# Model Selection ensuring Practical Identifiability for Models of Electric Drives with Coupled Mechanics

**Mathias Tantau** * **Eduard Popp** * **Lars Perner** **
**Mark Wielitzka** * **Tobias Ortmaier** *

* *Leibniz University Hanover, Institute of Mechatronic Systems, An der Universität 1, 30823 Garbsen, Germany (e-mail: mathias.tantau@imes.uni-hannover.de)*
** *Lenze Automation GmbH, Am Alten Bahnhof 11, 38122 Braunschweig, Germany*

**Abstract:** Physically motivated models of electric drive trains with coupled mechanics are ubiquitous in industry for control design, simulation, feed-forward, model-based fault diagnosis etc. Often, however, the effort of model building prohibits these model-based methods. In this paper an automated model selection strategy is proposed for dynamic simulation models that not only optimizes the accuracy of the fit but also ensures practical identifiability of model parameters during structural optimization. Practical identifiability is crucial for physically motivated, interpretable models as opposed to pure prediction and inference applications. Our approach extends structural optimization considering practical identifiability to nonlinear models. In spite of the nonlinearity, local and linear criteria are evaluated, the integrity of which is investigated exemplarily. The methods are validated experimentally on a stacker crane.

*Keywords:* model selection, structure and parameter identification, practical identifiability, sensitivity analysis.

## 1. INTRODUCTION

Physically motivated models, mostly called bright grey-box models of electric drive trains with coupled mechanics are commonly used in industry for control design (Schütte (2003)), simulation, feed-forward, model-based fault diagnosis (Witczak et al. (2002)) etc. Often, however, the effort of modelling prohibits these model-based methods. As to Mittelmann et al. (2007) 75 % of the cost of a modern control application is attributed to model development.

It would be appealing to automatize the process of model building, but also a very complex task. As a first step in this direction one could derive a model that includes only those complexities which are justified by the data. The selection procedure should choose the model with the smallest prediction error but it should also maintain practical identifiability (Vu (2015)) of the bright grey-box models because otherwise interpretation and control design would be pointless.

Previous works on model selection are mainly restricted to data-driven, static models of other disciplines such as biology (Volinsky et al. (1996)) and finance (Draper (1995)). A limited model selection in the field of dynamic models for electric drives is performed in Schütte et al. (1997), but the process is not fully automatic. The final decision is left for a commissioning engineer, while the algorithm merrily extracts relevant information. Also, the analysis of practical identifiability of nonlinear dynamic models, see for example Vu (2015); Gábor et al. (2017); Popp et al. (2019), has rarely been combined with model structure optimization. It is important that each generated model must be optimized separately before it can be evaluated in the nonlinear case. In many model selection problems parameters can be included or excluded on a per-parameter basis into nested models. In this case explicitly verifying practical identifiability may be superficial because an increase in the fit with each new parameter ensures uniqueness of the parameters naturally. This is different in the case of our physically motivated models where the decisions are made on a per-submodel basis. A submodel describes a certain physical effect, e.g. friction, gravity and includes several parameters in a cluster.

One is often confronted with the following difficulties in automatic model selection:

(1) Models should be selected based on their purpose, not only on a fit measure (Hansen (2005)), but formulating purpose-optimal models, e.g. for control design, by interweaving modelling and model based control optimization would be prohibitively complex.
(2) Manual model selection involves hardly comprehensible activities like identifying characteristic responses and outliers (Nelles (2001)). Insisting on purely objective criteria for automatic selection would impose too much inflexibility (Chatfield (1995)).
(3) Data-driven model design, possibly in an iterative way, may lead to treacherously good fits due to spurious correlations, especially if the set of candidate models incorporates only little experience with the

subject matter (data dredging) (Burnham and Anderson (1998))
(4) Parameter optimization as part of the model selection requires user interaction in the nonlinear case making the integration into a fully automatic chain of tools difficult (Nelles (2001))
(5) Uniqueness: Data admit multiple interpretations of observed symptoms (Stigter and Beck (1994); Chatfield (1995))
(6) Typically long runtimes (Nelles (2001))

We respond to these concerns one by one as follows:

(1) It is advisable to decouple model identification and utilization to reduce complexities and dependencies. Model validation should be carried out prior to taking any further steps of the destined purpose (Nelles (2001)).
(2) The consequence of this procedure is that preparatory steps must forego the formal automatism. It is also expectable that fiddle parameters (Nelles (2001)) will remain to be adjusted by the operator.
(3) Modelling and identification always requires defining and selecting candidate models, only here it is made explicit. Data dredging must be avoided by specifying a small set of candidate models based on expertise (Stigter and Beck (1994)).
(4) For success a certain amount of prior knowledge is required, especially regarding the parameter ranges.
(5) It is important that the operator gets feedback about model uncertainty after the optimization, which may even necessitate a reformulation of the problem.
(6) We consider runtimes up to 10 hours (over night) on modern PC hardware acceptable. Still, time-consuming techniques are precluded.

