according to (32), (34) it implies that both systems have identical Markov parameters.

Often, only the first few Markov parameters, less than $N_{\mathrm{S1}} + N_{\mathrm{S2}}$, can be calculated due to computational complexity.

Utilizing cond. 1 to 7 from above to investigate distinguishability with the MPA, it must be noted that cond. 1 is obsolete, since the number of Markov parameters, resp. $G_{1,i}, G_{2,i}$ terms should be chosen identical. Cond. 2 is not applicable because the distinction between numerator and denominator does not exist. The sufficient conditions (6, 7) cannot be used if only a subset of the Markov parameters has been calculated, less than $N_{\mathrm{S1}} + N_{\mathrm{S2}}$. It will be investigated in section 5.2 if the MPA can make worthy contributions to the analysis in spite of these limitations.

4.4 Criteria based on submodels

Two sufficient conditions for structural indistinguishability on certain open sets of the parameter space are given in Vajda and Rabitz (1988) that can be evaluated without calculation by inspecting the model structures visually:

- Cond 8: One model is a submodel of the other and both have the same number of determinable parameters. Submodel means that model structures are identical except for missing elements such as springs, dampers and masses.
- Cond 9: The two models are submodels of a third model and all three models have the same number of determinable parameters.

Further easy to evaluate criteria based on the model structure are the often cited geometrical rules. They are not applied here, because they make restrictive assumptions, e.g. controllability, observability, invariant matrices $\boldsymbol{B}_\mathrm{S}, \boldsymbol{C}_\mathrm{S}$ over the compared models, ... (Godfrey and Chapman (1990)). The servo system models considered in this paper are observable only in the case of position sensors and often the matrices $\boldsymbol{B}_\mathrm{S}$ and $\boldsymbol{C}_\mathrm{S}$ have different structures.

5 Results of the distinguishability analysis

Rather than directly investigating the distinguishability of the candidate models from Fig. 1 a more general class of multiple mass models is considered as they are created by placing actuator and sensor in different positions. Figure 1 represents a special case with single-sided force input and collocated measurement. However, depending on the construction it might be necessary in some cases to consider elasticity of the base and at other positions. So, the aim of the structure identification could be to determine at which point structural compliances and masses are significant, or can be neglected. This information is useful for detecting weaknesses in the construction and faults that occur in operation. The more general class of models is derived in Sec. 5.1, before the criteria for distinguishability are applied in Sec. 5.2.

5.1 Combinatory generation of multiple mass models

Attention is restricted to linear chains of masses without loops or branching and with only one actuator and one sensor. An example is shown in Fig. 9. Translational elements are displayed, but the chain could equally well consist of rotary elements.

For a given number N_B of masses the complete set of possible models can be created by rearranging the sensor, the actuator and the position in the chain where the spring is missing (gap), representing the degree of freedom of the positioner/motion system.

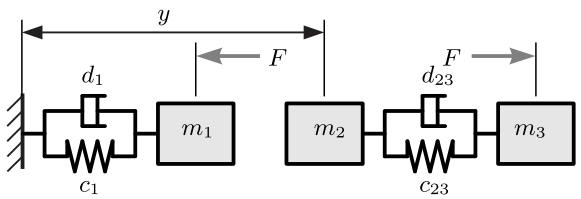


Figure 9: Example of multipe mass model with $N_\mathrm{B} = 3$, $n_\mathrm{gap} = 1$, $n_{\mathrm{act}+} = 3$, $n_{\mathrm{act}-} = 1$, $n_{\mathrm{meas}+} = 2$, $n_{\mathrm{meas}-} = 0$

It is located between masses $n_\mathrm{gap} \in \mathbb{N}$ and $n_\mathrm{gap} + 1$. Actuator and sensor have one point of actuation $n_{\mathrm{act}+} \in \mathbb{N}$, resp. measurement $n_{\mathrm{meas}+} \in \mathbb{N}$ and a reference point $n_{\mathrm{act}-} \in \mathbb{N}$, $n_{\mathrm{meas}-} \in \mathbb{N}$. An index of 0 refers to the rigid environment instead of a mass. The set of possible models can be constructed by varying these five indices under the following constraints:

$$\begin{cases} 0 \le n_\mathrm{gap} \le N_\mathrm{B} - 1, \\ 0 \le n_{\mathrm{act}-} \le n_\mathrm{gap}, \\ n_\mathrm{gap} + 1 \le n_{\mathrm{act}+} \le N_\mathrm{B}, \\ 0 \le n_{\mathrm{meas}-} \le n_\mathrm{gap}, \\ n_\mathrm{gap} + 1 \le n_{\mathrm{meas}+} \le N_\mathrm{B}. \end{cases} \tag{36}$$

