**Proceedings Paper:**

O'Connell, B.J. orcid.org/0000-0001-6042-927X and Rogers, T.J. orcid.org/0000-0002-3433-3247 (2024) On improving the efficiency of Bayesian stochastic subspace identification. In: Rainieri, C., Gentile, C. and Aenlle López, M., (eds.) Proceedings of the 10th International Operational Modal Analysis Conference (IOMAC 2024): Volume 1. 10th International Operational Modal Analysis Conference (IOMAC 2024), 22-24 May 2024, Naples, Italy. Lecture Notes in Civil Engineering, LNCE 514 . Springer Nature Switzerland , pp. 609-617. ISBN 9783031614200

https://doi.org/10.1007/978-3-031-61421-7_59

# On improving the efficiency of Bayesian stochastic subspace identification

Brandon J. O'Connell[1 [0000−0001−6042−927X]] and
Timothy J. Rogers[1 [0000−0002−3433−3247]]

Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield City Centre, Sheffield, S1 3JD
BJOConnell1@sheffield.ac.uk

**Abstract.** The recent development of a Bayesian stochastic subspace identification (SSI) algorithm for OMA has provided a new systematic and principled way of recovering posterior distributions over desired modal characteristics in an operational setting. Despite their many advantages, there is often a reluctance to adopt Bayesian methodologies in engineering practice because of their higher computational requirements. In the case of Bayesian SSI, this problem is even more relevant given the inherent speed of the traditional SSI algorithm. This has highlighted the need for a computationally efficient implementation of the Bayesian SSI algorithm, required to make Bayesian SSI a more competitive choice when considering multiple OMA approaches. This paper presents a novel solution, based on stochastic variational inference, and develops upon existing methods to speed up the Bayesian SSI algorithm. This method is evaluated using a simulated case study and subsequently compared to that of classical SSI and the current Bayesian SSI implementation.

**Keywords:** Bayesian · Stochastic Variational Inference · Stochastic Subspace

## 1 Introduction

As demand for models and data become more strict, there is an increasing desire in the engineering community to obtain a measure of uncertainty in order to better assess the quality of any chosen methodology and ultimately evaluate the risk of various outcomes. There has been growing interest in the OMA community to explore uncertainty based methods in an effort to improve operational modal identification and its derivatives. By their nature, OMA algorithms operate under inherently stochastic input regimes, lending themselves to higher signal to noise ratios and therefore, typically more difficult to identify datasets. Consequently, engineers are developing methodologies that account for this stochasticity and quantify some form of uncertainty on the learnt modal properties. Stochastic subspace identification (SSI) has seen particular attention, given its high performance in many OMA settings and suitability to this problem [13, 14, 4, 3].

Recently the authors proposed a probabilistic interpretation of canonical variate-weighted covariance-driven SSI (SSI-Cov) [12], realised using the theory of probabilistic projections [1]. This alternative probabilistic representation conveniently lends itself to familiar hierarchical extensions, e.g. robust formulations [12].

This was later followed by a Bayesian SSI algorithm for uncertainty quantification [10, 11], where the inclusion of prior distributions over the defined model parameters results in the recovery of a posterior on the observability matrix and, by extension, posteriors on the modal parameters. This recovery was demonstrated through two different inference schemes; Markov-chain Monte-Carlo (MCMC), and variational inference (VI).

Despite their many advantages, there is often a reluctance to adopt Bayesian methodologies in engineering practice because of their higher computational requirements. In the case of Bayesian SSI, this problem is even more relevant given the inherent speed of the singular value decomposition in the classic SSI algorithm. Computational efficiency is important when considering the application of Bayesian SSI as a suitable alternative to traditional techniques and as a competitive choice for uncertainty quantification. This paper introduces a solution to this problem, employing a batch stochastic variational inference (SVI) scheme, known to improve the overall efficiency of VI [5]. It will then be shown how batch SVI reduces the total number of operations required to reach convergence to the posterior.