In this paper an automated model selection strategy is proposed for dynamic simulation models that not only optimizes the accuracy of the fit in time domain but also ensures practical identifiability of model parameters during structural optimization. The resulting models contain no unnecessary complexities, which should speed up simulations and facilitate system interpretation. It is believed that the six implications from above are considered adequately in the procedure.

## 2. MODEL SELECTION AND VERIFICATION OF PRACTICAL IDENTIFIABILITY

In this section the methodology for optimizing the model structure is explained along with the set of candidate submodels and the criteria for practical identifiability.

### *2.1 Class of models*

The class of models considered is shown in Fig. 1. It is a chain of elastically coupled rotary inertias, or equivalently masses of which the first one is also subject to friction, constant gravity $M_{\mathrm{Grav}}$ and the torque of the driving motor $M_{\mathrm{M}}$, which is subject to input delay time $T_{\mathrm{dead}}$.

The friction torque is given by Schütte (2003):

$$M_{\mathrm{F}} = \underbrace{-c_{\mathrm{vis}}\dot{q}_i}_{\text{viscous}} \underbrace{-\tanh\left(f_{\tanh}\dot{q}_i\right)\left[M_{\mathrm{C}} + M_{\mathrm{S}}\mathrm{e}^{-\dot{q}_i/\Delta_{\mathrm{vel}}}\right]}_{\text{Coulomb and Stribeck}}. \quad (1)$$

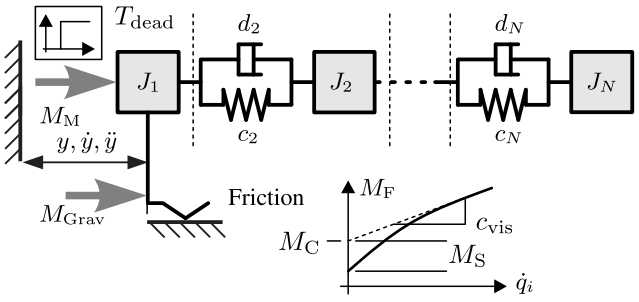


Fig. 1. Class of models

All three parts (viscous, Coulomb and Stribeck) are considered individual submodels to be included or excluded [1] with a total of four estimation parameters: $\{c_{\mathrm{vis}}, M_{\mathrm{C}}, M_{\mathrm{S}}, \Delta_{\mathrm{vel}}\}$. The other possible submodels are delay time and gravity with the estimation parameters $T_{\mathrm{dead}}$ and $M_{\mathrm{Grav}}$, resulting in $2^5$ possible combinations. In the case of a multiple mass system each additional mass is a submodel with the parameters $J_{2...N}$, $c_{2..N}$ and $d_{2...N}$. Here, the maximum number of masses is limited to $N = 3$, giving a total of 96 different overall models. Further details on the implementation of this construction kit of submodels can be found in Tantau et al. (2019). More diversity could be included easily but a large number of candidate models would contradict point 3 from above.

The resulting dynamic models can be written in the following continuous-time form:

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x}(t), u(t), \hat{\boldsymbol{p}}) + \boldsymbol{v}(t), \\ y &= h(\boldsymbol{x}(t), u(t), \hat{\boldsymbol{p}}) + w(t) \end{aligned} \quad (2)$$

with $u$: input, $\boldsymbol{x}$: states, $\hat{\boldsymbol{p}}$: estimation parameters. $\boldsymbol{v}(t)$ and $w(t)$ are mean-free noise. For simulation this description is discretized in time using Euler discretization with sampling time $T_{\mathrm{sys}}$:

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + T_{\mathrm{sys}}\boldsymbol{f}(\boldsymbol{x}_k, u_k, \hat{\boldsymbol{p}}) + \boldsymbol{v}(t_k), \\ y_k &= h(\boldsymbol{x}_k, u_k, \hat{\boldsymbol{p}}) + w(t_k). \end{aligned} \quad (3)$$

More accurate solvers could be used but since calculation time is critical here and accuracy is not the main concern in short excitations a simple solution is chosen.