In addition, two restrictions are necessary to exclude models with unexcited masses or states that have no output sensitivity for any type of sensor. Such models would depict a degradation to a simpler model, independent of the parameters:

$$\begin{cases} n_\mathrm{gap} = 0 \vee n_{\mathrm{act}-} > 0, \\ n_\mathrm{gap} = 0 \vee n_{\mathrm{meas}-} > 0. \end{cases} \tag{37}$$

The first condition ensures excitability of all masses, while the second ensures sensitivity of the measurement to all masses. In the example in Fig. 9 the second condition is violated: Mass 1 has no effect on the position measurement.

Finally, models are excluded that are created from other models by renaming/renumbering of some elements. This trivial way of generating additional models is prevented by demanding the following restrictions:

$$\begin{cases} n_{\mathrm{act}+} \le n_{\mathrm{meas}+}, \\ n_{\mathrm{act}+} \le n_\mathrm{gap} + \lceil (N_\mathrm{B} - n_\mathrm{gap})/2 \rceil. \end{cases} \tag{38}$$

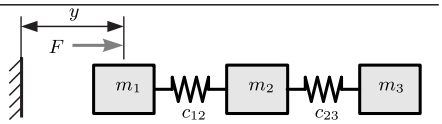
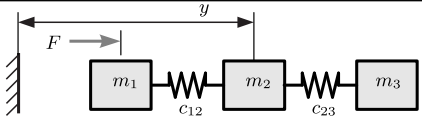
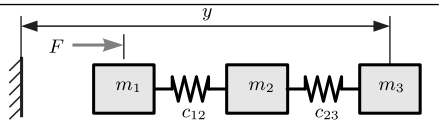
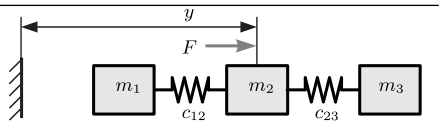
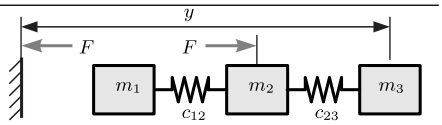
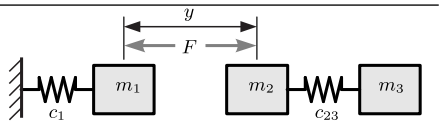
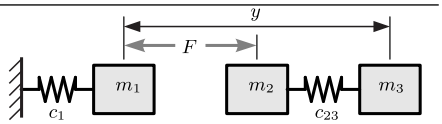
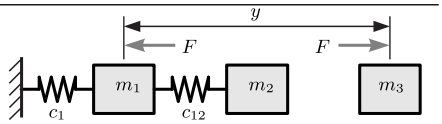
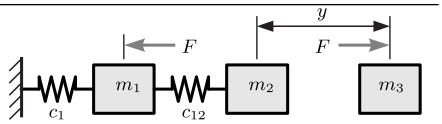
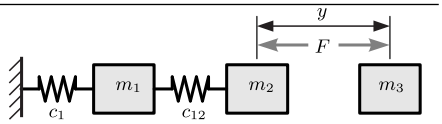
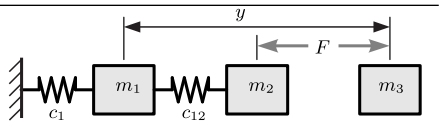
The symbol $\lceil\ \rceil$ means rounding up to full integers. In Fig. 9 both rules are violated. Table 2 states the resulting number of models for a given number of masses.

Table 2 Number of possible models for different numbers of masses

Masses N_B	1	2	3	4	5
Number of models	1	3	11	29	73

For three masses the eleven systems are given in Tab. 3. Collocated systems, i.e. sensor and actuator at the same position, are marked with an asterisk.