## 2 Covariance-Driven Stochastic Subspace Identification

Given standard derivations of canonical variate-weighted SSI-Cov [7, 15], it is widely known that for a classically-defined, output-only state-space model, the extended observability ($\mathcal{O}$) and controllability ($\mathcal{C}$) matrices can be computed using the singular value decomposition (SVD),

$$\boldsymbol{\Sigma}_{ff}^{-1/2}\boldsymbol{\Sigma}_{fp}\boldsymbol{\Sigma}_{pp}^{-1/2} = \mathbf{V}_1\boldsymbol{\Lambda}\mathbf{V}_2 \backsimeq \breve{\mathbf{V}}_1\breve{\boldsymbol{\Lambda}}\breve{\mathbf{V}}_2^{\mathsf{T}} \tag{1}$$

where $\boldsymbol{\Sigma}_{ff}$ and $\boldsymbol{\Sigma}_{pp}$ are block auto-covariances and $\boldsymbol{\Sigma}_{fp}$ is the block cross-covariance between Hankel matrices of the future $\mathbf{Y}_f$ and past $\mathbf{Y}_p$, constructed using lags of the measured time series. Vectors $\mathbf{V}_1$ and $\mathbf{V}_2$ correspond to the left and right singular vectors of the SVD respectively, and $\breve{\boldsymbol{\Lambda}}$ neglects small singular values in $\boldsymbol{\Lambda}$ such that the resulting vector has dimension $d = \dim(\breve{\boldsymbol{\Lambda}})$. This is equivalent to the well known statistical concept, canonical correlation analysis (CCA) [6]. The cross-covariance matrix, $\boldsymbol{\Sigma}_{fp}$, can be then decomposed into a product of the extended observability and controllability matrices $\boldsymbol{\Sigma}_{fp} = \mathcal{O}\mathcal{C}$ such that,

$$\mathcal{O} \;=\; \boldsymbol{\Sigma}_{ff}^{1/2}\breve{\mathbf{V}}_1\breve{\boldsymbol{\Lambda}}^{1/2} \quad , \quad \mathcal{C} \;=\; \breve{\boldsymbol{\Lambda}}^{1/2}\breve{\mathbf{V}}_2^{\mathsf{T}}\boldsymbol{\Sigma}_{pp}^{\mathsf{T}/2} \tag{2}$$

where $\mathrm{rank}(\mathcal{O}) = \mathrm{rank}(\mathcal{C}) = d$. From this standard result, the state matrix and by extension the modal properties can be recovered in the usual way for SSI-Cov [7].

## 3 Probabilistic Stochastic Subspace Identification

In previous work by the authors [12], it was shown that SSI-Cov could be rewritten probabilistically, substituting CCA with its probabilistic equivalent, defined by Bach and Jordan using a latent variable model [1]. This latent model assumes two datasets $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ originate from a shared latent space $\mathbf{z}$, in which they are maximally correlated, and are obtained through a linear transformation (or weight) $\mathbf{W}$, mean offset $\boldsymbol{\mu}$, and noise precision $\boldsymbol{\Psi}$. The overall model is given by Equations (3 - 5), where m = 1,2.

$$\mathbf{z}_n \sim \mathcal{N}(0, \mathbb{I}) \tag{3}$$

$$\mathbf{x}_n^{(m)}|\mathbf{z}_n \sim \mathcal{N}(\mathbf{W}^{(m)}\mathbf{z}_n + \boldsymbol{\mu}^{(m)}, \boldsymbol{\Sigma}^{(m)}) \tag{4}$$

$$\mathbf{x}_n|\mathbf{z}_n \sim \mathcal{N}(\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{5}$$

In this probabilistic form, assuming $\mathbf{x}^{(1)} = \mathbf{Y}_f \in \mathbb{R}^{D_1 \times N}$ and $\mathbf{x}^{(2)} = \mathbf{Y}_p \in \mathbb{R}^{D_2 \times N}$, the maximum likelihood estimates of the weight matrices was shown to be equivalent to the observability matrix and controllability matrix transposed [12].

$$\mathbf{W}^{(1)} = \boldsymbol{\Sigma}_{ff}^{1/2}\mathbf{V}_1\mathbf{P}^{1/2}\mathbf{R} = \mathcal{O}, \qquad \mathbf{W}^{(2)} = \boldsymbol{\Sigma}_{pp}^{1/2}\mathbf{V}_2\mathbf{P}^{1/2}\mathbf{R} = \mathcal{C}^{\mathsf{T}} \tag{6}$$

where $\mathbf{P}^{1/2}$ is the square root of the canonical correlations (singular values) and $\mathbf{R}$ is an abritrary rotation to be recovered.

