International Operational Modal Analysis Conference
11th International Operational Modal Analysis Conference
20 - 23 May 2025 | Rennes, France

Cross-Validation in Stochastic Subspace Identification

Simon Marwitz [1], Tom Lahmer [2], Volkmar Zabel [3]

[1] Bauhaus-Universität Weimar, simon.jakob.marwitz@uni-weimar.de
[2] Bauhaus-Universität Weimar, tom.lahmer@uni-weimar.de
[3] Universität Rostock, volkmar.zabel@uni-rostock.de

ABSTRACT

Operational Modal Analysis (OMA) often relies on user-specified parameters, such as model order and input sequence lengths. These are typically selected based on expert knowledge. While stabilization diagrams effectively determine the model order, they are unsuitable for selecting other sensitive parameters. The existing criteria for the assessment of parameter sets in OMA fall short of meeting key requirements, including the independence from system dynamics, and applicability across different OMA methods.
This study proposes a common validation criterion based on input reconstruction and cross-validation techniques. The covariance- and data-driven stochastic subspace identification methods, are extended for the purposes of multi-block estimation and synthesis. Moreover, a modally decomposed formulation is selected to facilitate an additional assessment of the identified modal parameters.
The proposed method is verified through the use of synthetic and experimental vibration signals, as well as a full-factorial parameter study, which collectively demonstrate the robustness and consistency of the proposed criterion.
The key objective of the developed methodology is to enhance confidence in identified modal parameters, thereby facilitating more reliable assessments in e. g. structural health monitoring or life-cycle analyses.

Keywords: Cross Validation, Synthetization, Reconstruction, Uncertainty, Stochastic Subspace Identification, Quality

1. INTRODUCTION

Operational Modal Analysis (OMA) applications frequently depend on user-specified parameter sets, such as the order of identified models and the length of input sequences, among other factors. These parameters are often selected based on expert knowledge and experience.Although the model order can be determined using a stabilization diagram, this method is not sufficient for selecting other sensitive parameters.

Criteria for the assessment of a chosen set of parameters for OMA exist under the name of modal indicators, mode validation criteria, etc. However, they are typically

- specific to a single OMA method, e. g. the Consistent Mode Indicator [1] for the ERA, or
- specific to particular structures, e. g. the Modal Phase Colinearity [1] is only applicable to proportionally damped structures with properly spaced modes,
- not normalized, e. g. the Mean Phase Deviation [2] may take any value in $[0, 180]°$, or
- suffer from overfitting and / or sample bias, e. g. the Modal Contributions [3] use all data for training and validation.

The proposed criterion is derived from the identified systems in various OMA methods, assuming a direct correspondence to the modal parameters. The reconstruction error [4, 5] is a common quality measure based on reconstructed input signals, synthesized by the identified system. To mitigate overfitting and/or sample bias if the same input signals are used for identification and validation, cross-validation methods [6], a class of well-known techniques in statistics, are employed.

In this study, synthetization methods for two stochastic state-space models are extended to accommodate multi-block estimation and cross-validation. The method is verified on synthetic signals and demonstrated with experimental vibration measurements from a guyed tower. A full-factorial parameter study demonstrates the unbiased behavior and numerical robustness of the developed criterion.

2. CROSS-VALIDATION (CV) TECHNIQUES

The reconstruction error of any system identification method can be estimated by comparing the identification input sequences with those synthesized by the identified system. However, it is recognized that training and evaluating a model on the same data can lead to over-optimistic validation estimates. CV methods [6] try to overcome these limitations. These methods divide the available data into a training set, which is used for system identification, and a validation set, against which the reconstructed system output is validated. The input and output of the system are not necessarily required to be timestep signals; they may also be correlation functions, spectral densities, or other derived quantities, depending on the identification method used.

The underlying idea of CV is to divide the available input data into n_{bl} blocks and subsequently assemble them in a set $D = \{\boldsymbol{d}_1, \dots, \boldsymbol{d}_{n_{\mathrm{bl}}}\}$. A portion of these blocks is allocated for training a model (the training sample $D_{\mathrm{t}} \subset D$), while the remainder is used for evaluating its performance (the validation sample $D_{\mathrm{v}} = D_{\mathrm{t}}^{\complement}$). A measure of fit $\gamma(\mathcal{M}_{D_{\mathrm{t}}}; \boldsymbol{d}_i)$ quantifies the discrepancy between the output of an identified model $\mathcal{M}_{D_{\mathrm{t}}}$ and a validation sample $\boldsymbol{d}_i \in D_{\mathrm{v}}$. One such measure is developed in Sect. 4. The loss function is defined as the expected value of the validation measure

$$\mathcal{L}(\mathcal{M}_{D_{\mathrm{t}}}, D_{\mathrm{v}}) = \mathbb{E}[\gamma(\mathcal{M}_{D_{\mathrm{t}}}, D_{\mathrm{v}})] . \tag{1}$$

Different estimators exist depending on the employed method for assembling the data blocks into training and validation sets.