### *2.2 Model selection strategy*

Model selection means finding the optimal model out of the 96 possible models. In this paper an exhaustive search over all combinations is performed. Heuristic optimizers like genetic programming (Tantau et al. (2019)) are not considered here in order to minimize random effects. For each model the parameters are optimized once via particle swarm optimization by matching time-domain simulation output $\boldsymbol{y}$ and measured system response $\boldsymbol{y}_{\mathrm{m}}$ of the training trajectory:

$$J = ||(y_{\mathrm{m},1}, y_{\mathrm{m},2}, ...y_{\mathrm{m},N_{\mathrm{S}}}) - (y_1, y_2, ..., y_{N_{\mathrm{S}}})||_2^2. \quad (4)$$

$N_{\mathrm{s}}$ is the number of samples. Excitation signals are shown in section 3. After optimization of each model's parameters, all the models that fail to meet certain criteria of practical identifiability, see next section, are discarded.

[1] The gain $f_{\tanh}$ is defined upfront. Choosing a high value will make the simulation more accurate but eventually also prone to instability.

Out of the remainder the model with the best fit on a separate validation experiment, similar to (4) is chosen.

### *2.3 Assessing practical identifiability of linear models*

In linear regression models of the form:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{p} + \boldsymbol{w} \tag{5}$$

with $\boldsymbol{y}$ being the set of observations and $\boldsymbol{w}$ an error term various criteria have been described for the matrix of regressors $\boldsymbol{X}$. They can largely be grouped into three categories: i. hypothesis testing, e.g. Farrar-Glauber tests for significant departures from orthogonality of the vectors in $\boldsymbol{X}$ (Farrar and Glauber (1967)), ii. separate treatment of sensitivity and (multi)collinearity (Stewart et al. (1987); Belsley (1991); Brun et al. (2001); Kovács et al. (2005)), and iii. judgement of parameter variance or variance inflation (Marquardt (1970); Kuczera and Mroczkowski (1998)). Hypothesis testing is criticised by Burnham and Anderson (1998) because significant departures from the null-hypothesis do not necessarily indicate the opposite. For example departures from orthogonality may still allow identification. Therefore hypothesis testing is not followed here. The idea behind category ii is that a parameter can be practically unidentifiable if it has either a too small influence on the output (small sensitivity, small corresponding column vector in $\boldsymbol{X}$) or its influence can be compensated by changes in other parameters (collinearity in $\boldsymbol{X}$). A further refinement is possible when the exact pattern of correlation among parameters is of interest (Farrar and Glauber (1967); Belsley (1991); Nienałtowski et al. (2015)).

Criteria from the last two categories are reviewed in the following as far as they are employed in this paper. For category ii Brun et al. (2001) introduced the mean squared residual (msqr) parameter importance index (PII):

$$\delta_j^{\mathrm{msqr}} = \left|\left| \boldsymbol{X}_j \Delta p_j \right|\right|_2 \tag{6}$$

with $\boldsymbol{X}_j$ the column of the regressor matrix corresponding to parameter $j$ together with the collinearity index $\gamma_k$:

$$\gamma_k = \frac{1}{\sqrt{\lambda_k}}, \tag{7}$$

in which $\lambda_k$ is the smallest singular value of $\tilde{\boldsymbol{X}}$, which is the matrix $\boldsymbol{X}$ with all columns normalized to unit length. $\Delta p_j$ is a normalization constant in the unit of parameter $p_j$. Collinearity is critical if $\gamma_k$ exceeds 5...20 (Brun et al. (2001)).

Related are also the scaled condition indices by Belsley (1991):

$$\tilde{\eta}_k = \frac{\mu_{\max}}{\mu_k}. \tag{8}$$

$\mu_k$ and $\mu_{\max}$ are the $k$-th and the maximum singular value of $\tilde{\boldsymbol{X}}$, respectively.

A criterion for category iii is the scaled parameter covariance matrix (Nelles (2001)):

$$\mathrm{cov}\left(\hat{\boldsymbol{p}}\right) = \sigma^2 \left( \boldsymbol{X}_{\mathrm{n}}^{\mathrm{T}} \boldsymbol{X}_{\mathrm{n}} \right)^{-1}, \tag{9}$$

where $\sigma^2$ is the variance of the output noise, $\boldsymbol{X}_{\mathrm{n}}$ is the matrix $\boldsymbol{X}$ with all columns multiplied by the range of the respective parameter. Considering the parameter range ensures that the parameter covariance is derived relative to this range and not in physical units.

The exclusion criterion for practical identifiability is defined as follows: Only if the smallest $\delta_j^{\mathrm{msqr}}$ is at least [2] 1 % of the largest $\delta_j^{\mathrm{msqr}}$ and if all $\tilde{\eta}_k$ are less than 30, as suggested by Belsley (1991), the model is accepted. In (6) the parameter ranges for normalization $\Delta p_j$ correspond to the bounds of the identification. The other criteria are not directly evaluated in the model selection process but will be reported for comparison.

