



This is a repository copy of *A robust probabilistic approach to stochastic subspace identification*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/202072/>

Version: Submitted Version

---

**Preprint:**

O'Connell, B.J. [orcid.org/0000-0001-6042-927X](https://orcid.org/0000-0001-6042-927X) and Rogers, T.J. (Submitted: 2023) A robust probabilistic approach to stochastic subspace identification. [Preprint - arXiv] (Submitted)

<https://doi.org/10.48550/arXiv.2305.16836>

---

© 2023 The Author(s). This preprint is made available under a Creative Commons Attribution 4.0 International License. (<https://creativecommons.org/licenses/by/4.0/>)

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# A Robust Probabilistic Approach to Stochastic Subspace Identification

B. J. O'Connell <sup>a,\*</sup>, T. J. Rogers <sup>a</sup>

<sup>a</sup>*Dynamics Research Group, Department of Mechanical Engineering, The University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK*

---

## Abstract

Modal parameter estimation of operational structures is often a challenging task when confronted with unwanted distortions (outliers) in field measurements. Atypical observations present a problem to operational modal analysis (OMA) algorithms, such as stochastic subspace identification (SSI), severely biasing parameter estimates and resulting in misidentification of the system. Despite this predicament, no simple mechanism currently exists capable of dealing with such anomalies in SSI. Addressing this problem, this paper first introduces a novel probabilistic formulation of stochastic subspace identification (Prob-SSI), realised using probabilistic projections. Mathematically, the equivalence between this model and the classic algorithm is demonstrated. This fresh perspective, viewing SSI as a problem in probabilistic inference, lays the necessary mathematical foundation to enable a plethora of new, more sophisticated OMA approaches. To this end, a statistically robust SSI algorithm (robust Prob-SSI) is developed, capable of providing a principled and automatic way of handling outlying or anomalous data in the measured timeseries, such as may occur in field recordings, e.g. intermittent sensor dropout. Robust Prob-SSI is shown to outperform conventional SSI when confronted with ‘corrupted’ data, exhibiting improved identification performance and higher levels of confidence in the found poles when viewing consistency (stabilisation) diagrams. Similar benefits are also demonstrated on the Z24 Bridge benchmark dataset, highlighting enhanced performance on measured systems.

*Keywords:* Probabilistic, System Identification, Stochastic Subspace Identification, Robust, Operational Modal Analysis

---

\*Corresponding Author

*Email address:* [bjconnell11@sheffield.ac.uk](mailto:bjconnell11@sheffield.ac.uk) (B. J. O'Connell )

## 1. Introduction

The characterisation of structural dynamic systems remains to be a key feature of modern engineering practice. Modal analysis is an established field of system identification that is frequently employed across industry and research to practically analyse (linear) dynamic systems with a view to recover estimates for the underlying invariant (modal) properties. These are familiar to the dynamicist as the natural frequencies, damping ratios and mode shapes. The availability of this modal information helps facilitate more informed decision-making throughout the entire lifecycle of engineering structures.

Subspace identification algorithms have gained increasing attention in the modal analysis community [1], predominantly due to their ability to deal with a considerable number of inputs and outputs. Stochastic subspace identification (SSI) is a prominent method capable of estimating a linear time-invariant state-space model from correlated sequences of observed data, using traditional linear algebra techniques. Once recovered, an eigen-decomposition of the state-space model yields the desired modal properties. Frequently appearing in both academic and industrial literature, SSI (and its many algorithmic variants [2, 3]) continues to be employed as reliable means of operational modal analysis (OMA) [1, 4–7]. In this particularly paper, we focus attention to covariance-driven stochastic subspace identification (Cov-SSI). Schematically, the Cov-SSI procedure is shown on the left hand side of Figure 1 and described in full in Section 2.2.