The simple *Hold-Out* (HO) estimator consists of a single $(b = 1)$ training and validation set. The validation results are then averaged over the validation samples

$$\widehat{\mathcal{L}}_{\mathrm{HO}}(\mathcal{M}_{D_{\mathrm{t}}}, D_{\mathrm{v}}) = \frac{1}{n_{\mathrm{v}}} \sum_{d_i \in D_{\mathrm{v}}} \gamma(\mathcal{M}_{D_{\mathrm{t}}}, \boldsymbol{d}_i) . \tag{2}$$

However, the validation results are dependent on the selection of the training and validation sets, which introduces sample bias. To circumvent this, CV methods assemble the training and validation sets in b distinct ways and average the validation measures. The Leave-One-Out and Leave-p-Out validators are examples of this category. The former employs all but one set for training and a single validation set, repeating this process across each set as the validation set, i. e. $b = n_{\mathrm{bl}}$. It should be noted that both of these methods are exhaustive, and as such, they may become computationally expensive if the training and/or validation procedure is so. A balanced choice that is not exhaustive but still reduces sample bias is the k-fold CV method. It involves the grouping of the data sets into k approximately equally sized validation subsets which are then successively used for training and validation. The CV estimator is subsequently averaged over each of these k groups.

$$\widehat{\mathcal{L}}_k(\mathcal{M}; D) = \frac{1}{k} \sum_{j=1}^{k} \widehat{\mathcal{L}}_{\text{HO}}(\mathcal{M}_{D_{\text{t},j}}; D_{\text{v},j}) \,. \tag{3}$$

3. MULTI-BLOCK STOCHASTIC SUBSPACE IDENTIFICATION

The *Stochastic Subspace Identification* (SSI) technique is a well-known method in operational modal analysis for identification of discrete, stochastic state-space models from measured signals. For a detailed description of the algorithms used in this study the reader is referred to [7–9]. The discrete stochastic state space system is defined by

$$\begin{aligned} \boldsymbol{x}[n+1] &= \boldsymbol{A}\boldsymbol{x}[n] + \boldsymbol{w}[n]\,, && \text{the state equation, and} \\ \boldsymbol{y}[n] &= \boldsymbol{C}\boldsymbol{x}[n] + \boldsymbol{v}[n]\,, && \text{the output equation,} \end{aligned} \tag{4}$$

where $\boldsymbol{x}(t)$ is the state vector, $\boldsymbol{A}$ is the state matrix, $\boldsymbol{C}$ is the output matrix, and the stochastic white noise vectors $\boldsymbol{w}[n]$ and $\boldsymbol{v}[n]$ account for the unknown input terms. The eigendecomposition $\boldsymbol{A\Psi} = \boldsymbol{\Psi\Lambda}$, is the basis for the estimation of modal frequencies f_j, modal damping ratios ζ_j, and mode shapes $\boldsymbol{\phi}_j$, as well as the modal decoupling of the state space system.

The required extensions to apply multi-block cross-validation techniques with the SSI are given in the following. In the covariance-driven case, a block-modified form of the unbiased correlation function estimator between the channels $i_\text{l} = 1 \dots n_\text{l}$ and reference channels $i_\text{r} = 1 \dots n_\text{r}$ of the input signal $\boldsymbol{y}[n]$ is

$$\hat{R}_{i_\text{l}i_\text{r},i_\text{bl}}[m] = \frac{1}{N_\text{bl} - m} \sum_{n=0}^{N_\text{bl}-m-1} y_{i_\text{l}}[i_\text{bl}N_\text{bl} + n] y_{i_\text{r}}[i_\text{bl}N_\text{bl} + n + m] \,, \tag{5}$$

where $i_\text{bl} = 1 \dots n_\text{bl}$ is the block number, N_bl is the block length in timesteps, and $m = 1 \dots M$ is the lag step of the correlation function. For short block lengths N_bl and high lags m, large confidence intervals may occur [10]. These can be circumvented by allowing the second input signal $y_{i_\text{r}}[n+m]$ to overlap into the consecutive block by m timesteps, at the cost of reduced block independence. The identification of $\boldsymbol{A}$, $\boldsymbol{C}$ and $\boldsymbol{G}_\text{ref} = \mathbb{E}\left[\boldsymbol{x}[n+1] \quad \boldsymbol{y}_\text{ref}[n]^\top\right]$ then proceeds as shown in [7] using an averaged correlation function estimate over a (training) subset of these n_bl correlation function estimates.

In the data-driven case, the projection matrix [11] must be estimated from multiple blocks of signals. Döhler [9] has developed a method to compute independent and identically distributed estimates of reduced projection matrices. This method is extended here for full projection matrices, which are required for input reconstruction.

A block Hankel matrix of the so-called past and future signals is split into n_bl blocks along its block-columns as shown in [9] such that

$$\begin{bmatrix} \boldsymbol{Y}_{-,1} & \boldsymbol{Y}_{-,2} & \cdots & \boldsymbol{Y}_{-,n_\text{bl}} \\ \boldsymbol{Y}_{+,1} & \boldsymbol{Y}_{+,2} & \cdots & \boldsymbol{Y}_{+,n_\text{bl}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_1 & \boldsymbol{L}_2 & \cdots & \boldsymbol{L}_{n_\text{bl}} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{Q}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{Q}_{n_\text{bl}} \end{bmatrix} \,, \tag{6}$$

where $\boldsymbol{L}_j$, $\boldsymbol{Q}_j$ are the individual LQ decompositions of each block.