## 4 Bayesian Stochastic Subspace Identification

Klami and Kaski [8], and Wang [16], extended probabilistic CCA to a Bayesian form through inclusion of prior distributions on the model parameters, as defined in Equations (7 - 9). The resulting Bayesian model is shown graphically in Figure 1.

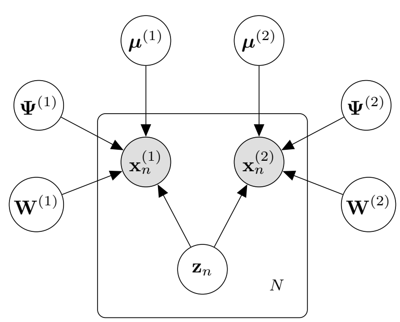


**Fig. 1.** Graphical model of Bayesian CCA

$$\mathbf{w}_i \sim \mathcal{N}(\boldsymbol{\mu}_{w_i}, \boldsymbol{\Sigma}_{w_i}) \tag{7}$$

$$\boldsymbol{\Psi} \sim \mathcal{W}(\mathbf{K}_0, \nu_0) \tag{8}$$

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_\mu, \boldsymbol{\Sigma}_\mu) \tag{9}$$

where $\mathbf{w}_i$ denotes the $i$th column of the full weight matrix $\mathbf{W}$, where each column is considered independent with its own prior of the same form, $\mathcal{W}$ is a Wishart distribution, and the priors over the mean $\boldsymbol{\mu}$ and block precision matrix $\boldsymbol{\Psi}$ are conventional conjugate priors.

Replacing CCA for Bayesian CCA, the authors redefined SSI-Cov as a problem in Bayesian inference, such that the posterior estimates for the weights $\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$ — analogous to posteriors on the observability matrix and controllability matrix — from which, posteriors on the modal properties are obtainable. There are multiple inference schemes available to compute the posteriors for this type of Bayesian problem. In this paper, the authors compare VI, and batch SVI.

## 5 Inference Schemes

### 5.1 Variational Inference

Variational inference (VI) is a common inference scheme used to approximate intractable posterior distributions with variational (surrogate) posteriors. This is achieved by minimising the Kullback-Leibler (KL) divergence from the surrogate to the true posterior distribution: Maximising the evidence lower bound (ELBO), which is equal to the negative KL divergence up to an additive constant. The ELBO is defined as the sum of the expected log of the joint and the entropy of the variational distribution [9, 2].

$$\mathcal{L}(\boldsymbol{\phi}, \lambda) = \mathbb{E}_{q(\mathbf{z},\boldsymbol{\theta})}\left[\log p(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) - \log q_{\boldsymbol{\phi},\lambda}(\mathbf{z}, \boldsymbol{\theta})\right] \tag{10}$$

Using a *mean-field* variational family, in which each latent variable is independent and governed by their own parameters, the surrogate posteriors take the factorised form in Equation (11) [16]

$$q(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\Psi}, \mathbf{W}) = \prod_{n=1}^{N} q(\mathbf{z}_n) \prod_{i=1}^{d} q(\mathbf{w}_i) q(\boldsymbol{\mu}) q(\boldsymbol{\Psi}) \tag{11}$$

with the following definitions for Bayesian SSI

$$q(\mathbf{z}_n) = \mathcal{N}\left(\mathbf{z}_n | \breve{\boldsymbol{\mu}}_{\mathbf{z}_n}, \breve{\boldsymbol{\Sigma}}_{\mathbf{z}_n}\right) \tag{12}$$

$$q(\boldsymbol{\Psi}) = \mathcal{W}\left(\boldsymbol{\Psi} | \breve{\mathbf{K}}, \breve{\boldsymbol{\nu}}\right) \tag{13}$$

$$q(\mathbf{w}_i) = \mathcal{N}\left(\mathbf{w}_i | \breve{\boldsymbol{\mu}}_{\mathbf{w}_i}, \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}\right) \tag{14}$$

$$q(\boldsymbol{\mu}) = \mathcal{N}\left(\boldsymbol{\mu} | \breve{\boldsymbol{\mu}}_{\boldsymbol{\mu}}, \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}\right) \tag{15}$$