Table 3 Complete set of 3-mass models to be considered in the analysis. Collocated systems are marked by an asterisk *.

ID	Sketch
A1*	
A2	
A3	
A4*	
A5	
B1*	
B2	
C1*	
C2	
C3*	
C4	

5.2 Distinguishability analysis

Distinguishability of the eleven 3 mass models is investigated by the help of the necessary and sufficient conditions from section 4. The sensors are assumed to measure accelerations. For velocity (position) sensors, the same results can be obtained, except that the Markov parameters and TF coefficients are shifted by one (two) with no difference in the computational effort. Firstly, damping is neglected. The resulting pattern of distinguishable and indistinguishable model combinations is shown in Fig. 10 for the TFA and for the MPA with 8 Markov parameters. The number in each field indicates the rule from section 4 that caused the decision.

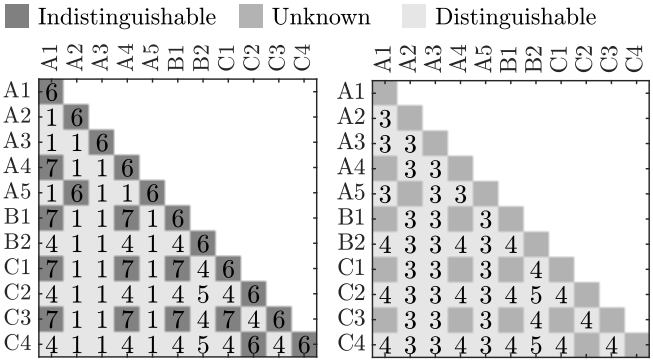


Figure 10: Distinguishability analysis without damping. Left: TFA, right: MPA with 8 Markov parameters

It can be seen that without damping a conclusive result is obtained for all model combinations with the TFA. The MPA with 8 Markov parameters produces no contradictions but it cannot prove indistinguishability of the indistinguishable models: Condition 6 is not applicable because only a subset of the $N_{S,1} + N_{S,2} = 12$ Markov parameters are calculated. Condition 7 is not applicable because there are typically more non-constant moment invariants than the 5 determinable parameters given that the overall number of Markov parameters is 12. For the case of calculating fewer Markov parameters than 8 the unknown area expands and for significantly more than 8, up to 12 the computation time would be formidable. From this comparison the TFA seems more powerful and the MPA provides no additional information.

Now, damping is included, see Fig. 11. In this case, neither the TFA, nor the MPA with 7 Markov parameters can derive an answer to the distinguishability question for all combinations. The reason is that condition 7 is not applicable in the case with damping, here. This is because the number of non-constant TF coefficients (7 . . . 9, 7 only for model A3) is almost always larger than the number of model parameters (7). Indeterminate areas result. Without damping the number of non-constant TF coefficients (3 . . . 5) is less than or equal to the number of model parameters (5). For the other conditions the situation is the same with and without damping in these areas: Conditions 1 to 5 are satisfied and 6 is not applicable.

It is important to note that in some cases the MPA proves indistinguishability that cannot be proved with the TFA. One such case is shown in Fig. 12, where system A4 has been optimized to resemble the TF of A1.

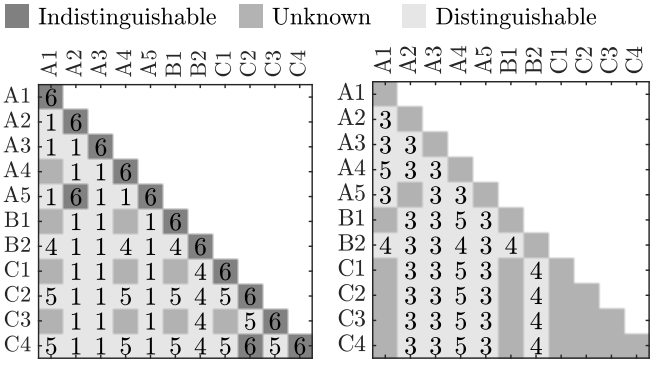


Figure 11: Distinguishability analysis incl. damping. Left: TFA, right: MPA with 7 Markov parameters

While the frequencies of resonances and antiresonances are in agreement, it is not possible to match the damping at all four (anti)resonances simultaneously. This result suggests that the evaluation of the criteria for the MPA is a valuable additional test. For 8 Markov parameters the same pattern results and for much higher numbers the computational complexity gets excessively large. 6 or fewer Markov parameters lead to larger unknown areas.