Another thin LQ decomposition of the column-stacked triangular matrices is defined

$$\begin{bmatrix} \boldsymbol{L}_1 & \boldsymbol{L}_2 & \cdots & \boldsymbol{L}_{n_\text{b}} \end{bmatrix} = \breve{\boldsymbol{L}}\breve{\boldsymbol{Q}} \tag{7}$$

where uniqueness is achieved by restraining the diagonal of $\breve{\boldsymbol{L}}$ to positive values, i. e. by multiplication of the rows / columns of $\breve{\boldsymbol{L}}, \breve{\boldsymbol{Q}}$ with $\mathrm{sgn}(\mathrm{diag}(\breve{\boldsymbol{L}}))$. Döhler [9] has shown, that

$$\boldsymbol{L} = \breve{\boldsymbol{L}} \,, \qquad \boldsymbol{Q} = \breve{\boldsymbol{Q}} \begin{bmatrix} \boldsymbol{Q}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{Q}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{Q}_{n_{\mathrm{bl}}} \end{bmatrix} \tag{8}$$

can be assumed. The identification then proceeds as detailed in [7] using only a (training) subset of these n_{bl} blocks in Eq (6). The system $\boldsymbol{A}, \boldsymbol{C}$ is finally identified by solving the following relation [7]

$$\begin{bmatrix} \hat{\boldsymbol{X}}_{q+1} \\ N^{-1/2} \left[\boldsymbol{y}[q+1] \quad \boldsymbol{y}[q+2] \quad \cdots \quad \boldsymbol{y}[q+N]\right] \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \hat{\boldsymbol{X}}_q + \begin{bmatrix} \boldsymbol{\rho}_w \\ \boldsymbol{\rho}_v \end{bmatrix} , \tag{9}$$

in a least-squares sense. The residuals yield the noise correlations $\boldsymbol{\Sigma}_{\boldsymbol{ww}}, \boldsymbol{\Sigma}_{\boldsymbol{wv}}, \boldsymbol{\Sigma}_{\boldsymbol{vv}}$, which are required for the construction of the Kálmán gain in the next section.

4. (MODAL) RECONSTRUCTION OF INPUT SEQUENCES

The states and corresponding system matrices are unique up to within a similarity transform [11]. Thus, a modally decoupled identified stochastic forward state-space model can be constructed as follows

$$\begin{aligned} \boldsymbol{x}_{\mathrm{m}}[n+1] &= \boldsymbol{\Lambda} \boldsymbol{x}_{\mathrm{m}}[k] + \boldsymbol{w}_{\mathrm{m}}[n] \\ \boldsymbol{y}[n] &= \boldsymbol{\Phi} \boldsymbol{x}_{\mathrm{m}}[n] + \boldsymbol{v}_{\mathrm{m}}[n] \,. \end{aligned} \tag{10}$$

Similarly, a forward innovation model, which is obtained by applying a Kálmán filter to the forward model, can be transformed. By eliminating the innovation sequences, the modally decoupled forward innovation model is expressed as

$$\begin{aligned} \hat{\boldsymbol{x}}_{\mathrm{m}}[n+1] &= (\boldsymbol{\Lambda} - \boldsymbol{K}_{\mathrm{m}} \boldsymbol{C}_{\mathrm{m}})\, \hat{\boldsymbol{x}}_{\mathrm{m}}[n] + \boldsymbol{K}_{\mathrm{m}} \boldsymbol{y}[n] \\ \boldsymbol{y}[n] &= \tilde{\boldsymbol{y}}_{\mathrm{m}}[n] + \tilde{\boldsymbol{y}}_{\boldsymbol{e}}[n] \,. \end{aligned} \tag{11}$$

In both cases the modal states and system matrices are

$$\boldsymbol{x}_{\mathrm{m}}[n] = \boldsymbol{\Psi}^{-1} \boldsymbol{x}[n] \,, \quad \boldsymbol{A} = \boldsymbol{\Psi} \boldsymbol{\Lambda} \boldsymbol{\Psi}^{-1} \,, \quad \boldsymbol{C}_{\mathrm{m}} = \boldsymbol{C} \boldsymbol{\Psi} \,, \text{and} \quad \boldsymbol{K}_{\mathrm{m}} = \boldsymbol{\Psi}^{-1} \boldsymbol{K} \,. \tag{12}$$

Modal Decomposition of Correlation Functions Given a modally decoupled identified stochastic forward state-space model, Reynders [4] has demonstrated the synthetization of correlation functions

$$\tilde{\boldsymbol{R}}_{\mathrm{m}}[m] = \sum_{j=1}^{n_{\mathrm{m}}} \left(\lambda_j^{m-1} \boldsymbol{\phi}_j \boldsymbol{g}_{\mathrm{m},j}^{\top} + \lambda_{j^*}^{m-1} \boldsymbol{\phi}_{j^*} \boldsymbol{g}_{\mathrm{m},j^*}^{\top} \right) \,, \tag{13}$$

where n_{m} the total number of modes, $j = 1 \ldots n_{\mathrm{m}}$ is the mode number and, by slight abuse of notation, j^* the index of its complex conjugate counterpart. The modal participation vector $\boldsymbol{g}_{\mathrm{m},j}$ is the j^{th} row of the modally decomposed next-state output covariance matrix, that is obtained by solving $\boldsymbol{G}_{\mathrm{ref}} = \boldsymbol{\Psi} \boldsymbol{G}_{\mathrm{m}}$ [12]. An example will be introduced in Sec. 5. and is shown in Fig. 1.