Despite the success of OMA algorithms, a typical weakness lies in their handling of atypical observations and non-Gaussian noise. In practice, in situ monitoring can often produce imperfect data containing unwanted features that may present a problem to OMA algorithms, severely biasing parameter estimates and inevitably leading to the misidentification of a system. These features arise due to the low level of excitation, the inherent stochastic nature of the forcing or from unpredicted events that are often independent of the system being measured. In the case of OMA, these ‘events’ refer to practically encountered scenarios in testing such as sensor drop-out or partial detachment.

This misidentification problem is also true of Cov-SSI. Unlike classical approaches to data analysis, where data may be pre-processed to remove outliers and ‘cleaned’ prior to use, such an approach cannot easily be applied to Cov-SSI due to its dependency on sequential data, as required for the Hankel matrix formation. This poses a significant dilemma. During application of SSI, any outliers remain present during the analysis, distorting the measurement of the response and ultimately affecting the identification procedure. Despite this discernible predicament, no simple mechanism currently exists capable of dealing with such anomalies in SSI.

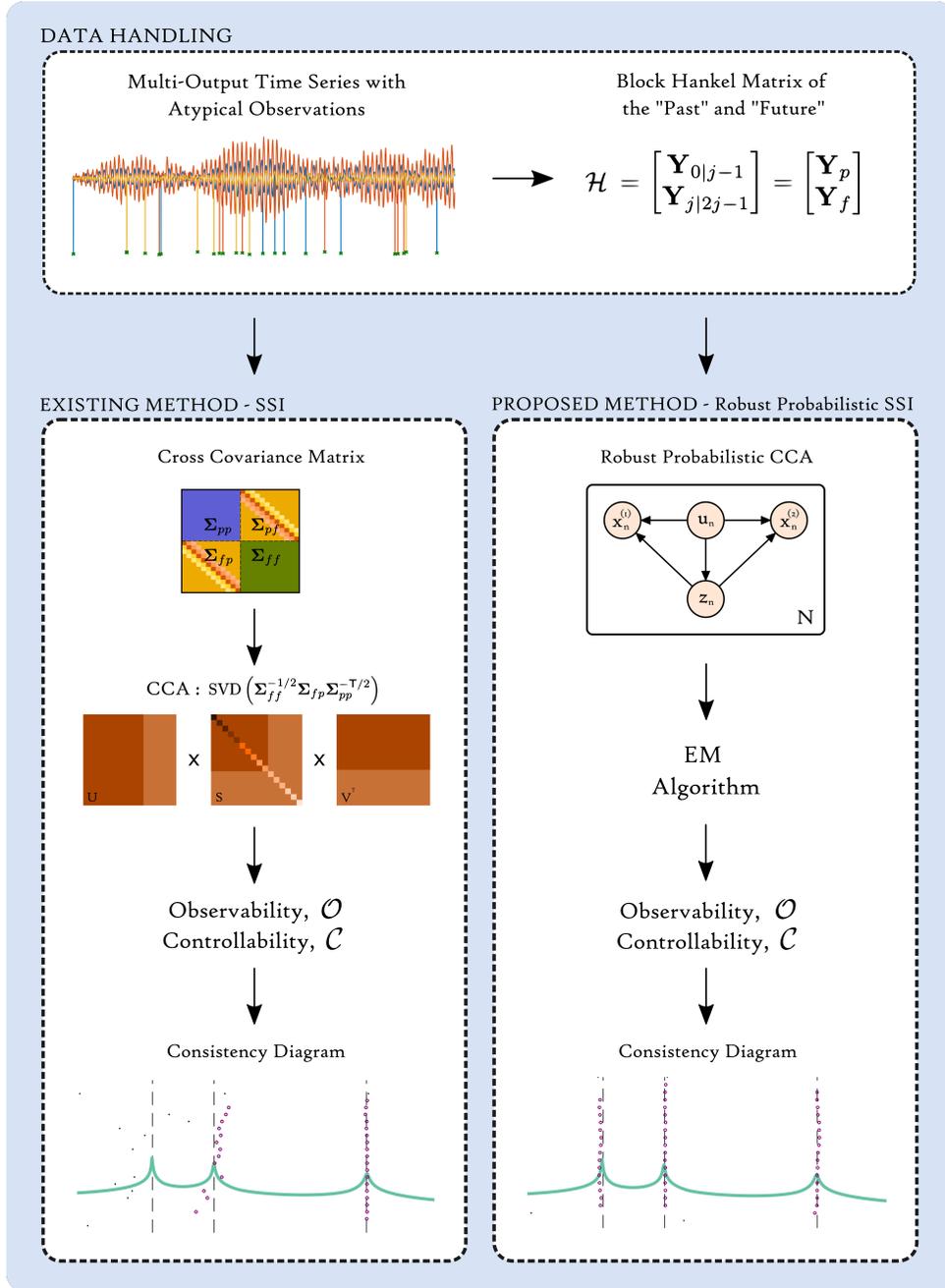


Figure 1: Schematic of the existing methodology, SSI, and the proposed methodology, Robust Prob-SSI, including an example of the observed relative improvement to identification performance from robust Prob-SSI.

Addressing this problem, this paper introduces a novel probabilistic formulation of SSI, referred to as Prob-SSI, realised using the theory of probabilistic projections [8] and latent variable models. This new formulation is achieved through substitution of

canonical correlation analysis (CCA) (i.e. the singular value decomposition (SVD)) with probabilistic CCA [8] within the SSI procedure. This fresh perspective, viewing SSI as a problem in probabilistic inference, now lays the appropriate mathematical foundation to enable a plethora of new SSI-based OMA algorithms, constructed using more sophisticated probabilistic and hierarchical techniques.