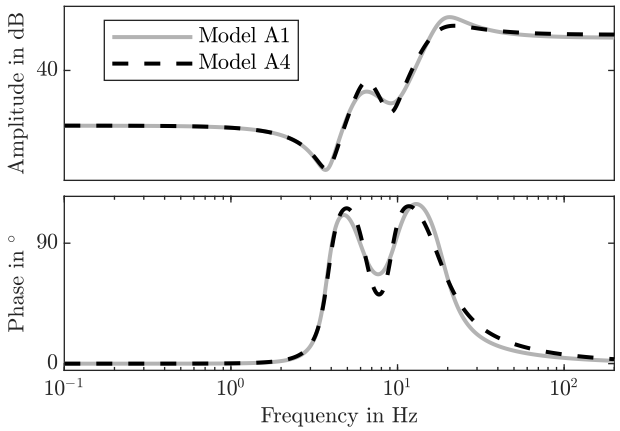


Figure 12: System A4 fitted against A1 with arbitrary parametrization, including damping

Now, the criteria based on submodels are considered, see Sec. 4.4. For the comparison of two models with the same complexity only the condition 9 is relevant. For the considered class of models it is applicable if the two models have the same actuator and sensor location, but a different position of the gap. For three masses there is no such occurrence, for four masses the criterion proves indistinguishability of one combination (not covered by any other criterion), see below and for five masses this is true for six combinations.

For the 29 4-mass models the resulting pattern considering all criteria is shown in Fig. 13. Damping is neglected because otherwise similarly to the 3-mass systems condition 6 could never be applied. In eight cases the MPA proved indistinguishability which could not be

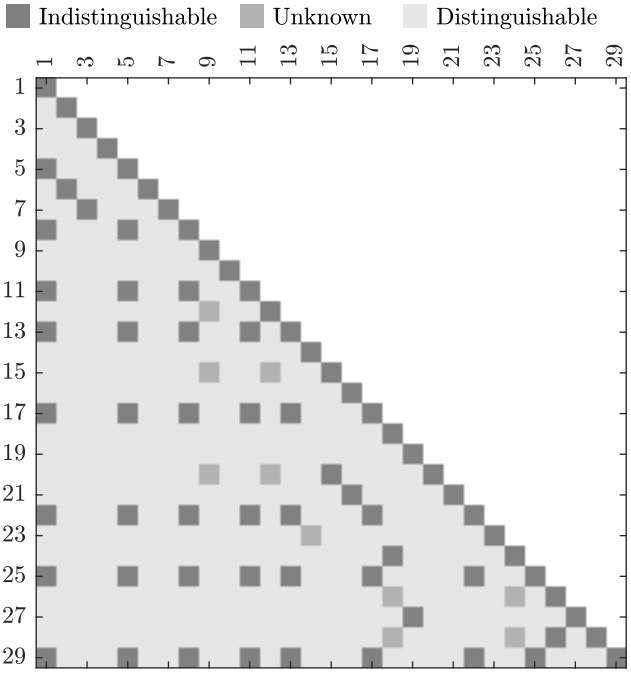


Figure 13: Combined TFA and MPA for 4-mass system, 8 Markov parameters, without damping

proved by the TFA. Equivalence of models 11 and 13 was shown by cond. 9.

Coming back to the candidate models from Fig. 1 the effect of delay time, the first-order hold model and the base dampers remains to be investigated. It is clear that delay time cannot be confused with a fractional polynomial TF, unless the order is very high. Further, the first-order hold model leads to augmented models that are distinguishable from the original models in all eleven cases. The inclusion of dampers has been investigated for the damped systems, see Fig. 14. For each of the eleven 3-mass systems it is shown which combinations of dampers are distinguishable, e.g. 0-1 means system 1 has no damper, system 2 has a base damper at mass 1. It can be seen that in the structure and parameter identification from above the set of candidate models was chosen appropriately with distinguishable models only, but that for other system, e.g. A3 it would not be possible to determine the correct positions of dampers.