Modal Decomposition of Kálmán-Filter State Sequences The modal decoupling of the Kálmán filter states of the input signal allows the estimation of each mode's contribution to the signal as shown by [3]. Under stationary conditions the Kálmán gain can be formulated in terms of the noise covariance matrices [3, 13] and the forward state covariance matrix follows from the solution of the Ricatti equation.

As the output matrix $\boldsymbol{C}_{\mathrm{m}}$ is also modally decoupled, the reconstructed signal in Eq. (11) can be obtained from the Kálmán filter states as

$$\tilde{\boldsymbol{y}}_{\mathrm{m}}[n] = \sum_{j=1}^{n_{\mathrm{m}}} \left(\hat{x}_{\mathrm{m},j}[n] \boldsymbol{C}_{\mathrm{m},j} + \hat{x}_{\mathrm{m},j^*}[n] \boldsymbol{C}_{\mathrm{m},j^*} \right) \tag{14}$$

An example will be introduced in Sec. 5. and is shown in Fig. 2.

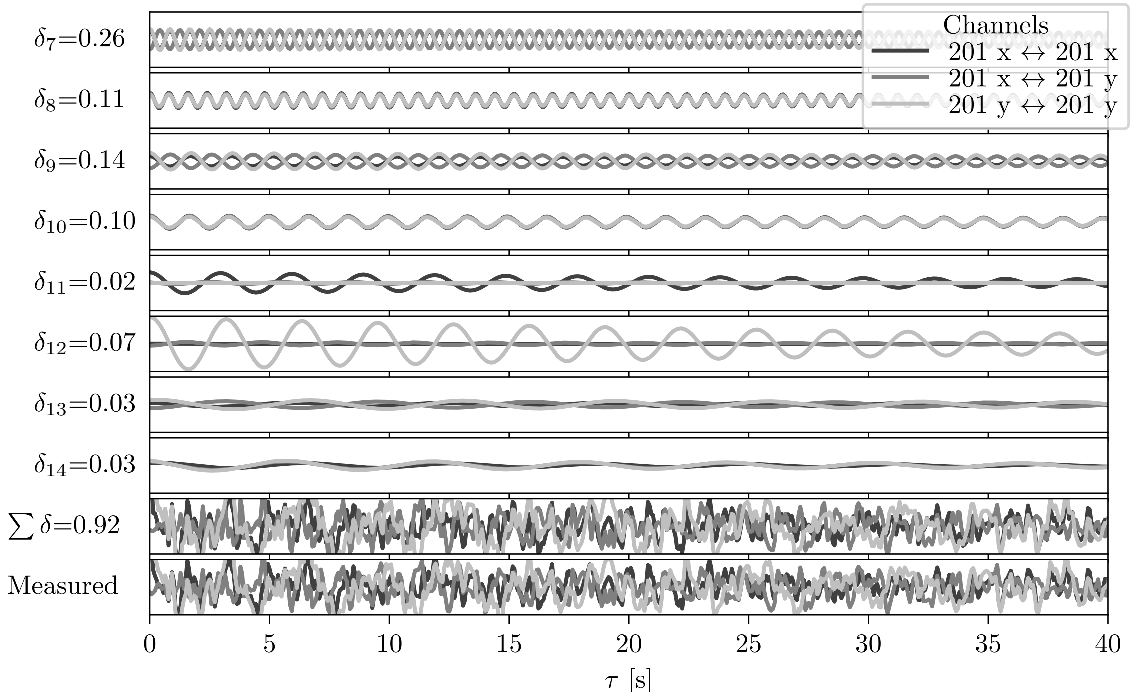


Figure 1: Verification Example: Synthetic correlation functions, modally decomposed by an identified, order-28 state-space model. Only the auto- and cross-correlations between the two reference channels (”201 x”, ”201 y”) are shown. Time lags up to $\tau = 40$ s are displayed and δ_j indicates the relative contribution of mode j to the input (measured) correlation function.