To that end, the authors also present the first of many new algorithms constructed on this unique model - a statistically robust SSI algorithm (robust Prob-SSI) capable of providing a principled and automatic way of handling atypical observations in multi-output timeseries responses. As will be demonstrated, when confronted with ‘corrupted’ data, this new approach to modal identification appears to outperform traditional SSI, exhibiting improved pole consistency and increased confidence; attributes reflected in the subsequent consistency diagrams. This new proposed methodology is shown on the right hand side of Figure 1 and discussed in Section 4. Further investigation and application to the well-documented Z24 Bridge benchmark is also conducted, with results highlighting similar benefits to identification performance.

### *1.1. Related Work*

The application of probabilistic methods in modern engineering practice, coupled with the demonstrative value associated with quantifying and harnessing uncertainty through probabilistic analysis, has led to a general shift across many engineering disciplines to adopt probabilistic tools and operate within more probabilistic frameworks. In the realm of structural dynamics, this is most evident in the areas of SHM [9], digital twins [10] and more increasingly in the field of system identification. The ability to reformulate engineering operations probabilistically not only simplifies their inclusion into probabilistic frameworks but can often provide an opportunity to improve existing methods, by means previously unattainable using a deterministic approach.

Probabilistic (and Bayesian) approaches to modal analysis are now becoming increasingly popular within the dynamics community [11]. The inclusion of uncertainty, in its various forms<sup>1</sup>, has proven increasingly useful; aiding the automatic selection of consistent poles in stabilisation diagrams [12, 13], quantifying the uncertainty over the modal properties [14–17] and by providing an alternative foundation upon which new methods can be developed [18].

Despite many alternative approaches to OMA and an increase in probabilistic identification techniques, few directly attempt to formulate statistically robust algorithms capable of handling atypical observations. Most OMA algorithms possess some level of inherent robustness given their dependency on certain mathematical operations, such as the singular

---

<sup>1</sup>It is important to note, that there are many types of uncertainty, whether that be a posterior uncertainty in a Bayesian sense, estimates for bounds, confidence intervals, fiducial intervals, etc. The discussion of their relative merits lies beyond this work.

value decomposition, but in certain cases, as will be shown, these algorithms are incapable of dealing with a considerable number of outliers.