6 Discussion

Our approach towards structure and parameter identification in frequency domain maximized the fit of the model while considering criteria of practical identifiability based on local sensitivity analysis. In the experiment the chosen model seems intuitively plausible, but the result depends strongly on the chosen thresholds. Especially when different interpretations of the data are admitted by visual inspection the algorithm will also suggest different models depending on the thresholds. It

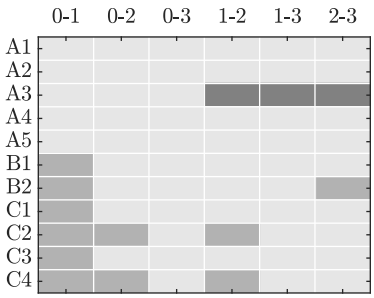


Figure 14: Distinguishability analysis of systems without additional damper (0) and damper positions 1,2 and 3

is questionable if it is possible to choose the thresholds in advance without carefully analysing all criteria for a given model, also all scaled condition indices, not only the maximum.

The selected model is not necessarily the best choice for a certain intended purpose such as feed-forward or control design, but it can be used with confidence as a practically identifiable model with good fit. Any results obtained by this method should be reviewed carefully regarding integrity.

It was chosen to include only those submodels that have exclusively identifiable parameters. Although this is sound, it sometimes leads to unintuitive conclusions, e.g. when an elastically coupled mass is not included only because the damping parameter cannot be identified securely. Here, different strategies could be further investigated for the transition from parameter investigations to submodel inclusion/exclusion decisions.

Distinguishability of the candidate models and of a broader class of multiple mass models has been analysed with the help of different necessary and sufficient criteria. The chosen composition of testing criteria is probably not complete and could be further extended to eliminate the unknown areas.

Conditions 2 and 8 never appeared to be pivotal. In the analysed setting with equally many masses of both systems the number of denominator coefficients was the same and the number of moment invariants was determined fully by the numerator. Nevertheless, these rules should not be skipped in the comparison of models with different complexity.

Future works should also consider parameter ranges. If equivalent sets of parameters include non-admissible parameter values, this can be utilized to decide in favour of one model. Furthermore, model uncertainty in the case of structurally distinguishable models should be investigated. The Akaike weights Burnham and Anderson (1998) can assess if the difference in the cost function is significant or lies in the order of measurement noise.

7 Conclusions

A procedure for model selection in frequency domain has been proposed that minimizes the Kullback-Leibler distance while also maintaining practical identifiability of all estimation parameters. It aims at mechanical models of servo systems including multiple-mass resonators. Criteria for practical identifiability are derived locally from the sensitivity matrix. Furthermore, the problem of distinguishability of model structures, which immediately arises, has been dealt with.

In experiments with two industry-like testbeds the model selection proves to reveal the characteristic mechanical properties of the two setups. Due to normalization thresholds for the criteria of practical identifiability can be chosen almost independently of the testbed properties but the results are still highly depended on the exact choices.

The set of models investigated in the distinguishability analysis is defined by structures that could describe servo positioning systems with a predefined number of elastically coupled masses. Most of the criteria are based on structural invariant vectors obtained from the transfer function approach and from the Markov parameter approach.

The transfer function approach is mostly superior over the Markov parameter approach, mainly because not all the required Markov parameters can be computed due to computational complexity. However, there are cases where the latter gives additional insights. Conditions based on the model structure helped to answer the distinguishability question for more complex models.

Acknowledgement

This work was sponsored by the German Forschungsvereinigung Antriebstechnik e.V. (FVA) and the AiF Arbeitsgemeinschaft industrieller Forschungsvereinigungen "Otto von Guericke" e.V.

References

Avdeenko, T. V. and Kargin, S. A. (2000). The problem of distinguishability of state space models. In *5th International Conference on Actual Problems of Electronic Instrument Engineering Proceedings. APEIE-2000*, pages 77–82. IEEE.

Belsley, D. A. (1991). A guide to using the collinearity diagnostics. *Computer Science in Economics and Management*, 4(1):33–50.