Validation Measure and Modal Contribution A validation measure is developed based on the reconstructed input sequences

$$\boldsymbol{d}[l] = \sum_{j=1}^{n_\mathrm{m}} \tilde{\boldsymbol{d}}_j[l] + \boldsymbol{d}^\epsilon[l] \quad , \tag{15}$$

where $\tilde{\boldsymbol{d}}_j[l]$ corresponds to either Eq. (13), (14) or an equivalent formulation for other OMA methods. Here, $l = 1 \dots L$ and $L = \{M, N\}$ is the sequence index (correlation / time), and $\boldsymbol{d}^\epsilon[l]$ is an additional error term resulting from an insufficient fit of the identified system. $\boldsymbol{d}[l]$ consist of $p = \{n_\mathrm{l}, n_\mathrm{l} n_\mathrm{r}\}$ rows that correspond to the signal channels or channel-pairs of the correlation function.

A validation measure of the identified system is then obtained from the normalized correlation coefficient of the input sequences with their synthesized counterpart, averaged over the rows and modes

$$\delta = \sum_{j=1}^{n_\mathrm{m}} \delta_j \quad \text{, where} \quad \delta_j = \frac{1}{p} \sum_p \frac{\sigma^2_{d\tilde{d}_j,p}}{\sqrt{\sigma^2_{\tilde{d}\tilde{d},p}\sigma^2_{dd,p}}} \ . \tag{16}$$

Here, δ_j is the contribution of each identified mode to the input data , which is also known as the modal participation factor [5] or the modal contribution [3]. The (co)-variances for each row p of the input and synthesized sequences, respectively, are defined as

$$\sigma^2_{\tilde{d}\tilde{d},p} = \sum_{l=1}^{L} \tilde{d}_p[l]\tilde{d}_p[l], \qquad \sigma^2_{dd,p} = \sum_{l=1}^{L} d_p[l] d_p[l], \qquad \sigma^2_{d\tilde{d}_j,p} = \sum_{l=1}^{L} d_p[l]\tilde{d}_{j,p}[l] \tag{17}$$

assuming zero mean sequences.

In order to adapt this measure to the cross-validation approach, the input sequences and their synthesized counterpart are split into n_bl blocks along the running index l. Subsequently, δ is computed on the validation subset only, according to the chosen validation estimator, as defined in Sec. 2..

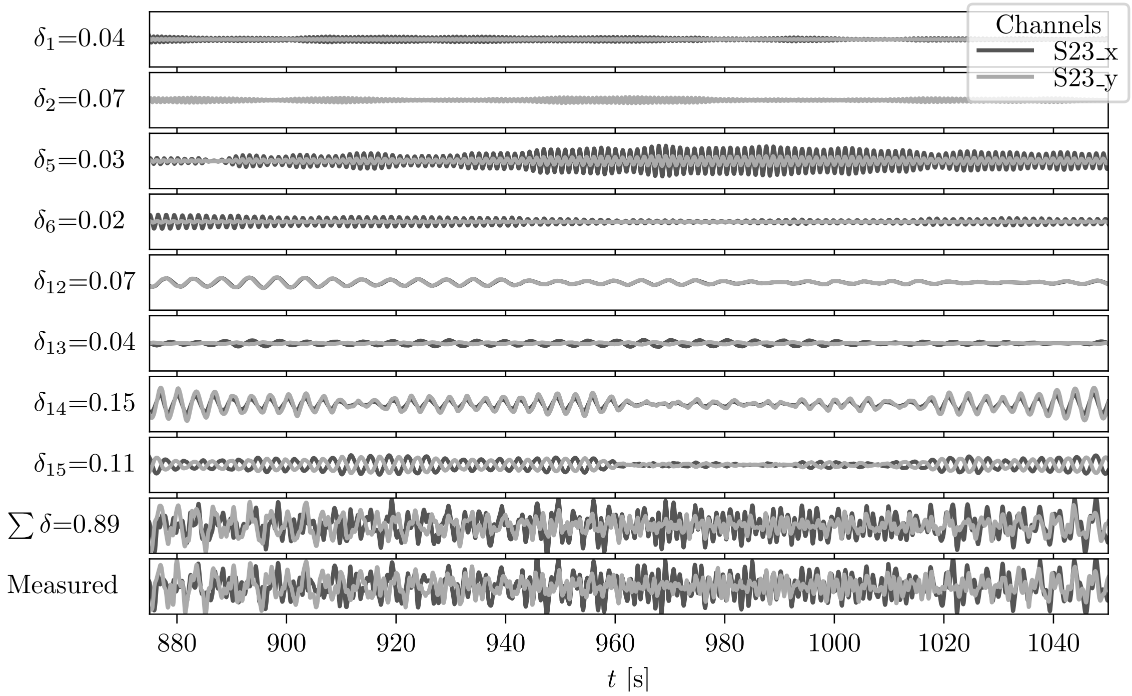


Figure 2: Application Example: Experimental vibration response signals, modally decomposed by an identified, order-30 state-space model. Only the signals at the two reference channels ("S23 x", "S23 y") are shown. A representative time window of $t = 175$ s is displayed and δ_j indicates the relative contribution of mode j to the measured signal.

5. NUMERICAL VERIFICATION AND EXPERIMENTAL APPLICATION

A guyed steel transmission mast with a height of 200 m has been selected to demonstrate the developed method, as shown in Figure 3. A *verification example* using synthetic acceleration response signals and an *application example* using experimental acceleration response measurements are shown.