### *2.4 Extension to nonlinear models*

Usually the profile likelihood method, see for example Raue et al. (2009) is the best choice for the identifiability analysis of nonlinear models. Instead the linear criteria from above are used as a local/asymptotic approximation in the vicinity of the previously obtained parameter estimate $\boldsymbol{p}_0$ as explained in this section. The profile likelihood method is used only to make the results plausible.

For the nonlinear regression model (Brun et al. (2001); Kitsos and Kolovos (2013)):

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{p}) + \boldsymbol{w} \tag{10}$$

with the objective function $J = (\boldsymbol{y}_{\mathrm{m}} - \boldsymbol{y})^{\mathrm{T}}(\boldsymbol{y}_{\mathrm{m}} - \boldsymbol{y})$ the standard assessment of identifiability for differentiable models is made via linearisation (Brun et al. (2001)):

$$\boldsymbol{g}(\boldsymbol{p}) \approx \boldsymbol{g}(\boldsymbol{p}_0) + \left.\frac{\partial \boldsymbol{g}(\boldsymbol{p})}{\partial \boldsymbol{p}^{\mathrm{T}}}\right|_{\boldsymbol{p}=\boldsymbol{p}_0} (\boldsymbol{p} - \boldsymbol{p}_0). \tag{11}$$

Thus, a substitute for $\boldsymbol{X}$ is given by the output sensitivity

$$\boldsymbol{X} = \left.\frac{\partial \boldsymbol{g}(\boldsymbol{p})}{\partial \boldsymbol{p}^{\mathrm{T}}}\right|_{\boldsymbol{p}=\boldsymbol{p}_0}. \tag{12}$$

A relation can be established between the local criteria and the curvatures of the profile likelihood plot for each parameter $p_i$. Corresponding to the linearisation (11) the quadratic approximation of the cost function $J(\boldsymbol{p})$ around the optimum is given by $J_0 + (\boldsymbol{p} - \boldsymbol{p}_0)^{\mathrm{T}} \cdot \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \cdot (\boldsymbol{p} - \boldsymbol{p}_0)$. According to the derivation in Press et al. (2007) the objective function plotted against one parameter while the others are optimized can be approximated locally by $J_i(p_i) \approx J_0 + (p_i - p_{i,0})^2 / C_{ii}$ with $\boldsymbol{C} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}$. So the observed, local curvature $2/C_{ii}$ of the profile likelihood plot is directly related to the non-normalized variance, compare (9) [3].

A further distinction between sensitivity and collinearity is not possible, but when the remaining parameters, other than $p_i$, are not optimized in each step, the resulting curvature around $\boldsymbol{p}_0$ is two times the element $ii$ of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$, so with this modification sensitivity can also be visualized and the difference between these modified plots and the original profile likelihood plots is attributed to collinearity. The collinearity index (7) characterizes this flattening out

[2] A rejection threshold of 1/10000 as recommended in Gábor et al. (2017) led to unreasonable results.

[3] Normalization is included implicitly into the profile likelihood method by plotting the functions against the parameter ranges.

caused by optimizing the other parameters, but it does not provide a quantitative measure for each parameter.

### *2.5 Sensitivity calculation*

In compliance with the optimization target each point in time is a separate entry in $\boldsymbol{y}$. The output sensitivity can be calculated readily from the output equation,

$$\boldsymbol{X}_y = \frac{\mathrm{d}y}{\mathrm{d}\hat{\boldsymbol{p}}^\mathrm{T}} = \frac{\partial h}{\partial \boldsymbol{x}^\mathrm{T}} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\hat{\boldsymbol{p}}^\mathrm{T}} + \frac{\partial h}{\partial \hat{\boldsymbol{p}}^\mathrm{T}}, \tag{13}$$

but it requires the state sensitivity $\boldsymbol{X}_x = \mathrm{d}\boldsymbol{x}/\mathrm{d}\hat{\boldsymbol{p}}^\mathrm{T}$, which cannot be calculated directly. But given continuous second partial derivatives of $\boldsymbol{x}(\hat{\boldsymbol{p}}, t)$ with respect to $\boldsymbol{x}$ and $t$ the partial derivatives are commutative according to Schwarz's theorem and the sensitivity differential equation can be formulated (Bohn (2000)):

$$\underbrace{\frac{\mathrm{d}\dot{\boldsymbol{x}}}{\mathrm{d}\hat{\boldsymbol{p}}^\mathrm{T}}}_{\dot{\boldsymbol{X}}_x} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}^\mathrm{T}} \underbrace{\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\hat{\boldsymbol{p}}^\mathrm{T}}}_{\boldsymbol{X}_x} + \frac{\partial \boldsymbol{f}}{\partial \hat{\boldsymbol{p}}^\mathrm{T}}. \tag{14}$$