Some ‘robust’ approaches to OMA do exist in literature, however, the term ‘robust’ is often used to loosely refer to methods capable of reliable performance, rather than statistical robustness. Most of these methods achieve consistency through automated process; using optimised metrics or machine learning [19], often with the inclusion of uncertainty in techniques such as clustering [13]. In contrast, this work achieves statistical robustness through direct augmentation of the underlying algorithm. Furthermore, upon closer inspection of the literature, only a small selection of papers appear to directly address the problem of outliers, including recently developed robust algorithms by Liu et.al. on correlation signal subset-based SSI (CoS-SSI) [20] and Goursat et.al. [21] on Crystal Clear SSI (CC-SSI); both concerned with addressing non-stationary and noisy signals.

## 1.2. Structure

To assist navigation through the remainder of the article, the structure is given as follows:

- **(Section 2)** provides the necessary theoretical background for canonical correlation analysis **(2.1)** and covariance-driven stochastic subspace identification **(2.2)**.
- **(Section 3)** introduces the theory of probabilistic canonical correlation analysis **(3.1)**. This is followed by a definition of the novel probabilistic stochastic subspace identification algorithm **(3.2)**.
- **(Section 4)** introduces the existing theory of robust probabilistic canonical correlation analysis **(4.1)**, including the necessary expectation-maximisation update equations **(4.1.1)**, followed by a definition of the new robust probabilistic stochastic subspace identification algorithm **(4.2)**.
- **(Section 5)** presents results and discussion from three individual case studies: a ‘clean’ benchmark case **(5.1)**, a ‘corrupted’ case containing atypical observations **(5.2)** and a real world case, using the Z24 bridge benchmark **(5.3)**.
- **(Section 6)** provides concluding remarks and an outward look at possible avenues for future work.

## 2. Theoretical background

### 2.1. Canonical Correlation Analysis

Canonical correlation analysis (CCA) is a well established statistical tool [22] that is used to analyse the mutual dependency between two sets of multidimensional variables. This dependency can be quantified by finding an appropriate set of orthogonal basis vectors,  $\mathbf{a}$  and  $\mathbf{b}$ , such that the correlation between the linear projections<sup>2</sup> of the variables is mutually maximised, subject to the constraint that the set of transformed variables are uncorrelated<sup>3</sup>. Given two co-occurring multidimensional random variables,  $\mathbf{x} \in \mathbb{R}^{d_1}$  and  $\mathbf{y} \in \mathbb{R}^{d_2}$ , this process can be represented mathematically as

$$(\mathbf{a}', \mathbf{b}') = \arg \max_{\mathbf{a}, \mathbf{b}} \text{corr}(\mathbf{a}^\top \mathbf{x}, \mathbf{b}^\top \mathbf{y}) = \arg \max_{\mathbf{a}, \mathbf{b}} \frac{\mathbf{a}^\top \Sigma_{xy} \mathbf{b}}{\sqrt{\mathbf{a}^\top \Sigma_{xx} \mathbf{a} \mathbf{b}^\top \Sigma_{yy} \mathbf{b}^\top}} \quad (1)$$

where  $\Sigma_{xy}$  is defined as the cross covariance matrix between  $\mathbf{x}$  and  $\mathbf{y}$ , with  $\Sigma_{xy} = \Sigma_{yx}^\top$ , and  $\Sigma_{xx}$  and  $\Sigma_{yy}$  are the auto covariance matrices. The canonical correlations can be computed easily by solving the generalised eigenvalue problem

$$\begin{pmatrix} 0 & \Sigma_{xy} \\ \Sigma_{yx} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \lambda \begin{pmatrix} \Sigma_{xx} & 0 \\ 0 & \Sigma_{yy} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \quad (2)$$

where  $\mathbf{a}$  and  $\mathbf{b}$  are the eigenvectors and  $\lambda$  is the eigenvalue or canonical correlation. In practice the set of eigenvalues and eigenvectors are computed using the singular value decomposition (SVD) of  $\Sigma_{xx}^{-\frac{1}{2}} \Sigma_{xy} \Sigma_{yy}^{-\frac{1}{2}} = \mathbf{V}_1 \mathbf{\Lambda} \mathbf{V}_2^\top$ .