The analysis of the acceleration vibration response measurements carried out in 2019 was presented in [14]. Figure 4 shows an extract of the identified modal characteristics.

A simplified numerical model is developed in this study, modeling the mast as a continuous beam with elastic supports, representing the pre-stressed guy cables, and a tuned mass damper, which introduces non-proportional damping; thus the modal parameters must be interpreted as non-classical [15]. Further details with respect to the modeling are described in [14, 16].

Synthetic acceleration response time histories are generated by an inverse Fourier transform of the frequency response function, utilizing generalized modal coefficients [5], multiplied with spectral stochastic wind excitation [17]. The signals are virtually sampled at 5 locations and two directions and varying sample rates and durations. Additional white Gaussian measurement noise is added s. t. SNR $= -10$ dB.

Modal Synthetization of Input Sequences The following example demonstrates the modal reconstruction of correlation functions and response signals from identified state-space models. The size of the inputs is chosen[1] as $M = 2048$ and $n_{\text{ord}} = 28$ in the covariance-driven case and $p = q = 1024$ and $n_{\text{ord}} = 30$ in the data driven case, where p and q are the number of block-rows and -columns of the Block-Hankel matrix and n_{ord} is the order of the identified stochastic state space model. At a duration $T = 3600$ s and sample rate $f_{\text{s}} = 14$ Hz the signal has a total of $N = 50400$ timesteps.

Fig. 1 shows the modal reconstruction of correlation functions for selected physical modes of the veri-

[1]For the sake of a smooth display of synthesized signals, a much longer correlation function than required for identification purposes was chosen.

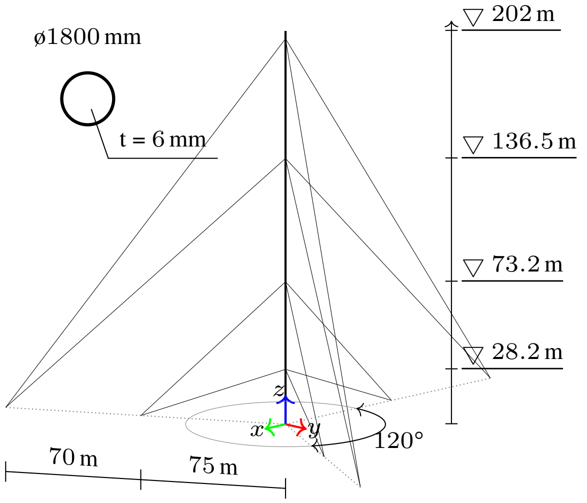


Figure 3: Geometry and dimensions of the antenna mast (Photograph: Wolfgang Greiner)

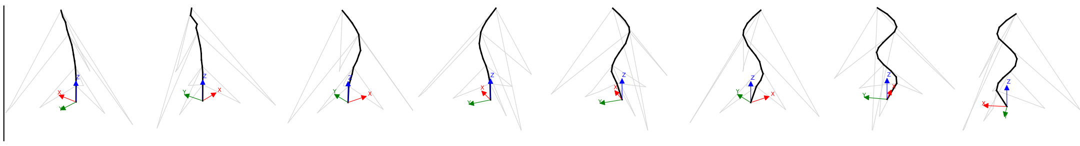
ϕ_j								
$f_{d,j}$ [Hz]	0.196	0.192	0.471	0.501	1.024	1.043	1.326	1.336
ζ_j [%]	0.466	0.479	0.615	0.365	0.47	0.25	0.29	0.23

Figure 4: Part of the identified modal parameters of the mast (Note the rotated coordinate systems)

fication (numerical) example. Similarly, Fig. 2 shows the modally reconstructed response signal of the application (measurement) example.

The examples demonstrate the effectiveness of the employed reconstruction methods in two scenarios. In the application example, the high overall reconstruction rate indicates low-noise signals. While noise significantly affects the response signal reconstruction, its impact is effectively reduced due to the averaging nature of correlation functions in their reconstruction. The examples also show that the modal contributions, can reliably be used as modal indicators, but are dependent on the excitation and experimental setup, highlighting the need for careful consideration of these factors.

Full-Factorial Study of the Cross-Validation Measure To demonstrate the developed cross-validation procedure, the total modal reconstruction indicator $\delta = \sum \delta_j$ from the previous example is used as the validation measure γ. The input signals are split into $n_{\mathrm{bl}} = 40$ blocks and assembled into $k = 10$ randomized subsets of cardinality $n_{\mathrm{t}} = 36$ for identification and $n_{\mathrm{v}} = 4$ for validation. The loss estimate, which is used in the following two full-factorial parameter studies, follows as

$$\widehat{\mathcal{L}}_k = \frac{1}{k} \sum_k \delta_k \,. \tag{18}$$

In the case of the experimental application example, the model order $n_{\mathrm{ord}} = 4 \dots 40$ and the number of block-rows $q = 20 \dots 200$ are each sampled at 37 steps. The resulting variations of the loss estimate in Figure 5 indicate, that the validation measure increases with increasing model order and increasing time lags. However, a closer investigation reveals the occurrence of modal splitting at higher model orders. This is not represented in the reconstruction criterion and points to a possible direction of further research.