Using the Euler discretisation again the sensitivity can be propagated over time from given initial values with sensitivity sample rate $T_\mathrm{sens}$:

$$\underbrace{\frac{\mathrm{d}\boldsymbol{x}_{k+1}}{\mathrm{d}\hat{\boldsymbol{p}}^\mathrm{T}}}_{\boldsymbol{X}_{x,k+1}} = \underbrace{\frac{\mathrm{d}\boldsymbol{x}_k}{\mathrm{d}\hat{\boldsymbol{p}}^\mathrm{T}}}_{\boldsymbol{X}_{x,k}} + T_\mathrm{sens} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}^\mathrm{T}} \underbrace{\frac{\mathrm{d}\boldsymbol{x}_k}{\mathrm{d}\hat{\boldsymbol{p}}^\mathrm{T}}}_{\boldsymbol{X}_{x,k}} + \frac{\partial \boldsymbol{f}}{\partial \hat{\boldsymbol{p}}^\mathrm{T}}. \tag{15}$$

In the partial derivatives $\boldsymbol{x}_k$ is inserted for $\boldsymbol{x}$ and $u_k$ for $u$, while $\hat{\boldsymbol{p}}$ is constant. This procedure was adopted from Bohn (2000) and it is also described in Gábor et al. (2017). Here, the initial values of the sensitivities are set to zero while the excitation trajectory also starts from standstill, independent of the parameters. So, the initial output sensitivity should actually be zero. For the parameter $T_\mathrm{dead}$ the difference quotient with linear sub-sample interpolation is used instead of the above procedure, since this parameter does not appear in the system or output function directly.

## 3. EXPERIMENTAL RESULTS

In this section we firstly validate the linear, local identifiability criteria against the profile likelihood method in a representative experiment. Then, the structure optimization is carried out. For all experiments the testbed shown in Fig. 2 is used. It is a stacker crane with three orthogonal, belt-driven axes. The mast is 5.6 m heigh and the horizontal axis along the shelf ($x$-axis) is 5 m long. Only experiments on the $x$-axis are reported, while the load handling device is held in a height of 2 m. From stepped sine frequency-domain measurements it is known that this axis shows the behaviour of a three-mass resonator, see Fig. 3.

For the simulation of all dynamic models the sampling rate is $T_\mathrm{sys} = 10\,\mathrm{kHz}$, which is approx. factor 100 above the expected max. eigenfrequencies of resonating masses ($\sim 100\,\mathrm{Hz}$), while the sensitivity simulation is even sampled at $T_\mathrm{sens} = 100\,\mathrm{kHz}$ to ensure stability of the latter for

Fig. 2. Experimental testbed: stacker crane

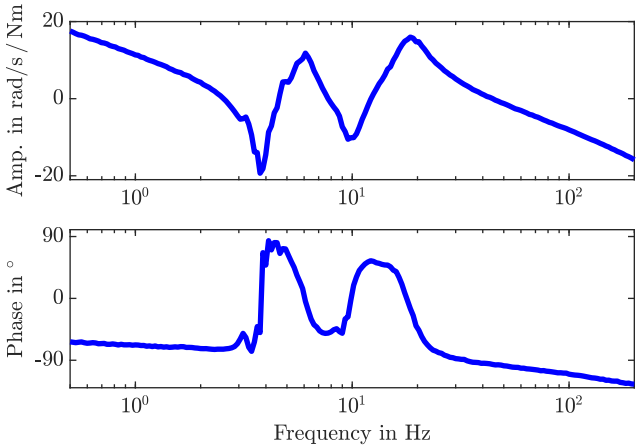


Fig. 3. Frequency response functions of the testbed measured with stepped sine excitation

all obtained parameter sets. The entire process of model selection and validation takes approximately three hours in our implementation on an i7 4-core computer running at 3.7 GHz with 16 GB DDR4 Ram, which complies with point 6 from above.

The choice of parameter ranges, see Fig. 5, is an important manual preparation step corresponding to point 4 from above. The inertias are known to be between 0 and the overall inertia. For the other parameters it is more difficult to specify bounds and the results should be interpreted on the background of these parameter ranges.

### *3.1 Comparison of profile likelihood and local identifiability criteria*

The training trajectory for optimizing the full model is shown in Fig. 4, a crest factor-optimized multisine signal (MFE) (Mittelmann et al. (2007)), along with the resulting model output. Full model means including all potential submodels. Fig. 5 shows the corresponding profile likelihood plot. It seems that all parameters, except