where $\phi_n = \{\breve{\boldsymbol{\mu}}_{\mathbf{z}_n}, \breve{\boldsymbol{\Sigma}}_{\mathbf{z}_n}\}$ are local variational parameters and $\lambda = \{\breve{\boldsymbol{\mu}}_{\mathbf{w}_i}, \breve{\boldsymbol{\Sigma}}_{\mathbf{w}_i}, \breve{\boldsymbol{\mu}}_{\boldsymbol{\mu}}, \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}, \breve{\mathbf{K}}^{-1}, \breve{\boldsymbol{\nu}}\}$ are global variational parameters. This allows the local and global parameters of the surrogate posteriors to be determined which, for mean-field

VI, is achieved using coordinate ascent on the gradient of the ELBO [9]. The algorithm first optimises the local parameters for all datapoints, and then re-estimates the global parameters, iterating until convergence of the ELBO.

### 5.2 Batch Stochastic Variational Inference

Stochastic variational inference (SVI), developed by Hoffman et.al. [5], is designed to remove the dependency of coordinate ascent VI on optimising the full set of local parameters using the entire dataset before re-estimating the global parameters. This is particularly useful when analysing larger datasets. SVI uses stochastic optimisation to form noisy estimates of the natural gradients of the ELBO. The method proceeds by first subsampling the dataset and finding the local variational parameters at a single point, $x_i$. A set of intermediate global parameters $\hat{\lambda}$ are then computed as though $x_i$ were repeated $N$ times. This intermediatory is then used to update the current estimate of the global parameters $\lambda^{(t-1)}$ according to, $\lambda^{(t)} = (1-\rho_t)\lambda^{(t-1)} + \rho_t\hat{\lambda}$, where $\rho$ is a decreasing step size [5]. It can be seen that this approach is equivalent to a stochastic optimisation scheme for $\lambda$. These steps are then repeated, uniformly sampling the dataset until convergence. Hoffman also showed how this can be extended to a 'mini-batch' algorithm to improve the algorithm's stability, reducing the variance in the estimates $\hat{\lambda}$. Instead, a batch of $S$ random points are subsampled at each iteration. The local variational parameters are once again computed for each data point and intermediate global parameters are computed in the same way $\hat{\lambda}_s$. However, before updating, the intermediate global estimates are averaged over the batch, such that

$$\lambda^{(t)} = (1-\rho_t)\lambda^{(t-1)} + \frac{\rho_t}{S}\sum_{s=1}^{S}\hat{\lambda}_s \tag{16}$$

This has significant benefits, such as amortising any computational expense associated with updating the global parameters (as is true in the case of Bayesian SSI, incurring this cost less frequently) and helping the algorithm find a better variational posterior.

## 6 Results and Discussion

To compare the performance of the two inference schemes response data from a three-degree of freedom linear dynamic system, described by the modal properties $\omega = \{10.54, 16.35, 24.34\}$ and $\zeta = \{0.0051, 0.0076, 0.0033\}$, were generated given a simple white noise excitation. The system was simulated at a sample rate of 1000Hz, generating 16384 datapoints. SSI-Cov, VI and batch SVI were then applied assuming a larger than true model order of 20 (10 unique modes). A simulated system was used to allow comparison to a known ground truth.

The following weakly informative, proper priors were chosen, $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \mathbb{I})$, $\boldsymbol{\mu} \sim \mathcal{N}\left(\mathbf{0}, \mathbb{I}\times 10^{-3}\right)$, $\boldsymbol{\Psi} \sim \mathcal{W}\left(\mathbb{I}\times 10^{1}, D+1\right)$, providing the model with sufficient flexibility. The priors on the mean and variance were lightly constrained given assumed prior knowledge of a zero mean and low measurement noise. Histograms of the prior modal properties, using samples drawn and propagated from the prior observability matrix, are presented in Figure 2.