In the second study, the numerical verification example, the signal duration $T = 600 \dots 7200$ s and sample rate $f_{\mathrm{s}} = 3.9 \dots 70$ Hz are varied similarly. A stochastic state space model of order $n_{\mathrm{ord}} = 28$ and the

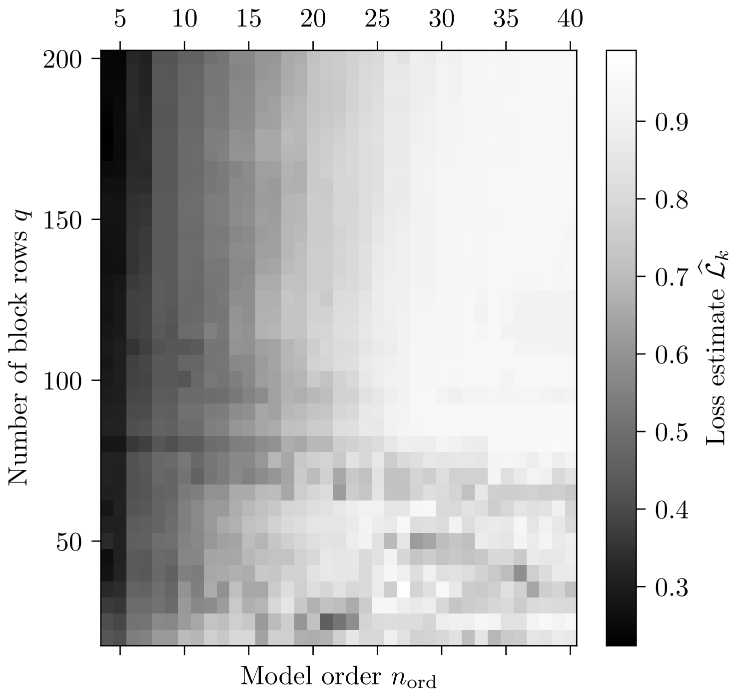


Figure 5: Application Example: Loss estimate $\widehat{\mathcal{L}_k}$ for cross-validation of identified stochastic forward innovation state space models (SSI-data) at different model orders n_{ord} and numbers of block-rows/-columns p, q.

corresponding validation measure is identified from correlation functions with length $\tau_{\mathrm{max}} = 8\,\mathrm{s}$. The loss estimate is displayed in Figure 6, which indicates, that the sample rate should be chosen about a factor four higher than the highest natural frequency and a longer signal duration is favorable.

The examples show, that the proposed method is consistent over a wide range of parameters and indicates its usefulness for purposes of optimization or uncertainty quantification.

6. CONCLUSIONS AND OUTLOOK

In summary, the present study proposes an improved validation criterion for selecting parameter sets in Operational Modal Analysis. It combines input reconstruction with cross-validation techniques to provide a more objective and reliable means for determining the influence of sensitive parameters in the measurement and analysis process.

Notably, the criterion is application- and method-agnostic, scalar and normalized, and robust with respect to overfitting and sample bias. It is designed to be independent of system and excitation dynamics. Extensions of two stochastic subspace identification methods have been developed to include multi-block estimation and reconstruction in a modally decomposed manner, for further use as a stabilization criterion of single modes. The method has been implemented within an object-oriented, open-source Python framework [18].

The effectiveness of this approach has been demonstrated through both numerical and experimental vibration signals. Its application to an experimental example, a 200-meter-tall guyed steel mast, demonstrates its applicability and highlights the critical need for consideration of noise and excitation conditions in the modal reconstruction process. A full-factorial study has further demonstrated the criterion's consistent performance over a broad range of parameters, indicating its usefulness for optimization and uncertainty quantification in modal analysis.