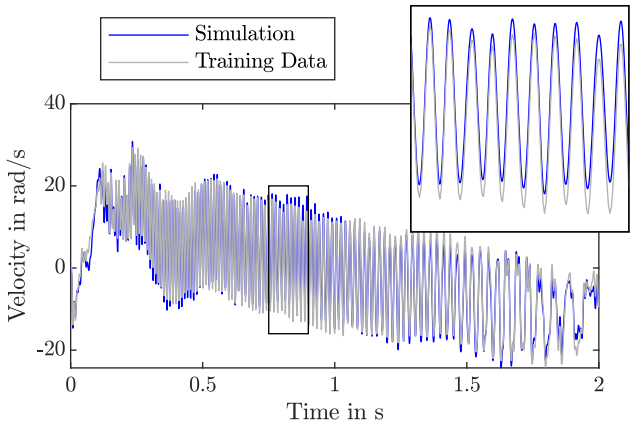


Fig. 4. Training data and corresponding model output after training with MFE excitation, full model

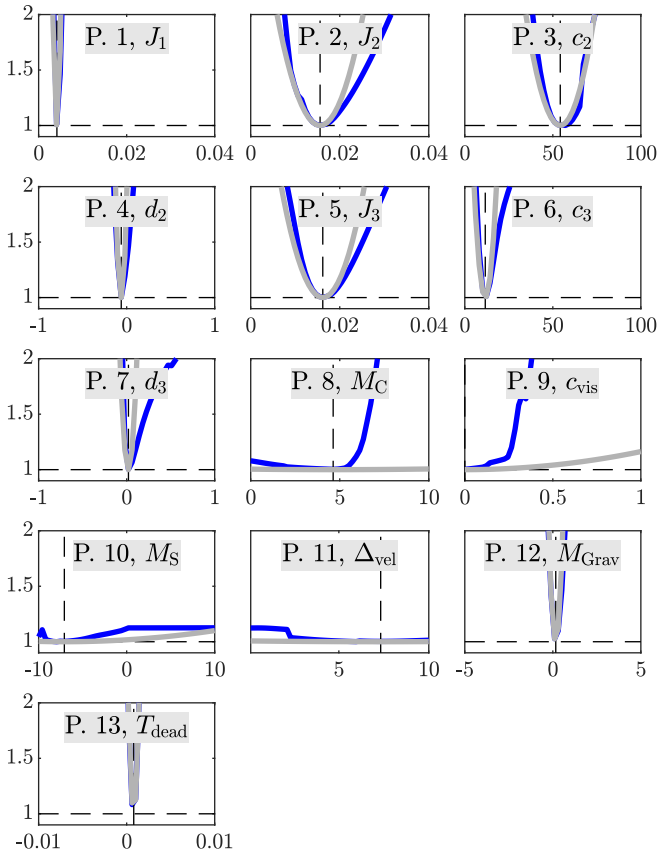


Fig. 5. Profile likelihood plot corresponding to Fig. 4. The abscissa spans over each parameter's range, the same as used for the identification. Thick blue line: profile likelihood cost function $J$ in multiples of the minimum, thick grey: asymptotic approximation, thin dashed: identification result.

the friction parameters $\{c_{vis}, M_C, M_S, \Delta_{vel}\}$ are practically identifiable. Apparently, for the given dynamic excitation friction is overparameterized.

Next, the local criteria are evaluated to find out if they support or confute the results of the profile likelihood. Firstly, it is striking that the asymptotic approximates, also shown in Fig. 5 show the same behaviour although the curves are not completely identical throughout. The PIIs

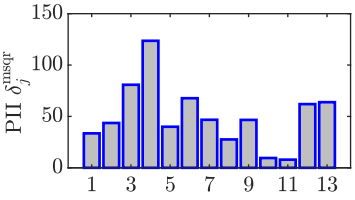


Fig. 6. Parameter importance indices for Fig. 4

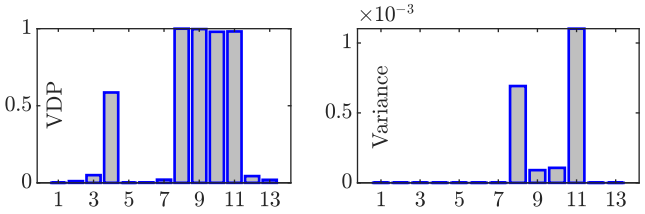


Fig. 7. Left: variance decomposition proportions for Fig. 4, right: diagonal of the scaled parameter covariance matrix according to (9)

$\delta_j^{msqr}$ of the 13 parameters are shown in Fig. 6. As they are all in the same order of magnitude, the identifiability problem with the friction parameters is not explained by the sensitivities.

The largest condition index $\tilde{\eta}_k$ is $281.3 > 30$, the collinearity index $\gamma_k$ is $111 > 20$, so the unidentifiability problem can be explained by collinearity. Taking this a step further the scaled variance decomposition proportions for the largest condition index according to Belsley (1991), are shown in Fig. 7 on the left. Each of them corresponds to one parameter. Although not always quite so clear, here they show large values for the questionable four parameters 8 to 11, indicating that these four are responsible for collinearity. The less pronounced but also visible involvement of parameter 4 could be the reason why its parabola in Fig. 5 is less steep than that of parameter 1, although it has the highest sensitivity in Fig. 6.