Bizeray, A. M., Kim, J.-H., Duncan, S. R., and Howey, D. A. (2018). Identifiability and parameter estimation of the single particle lithium-ion battery model. *IEEE Transactions on Control Systems Technology*, 27(5):1862–1877.

Brun, R., Reichert, P., and Künsch, H. R. (2001). Practical identifiability analysis of large environmental simulation models. *Water Resources Research*, 37(4):1015–1030.

Burnham, K. P. and Anderson, D. R. (1998). *Model selection and inference - a practical use of the information-theoretic approach*. Springer, New York.

Chapman, M. J. and Godfrey, K. R. (1989). A methodology for compartmental model indistinguishability. *Mathematical biosciences*, 96(2):141–164.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society: Series A*, 158(3):419–444.

Chen, Y.-Y., Huang, P.-Y., and Yen, J.-Y. (2002). Frequency-domain identification algorithms for servo systems with friction. *IEEE transactions on control systems technology*, 10(5):654–665.

Davidson, N. R., Godfrey, K. R., Alquaddoomi, F., Nola, D., and DiStefano III, J. J. (2017). Disting: A web application for fast algorithmic computation of alternative indistinguishable linear compartmental models. *Computer Methods and Programs in Biomedicine*, 143:129–135.

Diard, J. (2009). Bayesian model comparison and distinguishability. In *International Conference on Cognitive Modeling (ICCM 09)*, pages 204–209, Manchester, United Kingdom.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):45–70.

Evans, N. D., Chappell, M. J., Chapman, M. J., and Godfrey, K. R. (2004). Structural indistinguishability between uncontrolled (autonomous) nonlinear analytic systems. *Automatica*, 40(11):1947–1953.

Farrar, D. E. and Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, pages 92–107.

Gábor, A., Villaverde, A. F., and Banga, J. R. (2017). Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems. *BMC systems biology*, 11(1):54.

Galarza, R. J., Chow, J. H., and Degeneff, R. C. (1995). Transformer model reduction using time and frequency domain sensitivity techniques. *IEEE transactions on power delivery*, 10(2):1052–1059.

Godfrey, K. R. and Chapman, M. J. (1989). The problem of model indistinguishability in pharmacokinetics. *Journal of pharmacokinetics and biopharmaceutics*, 17(2):229–267.

Godfrey, K. R. and Chapman, M. J. (1990). Identifiability and indistinguishability of linear compartmental models. *Mathematics and Computers in Simulation*, 32(3):273–295.

Godfrey, K. R., Chapman, M. J., and Vajda, S. (1994). Identifiability and indistinguishability of nonlinear pharmacokinetic models. *Journal of pharmacokinetics and biopharmaceutics*, 22(3):229–251.

Hatakeyama, S., Pan, Y., and Furuta, K. (1999). Identification of continuous systems via markov parameters and time moments. *IFAC Proceedings Volumes*, 32(2):3998–4003.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 44(4):382–417.

Kovács, P., Petres, T., and Tóth, L. (2005). A new measure of multicollinearity in linear regression models. *International Statistical Review*, 73(3):405–412.

Lakhoua, M. N., Salem, J. B., Battikh, T., and Jabri, I. (2020). Review on modelling and design of mechatronic systems. *International Journal of Mechatronics and Automation*, 7(2):57–63.

Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591–612.

Meshkat, N., Rosen, Z., Sullivant, S., et al. (2018). Algebraic tools for the analysis of state space models. In *The 50th anniversary of Gröbner bases*, pages 171–205. Mathematical Society of Japan.

Motchon, K. M. D., Pekpe, K. M., and Cassar, J.-P. (2017). Conditions for strict distinguishability of single-output nonlinear control-affine systems. *Systems & Control Letters*, 105:20–26.

Motchon, K. M. D., Pekpe, K. M., Cassar, J.-P., and De Bievre, S. (2016). On the input-output distinguishability of single output continuous linear time-invariant systems. In *IEEE Transactions on Automatic Control*. Institute of Electrical and Electronics Engineers.

Nienałtowski, K., Włodarczyk, M., Lipniacki, T., and Komorowski, M. (2015). Clustering reveals limits of parameter identifiability in multi-parameter models of biochemical dynamics. *BMC systems biology*, 9(1):65.