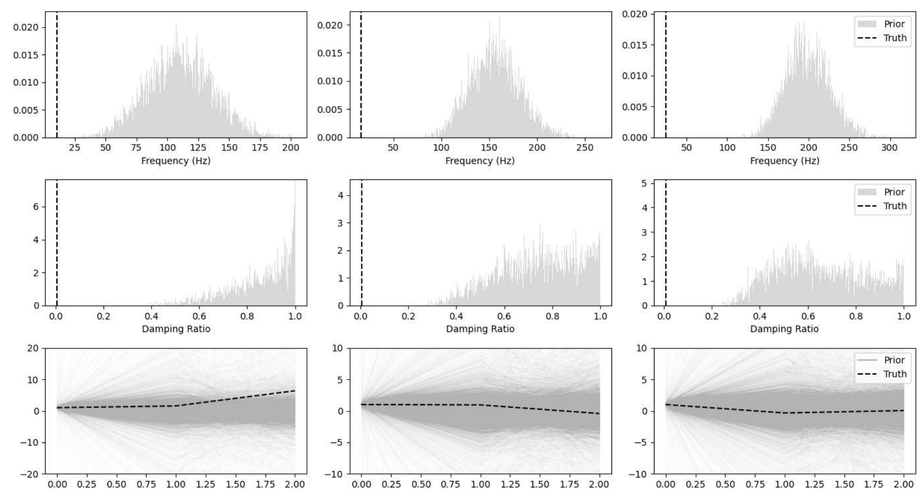


**Fig. 2.** Approximate prior distributions on the modal properties as histograms, propagated from samples of the prior distribution on the observability matrix. (top to bottom: natural frequency, damping ratio, normalised mode shape)

Rather than evaluating both algorithms using compute time, instead the required number of necessary repeats (i.e. the number of times the entire dataset was analysed) to reach convergence of the evidence lower-bound (ELBO) was used. Considering the number of updates on each parameter, avoids inconsistencies in implementation compared to "wall-time" comparison. For the batch SVI analysis, a batch size of $S = 1024$ was chosen, equivalent to 16 batches, with the common step size function $\rho = s^{-k}$ where $s$ is the current batch number and $k$ is the forgetting rate, chosen as $k = 0.95$. Both methods were then applied, with VI converging after 5 iterations (5 full sweeps of the data) and batch SVI converging after 48 batches, equivalent to 3 full sweeps of the data. As can be seen from Figure 3, the approximate posteriors of both algorithms are in relatively good agreement with one another, and with SSI-Cov. This is expected as the posterior mean should converge to the maximum-a-posteriori estimate. Furthermore, the posteriors from batch SVI demonstrate lower overall variance, despite two fewer full sweeps of the data. This is believed to be a result of the stochastic nature of batch SVI, which is often hypothesised to reduce possible stagnation at saddle points.

The reader may also notice the possibility for unrealistic estimates of the damping ratio, which for real mechanical systems should be bounded $\zeta = [0, 1]$, given the quantified uncertainty. Nevertheless, the authors note that classic

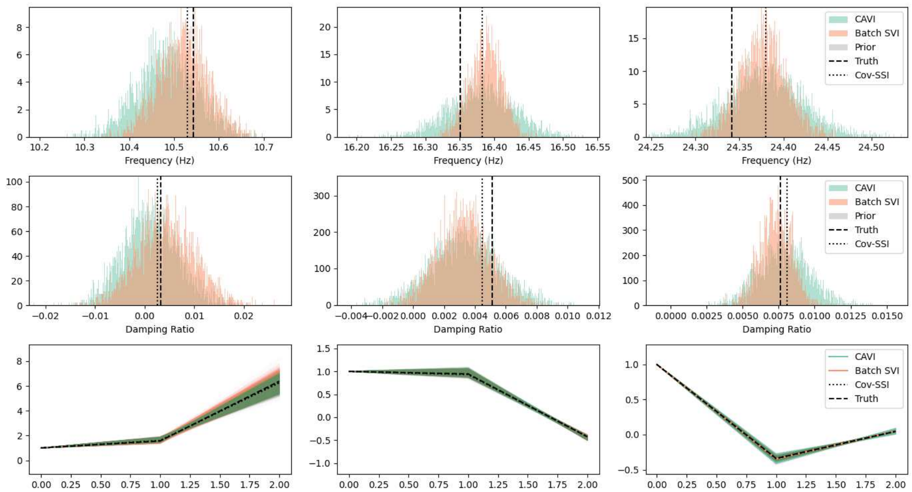


**Fig. 3.** Posterior distributions on the modal properties as histograms, made up of samples drawn from the closed form surrogate posterior on the observability matrix found using CAVI and Batch SVI. These samples are then propagated onto the modal properties. (top to bottom: natural frequency, damping ratio, normalised mode shape)

SSI-Cov algorithm is not mathematically constructed to enforce such a restriction and that negative damping estimates can occur. However, this does pose an interesting research problem which is discussed briefly in the concluding remarks.