As a last evaluation the scaled parameter variances are plotted in Fig. 7 on the right. For scaling the same parameter ranges have been used as before. The absolute scale of the variances is not meaningful because the variance of the output is arbitrarily set to one. It can be seen that the variances also indicate a problem in the variables 8 to 11.

As a conclusion of this examplary investigation, the local criteria are in satisfactory agreement with the likelihood plots and can be used for structural optimization in the next section.

### *3.2 Structural optimization*

Structure and parameter optimization is performed in experiments. The same MFE excitation as before is used for the training experiment, for validation an MFE excitation with different settings is applied to the testbed. In Fig. 8 and 9 the measured velocities as well as the prediction of the best model are shown. As a result of the structure identification a three-mass system with viscous friction and delay time is obtained. Comparing the figures visually, the fit of this model seems less good on validation data than on training data, but in both cases it is acceptable.

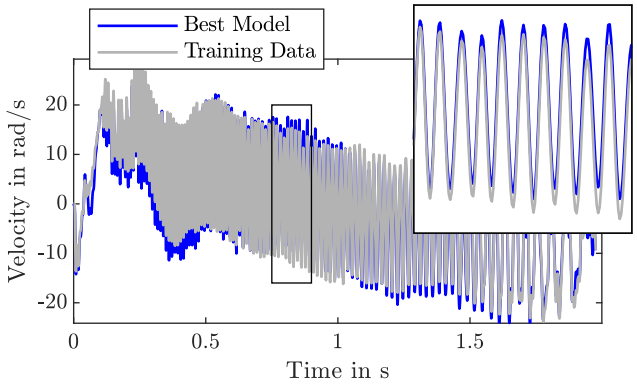


Fig. 8. Training data and best model

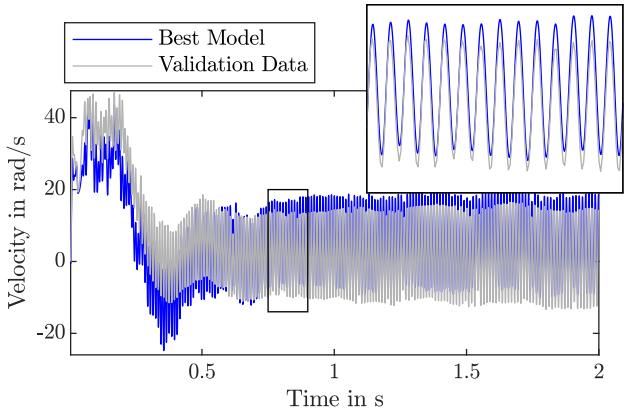


Fig. 9. Validation data and best model

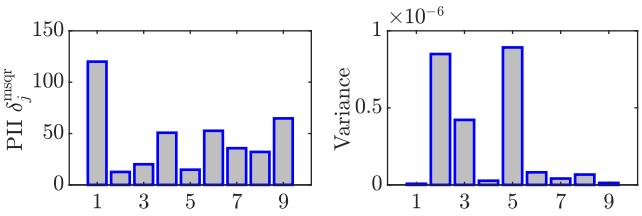


Fig. 10. PIIs and variances of the best model

The scaled parameter sensitivities and variances are shown in Fig. 10 and the scaled condition indices are $\tilde{\eta}_k$ : 1.0, 1.2, 1, 2, 1.6, 1.8, 2.1, 3.6, 4.7. As requested, all sensitivities are in the same order of magnitude and the largest $\tilde{\eta}_k$ is less than 30. The likelihood plot in Fig. 11 also indicates identifiability of this model, so again profile likelihood and the local, linear criteria are in agreement. It is striking that viscous friction is included in the optimal model although it was judged unidentifiable in combination with the other submodels in the full model.

When repeating the structural optimization it sometimes happens that Coulomb friction or viscous friction with parameters close to zero are also included in the resulting model. Such models have approximately the same fit and their is no direct penalty on the number of estimation parameters, only if clear overfitting occurs, the fit on validation data will deteriorate. The number of masses is always found to be three which is in agreement with the frequency response measurement in Fig. 3. It would be interesting to see what the algorithm returns in less obvious cases.