Phan, M. Q., Lim, R. K., and Longman, R. W. (1998). Unifying input-output and state-space perspectives of predictive control. Technical Report 3044, Princeton University, Princeton, NJ.

Pintelon, R. and Schoukens, J. (2012). *System identification: a frequency domain approach*. John Wiley & Sons.

Pohjanpalo, H. (1978). System identifiability based on the power series expansion of the solution. *Mathematical biosciences*, 41(1-2):21–33.

Pronzato, L. and Pázman, A. (2013). Design of experiments in nonlinear models. *Lecture notes in statistics*, 212.

Raksanyi, A., Lecourtier, Y., Walter, E., and Venot, A. (1985). Identifiability and distinguishability testing via computer algebra. *Mathematical biosciences*, 77(1-2):245–266.

Rosa, P. and Silvestre, C. (2011). On the distinguishability of discrete linear time-invariant dynamic systems. In *50th IEEE Conference on Decision and Control and European Control Conference*, pages 3356–3361, Orlando, FL, USA. IEEE.

Schütte, F. (2003). *Automatisierte Reglerinbetriebnahme für elektrische Antriebe mit schwingungsfähiger Mechanik*. Shaker.

Schütte, F., Beineke, S., Grotstollen, H., Fröhleke, N., Witkowski, U., Rückert, U., and Rüping, S. (1997). Structure-and parameter identification for a two-mass-system with backlash and friction using a self-organizing map. In *European Conference on Power Electronics and Applications*, volume 3, pages 3–358.

Stewart, G. W. et al. (1987). Collinearity and least squares regression. *Statistical Science*, 2(1):68–84.

Stigter, J. D. and Beck, M. B. (1994). A new approach to the identification of model structure. *Environmetrics*, 5(3):315–333.

Tantau, M., Perner, L., Wielitzka, M., and Ortmaier, T. (2019). Structure and parameter identification of process models with hard non-linearities for industrial drive trains by means of degenerate genetic programming. In *16th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, volume 1, pages 368–376, Prague, Czech Republic. 29-32 July.

Tantau, M., Popp, E., Perner, L., Wielitzka, M., and Ortmaier, T. (2020). Model selection ensuring practical identifiability for models of electric drives with coupled mechanics. In *21st International Federation of Automatic Control (IFAC) World Congress*, Berlin, Germany. 12.-17. Jul.

Vajda, S. (1981). Structural equivalence of linear systems and compartmental models. *Mathematical Biosciences*, 55(1-2):39–64.

Vajda, S. (1984). Structural equivalence and exhaustive compartmental modeling. *Mathematical biosciences*, 69(1):57–75.

Vajda, S. and Rabitz, H. (1988). Identifiability and distinguishability of first-order reaction systems. *The Journal of Physical Chemistry*, 92(3):701–707.

Villwock, S. (2007). *Identifikationsmethoden für die automatisierte Inbetriebnahme und Zustandsüberwachung elektrischer Antriebe*. PhD thesis, Universität Siegen.

Volinsky, C. T., Madigan, D., Raftery, A. E., and Kronmal, R. A. (1996). Bayesian model averaging in proportional hazard models: Assessing stroke risk. Technical Report 302, Department of Statistics, University of Washington.

Vu, Q. D. (2015). *Parameter Estimation in Complex Nonlinear Dynamical Systems*. Phdthesis, Technische Universität Ilmenau.

Walter, E., Lecourtier, Y., and Happel, J. (1984). On the structural output distinguishability of parametric models, and its relations with structural identifiability. *IEEE Transactions on Automatic Control*, 29(1):56–57.

Walter, E. and Pronzato, L. (1996). On the identifiability and distinguishability of nonlinear parametric models. *Mathematics and computers in simulation*, 42(2-3):125–134.

Yates, J. W. T., Jones, R. D. O., Walker, M., and Cheung, S. Y. A. (2009). Structural identifiability and indistinguishability of compartmental models. *Expert opinion on drug metabolism & toxicology*, 5(3):295–302.

Zhang, L.-Q., Collins, J. C., and King, P. H. (1991). Indistinguishability and identifiability analysis of linear compartmental models. *Mathematical biosciences*, 103(1):77–95.