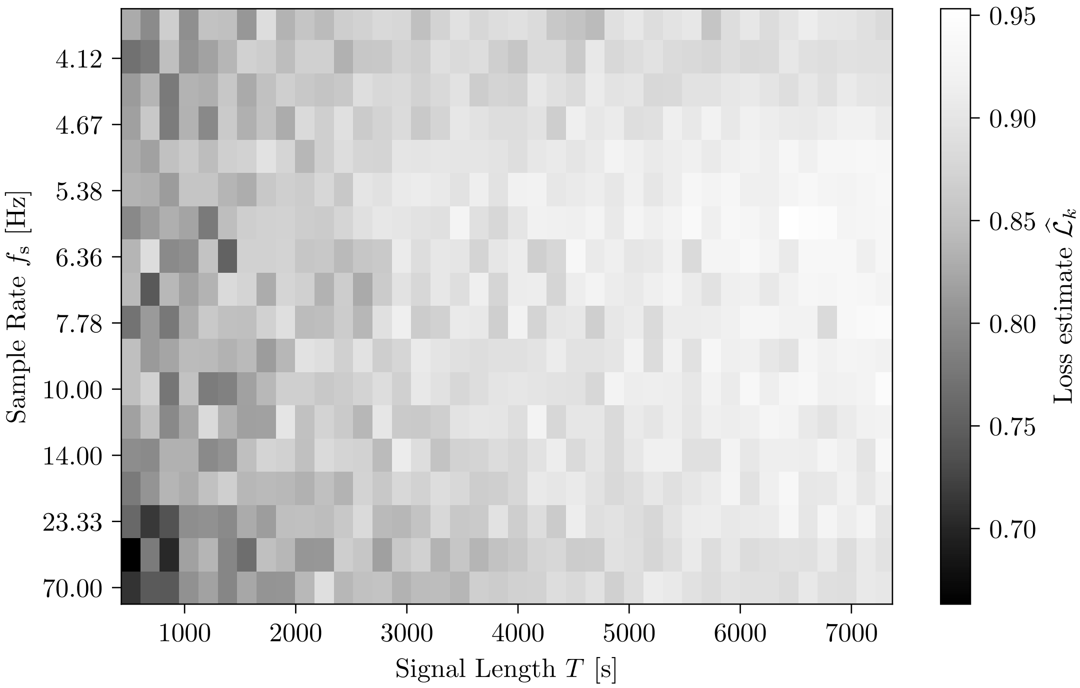


Figure 6: Verification Example: Loss estimate $\widehat{\mathcal{L}_k}$ for cross-validation of stochastic forward state space models (SSI-cov) identified from signals with different durations T and sample rates f_s.

It is acknowledged that the method exhibits certain limitations, which offer directions for future research. Specifically, the utilization of other Cross-Validation methods, such as Leave-One-Out (LOO) or Leave-K-Out (LKO), has not been explored, despite being theoretically applicable. The application examples also highlight challenges in the context of modal splitting, indicating potential areas for future research. While the method has been successfully demonstrated for the SSI methods and additionally implemented for the pLSCF method, its extension to other OMA methods, though conceptually possible, remains to be developed and implemented.

ACKNOWLEDGEMENTS

The financial support of the German Research Foundation (DFG) for the project *Assessment and Reduction of Uncertainties in Operational Modal Analysis* (RUN-OMA) is gratefully acknowledged.

DATA AVAILABILITY STATEMENT

The implementations developed as part of this study are available under a GPL-3.0 license at `https://github.com/pyOMA-dev/pyOMA` [18]. The data that support the findings of this study are available upon request from the corresponding author, SM.

REFERENCES

[1] R. S. Pappa, K. B. Elliott, and A. Schenk. "A consistent-mode indicator for the eigensystem realization algorithm". In: *AIAA Dynamics Specialists Conference* (Apr. 16, 1992).

[2] E. Reynders, J. Houbrechts, and G. De Roeck. "Fully automated (operational) modal analysis". In: *Mechanical Systems and Signal Processing* 29 (2012), pp. 228–250.

[3] F. J. Cara et al. "Modal contribution and state space order selection in operational modal analysis". In: *Mechanical Systems and Signal Processing* 38.2 (2013), pp. 276–298.

[4] E. Reynders. "System identification methods for (operational) modal analysis: review and comparison". In: *Archives of Computational Methods in Engineering* (2012), pp. 51–124.

[5] R. Brincker and C. Ventura. *Introduction to Operational Modal Analysis*. Wiley, 2015.

[6] S. Arlot and A. Celisse. "A survey of cross-validation procedures for model selection". In: *Statistics Surveys* 4.0 (2010), pp. 40–79.

[7] B. Peeters and G. de Roeck. "Reference-Based Stochastic-Subspace Identification for Output-Only Modal Analysis". In: *Mechanical Systems and Signal Processing* (1999).

[8] M. Döhler and L. Mevel. "Fast multi-order computation of system matrices in subspace-based system identification". In: *Control Engineering Practice* 20.9 (2012), pp. 882–894.

[9] M. Döhler. "Subspace-based system identification and fault detection: Algorithms for large systems and application to structural vibration analysis". Theses. Université Rennes, 2011.

[10] P. Stoica. *Spectral analysis of signals*. Upper Saddle River, N.J: Prentice Hall, 2005.

[11] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer Academic Publishers, 1996.

[12] B. Peeters and G. de Roeck. "System identification and damage detection in civil engineering". eng. PhD thesis. KU Leuven, 2000.

[13] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer International Publishing, 2017.

[14] V. Zabel, S. Marwitz, and A. Habtemariam. "Bestimmung von modalen Parametern seilabgespannter Rohrmasten". In: *Berichte der Fachtagung Baustatik-Baupraxis*. Vol. 14. Baustatik Baupraxis e. V. Mar. 2020, pp. 519–526.

[15] E. Kausel. *Advanced Structural Dynamics*. Cambridge University Press, Aug. 2017.

[16] S. Marwitz, T.Lahmer, and V. Zabel. "Quantification of Polymorphic Uncertainties - A Quasi-Monte Carlo Approach". In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* (June 1, 2024).

[17] M. Clobes. *Identifikation und Simulation instationärer Übertragung der Windturbulenz im Zeitbereich*. Aachen, 2008.

[18] Simon Marwitz, Volkmar Zabel, and Andrei Udrea. *pyOMA*. 2025.