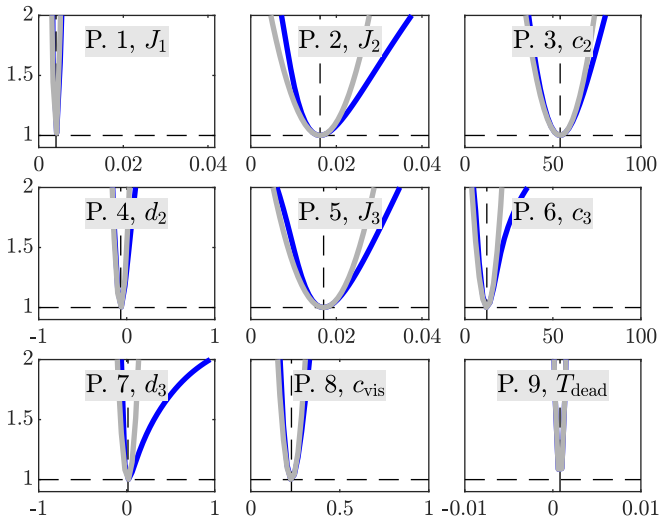


Fig. 11. Profile likelihood plot for excitation with multisine signal for the best model, compare Fig. 5

## 4. DISCUSSION

A procedure for structure and parameter identification of nonlinear dynamic models has been demonstrated. It complies with all six implications from above, except for the model uncertainty evaluation (point 5). Because the outputs of the simulation at different time steps are correlated, methods for model uncertainty evaluation like the Akaike weights from Burnham and Anderson (1998) cannot be used directly. Especially in the observed cases of different friction models in subsequent runs it would be nice to know if the best model could be chosen with certainty or if another model is just as good.

In addition to the six points from above a new problem has occurred: The commissioner is not limited to a given set of data but can design the excitation arbitrarily. This is a difficulty but also a chance, which is not fully utilized in this paper. Also, special, dedicated trajectories to identify each parameter separately as in Schütte (2003) could not readily be integrated into our holistic assessment of practical identifiability. In a similar spirit the adjustment of hyperparameters for the criteria of practical identifiability intensifies point 2 from above. Fortunately the methods applied here do not need many hyperparameters, only one for sensitivity and one for collinearity. They are widely independent of the model because of normalization.

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.

In any case, results obtained with the method should be reviewed carefully regarding integrity. Reasons for unreasonable models are often related to the formulation of the problem, e.g. candidate models, parameter bounds, excitation, hyperparameters. Model simplifications, although unavoidable, may lead to unrealistic values of the remaining parameters.

## 5. CONCLUSIONS

A procedure for model selection of nonlinear dynamic models of electric drive trains has been proposed. It selects the model with the lowest prediction error that satisfies certain criteria of practical identifiability and identifies its parameters. Because of the long simulation times only linear criteria for practical identifiability are evaluated, but it is shown in a few examples that these are in satisfactory agreement with the profile likelihood method, which is the more thorough nonlinear counterpart.

In experiments with an industry-like testbed the methodology proves to reveal the characteristic mechanical properties of the setup.

The method is easy to use as it requires only two time series measurements of for example 2 s each and the optimization of approx. three hours can be performed without user interaction. Expert knowledge is still required for setting initial parameter ranges and for checking sanity of the results.

## ACKNOWLEDGEMENTS

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

## REFERENCES

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

Bohn (2000). *Recursive parameter estimation for nonlinear continuous time systems through sensitivity model based adaptive filters.* phdthesis, Ruhr-Universität Bochum, Bochum, Germany.

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.

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

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

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

Hansen, B.E. (2005). Challenges for econometric model selection. *Econometric Theory*, 21(1), 60–68.

Kitsos, C.P. and Kolovos, K.G. (2013). A compilation of the d-optimal designs in chemical kinetics. *Chemical Engineering Communications*, 200(2), 185–204.

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

Kuczera, G. and Mroczkowski, M. (1998). Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. *Water Resources Research*, 34(6), 1481–1489.

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

Mittelmann, H.D., Pendse, G., Rivera, D.E., and Lee, H. (2007). Optimization-based design of plant-friendly multisine signals using geometric discrepancy criteria. *Comput. Optim. and Appl.*, 38(1), 173–190.

Nelles, O. (2001). *Nonlinear system identification: from classical approaches to neural networks and fuzzy models.* Springer-Verlag, Berlin.

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.

Popp, E., Tantau, M., Wielitzka, M., Ortmaier, T., and Giebert, D. (2019). Frequency domain identification and identifiability analysis of a nonlinear vehicle drivetrain model. In *ECC*, 237–242. IEEE.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (2007). *Numerical recipes 3rd edition: The art of scientific computing.* Cambridge university press.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15), 1923–1929.

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, 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., 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 *ICINCO*, volume 2, 368–376. Praque, Czech Republic.

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.

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

Witczak, M., Obuchowicz, A., and Korbicz, J. (2002). Genetic programming based approaches to identification and fault diagnosis of non-linear dynamic systems. *International Journal of Control*, 75(13), 1012–1031.