## 7 Concluding Remarks

This paper presented a new efficient implementation of Bayesian SSI by replacing variational inference (VI), with batch stochastic variational inference (SVI). The two inference schemes were applied to a simulated three degree-of-freedom linear dynamic system, demonstrating comparable performance in the recovery of posterior distributions on the modal properties. Notably, batch SVI was able to achieve convergence to a reasonable posterior in fewer full sweeps of the dataset (3 repeats) than traditional VI (5 repeats). It is believed that this improved version of Bayesian SSI makes it a competitive choice for Bayesian uncertainty quantification in industrial and research applications. Following this body of work, other future work will aim to address the issue of recovering physically meaningful posterior estimates to the modal properties. This may include the exploration and development of a novel approach to embedding physical understanding into the priors in this practical context. Other important points for consideration will be studying the effect of hyperparameters, such as batch size and forgetting rate, and considering how model order selection could be achieved efficiently in this framework.

## Acknowledgements

The authors gratefully acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC), UK through grant number EP/W002140/1. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

## References

1. Bach, F.R., Jordan, M.I.: A probabilistic interpretation of canonical correlation analysis. Dept. Statist., Univ. California, Berkeley, CA, Tech. Rep **688**, 1–11 (2005)
2. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Cambridge (2006)
3. Döhler, M., Lam, X.B., Mevel, L.: Uncertainty quantification for modal parameters from stochastic subspace identification on multi-setup measurements. Mechanical Systems and Signal Processing **36**(2), 562–581 (2013)
4. Döhler, M., Mevel, L.: Efficient multi-order uncertainty computation for stochastic subspace identification. Mechanical Systems and Signal Processing **38**(2), 346–366 (2013)
5. Hoffman, M.D., Blei, D.M., Wang, C., Paisley, J.: Stochastic Variational Inference. Journal of Machine Learning Research **14**, 1303–1347 (2013)
6. Hotelling, H.: Relations Between Two Sets of Variates. Biometrika **28**(3/4), 321 (1936)
7. Katayama, T.: Subspace Methods for System Identification: A Realisation Approach. Springer (2005)
8. Klami, A., Kaski, S.: Local dependent components. In: ACM International Conference Proceeding Series. vol. 227, pp. 425–432 (2007)
9. Murphy, K.P.: Probabilistic Machine Learning: An Introduction (Adaptive Computation and Machine Learning), vol. Massachuse (2021)
10. O'Connell, B.J., Cross, E.J., Tygesen, U.T., Rogers, T.J.: Bayesian canonical correlations for stochastic subspace identification. In: 30th ISMA, 9th USD - International Symposium on Modal Analysis and International Conference on Uncertainty in Structural Dynamics (2022)
11. O'Connell, B.J., Champneys, M.D., Cross, E.J., Rogers, T.J.: A novel variational Bayesian approach to stochastic subspace identification. In: 5th ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Science and Engineering (UNCECOMP). pp. 82–93 (2023)
12. O'Connell, B.J., Rogers, T.J.: A Robust Probabilistic Approach to Stochastic Subspace Identification. arxiv.org/abs/2305.16836 (Preprint submitted to the Journal of Sound and Vibration) (2023)
13. Reynders, E., Pintelon, R., Roeck, G.D.: Uncertainty bounds on modal parameters obtained from Stochastic Subspace Identification. Mechanical Systems and Signal Processing **22**(4), 948–969 (2008)
14. Reynders, E.P.: Uncertainty quantification in data-driven stochastic subspace identification. Mechanical Systems and Signal Processing **151**, 107338 (2021)
15. Van Overschee, P., De Moor, B.: Subspace Identification for Linear Systems. Springer (1996)
16. Wang, C.: Variational Bayesian Approach to Canonical Correlation Analysis. IEEE Transactions on Neural Networks **18**(3), 905–910 (2007)