### 2.2. Covariance-Driven Stochastic Subspace Identification (Cov-SSI)

This section provides a concise overview of the theory underpinning Cov-SSI for an output-only case. For the benefit of the reader, the subtle presence of CCA in the method is highlighted. This derivation follows descriptions given by Katayama [23], Van Overschee and De Moor [24] and Döhler and Mevel [25]. The reader is directed towards the aforementioned texts for a fuller derivation and for specifics. Only important details are repeated here. For ease, Cov-SSI will be synonymous with the abbreviation ‘SSI’ if used, throughout the remainder of the paper.

Consider an  $r^{\text{th}}$  order discrete state space model of a linear dynamic system, equivalent to a mechanical system with  $n_{\text{dof}}$  degrees of freedom, such that  $r = 2n_{\text{dof}}$ , in the form

---

<sup>2</sup>By assuming linear projections, the results of CCA are easily interpretable whilst keeping the problems of overfitting mostly manageable.

<sup>3</sup>One intuitive explanation of CCA is that it is performing PCA on two datasets  $\mathbf{x}$  and  $\mathbf{y}$  whilst simultaneously attempting to maximise the correlation of the two sets of principal components.

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}_d \mathbf{x}_k + w_k \\ \mathbf{y}_k &= \mathbf{C} \mathbf{x}_k + v_k \end{aligned}, \quad \mathbb{E} \left\{ \begin{bmatrix} w_q \\ v_q \end{bmatrix} \begin{bmatrix} w_s^\top & v_s^\top \end{bmatrix} \right\} = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^\top & \mathbf{R} \end{bmatrix} \delta_{qs} \quad (3)$$

where  $\mathbf{y}_k \in \mathbb{R}^l$  is the output vector at discrete time step  $k$ ,  $\mathbf{x}_k \in \mathbb{R}^p$  is the internal state vector,  $\mathbf{A}_d \in \mathbb{R}^{p \times p}$  is the discrete state matrix such that  $\mathbf{A}_d = \exp(\mathbf{A}_c \Delta t)$  where  $\mathbf{A}_c$  is the continuous state matrix,  $\Delta t$  is the sampling time and  $\mathbf{C}$  is the output matrix.  $w_k \in \mathbb{R}^p$  and  $v_k \in \mathbb{R}^l$  are samples of the process noise and measurement noise respectively,  $\mathbb{E}[\cdot]$  denotes the expectation and  $\delta_{qs}$  is the Kronecker delta for any two samples in time  $q$  and  $s$ . The process and measurement noise are both assumed to be stationary, white noise Gaussian with zero mean and covariance given by the second part of Equation 3.

Given output measurements from this stationary process, based on all  $l$  measurement channels, the data can be arranged into a block Hankel matrix, in the usual way, to form

$$\mathcal{H} = \mathbf{Y}_{0|2j-1} \in \mathbb{R}^{2lj \times N} = \begin{bmatrix} \mathbf{Y}_{0|j-1} \\ \mathbf{Y}_{j|2j-1} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_p \\ \mathbf{Y}_f \end{bmatrix} \quad (4)$$

with  $2j$  block rows and  $N$  columns, with every block consisting of  $l$  rows and where  $j > 0$  and  $N$  is sufficiently large (i.e much larger than  $2lj$ ) and where  $j > r$  (model order) and the number of columns of block matrices is  $N$ . The resultant cross-covariance matrix of the future  $\mathbf{Y}_f$  with the past  $\mathbf{Y}_p$  is therefore given by

$$\tilde{\Sigma} = \frac{1}{N} \begin{bmatrix} \mathbf{Y}_p \\ \mathbf{Y}_f \end{bmatrix} \begin{bmatrix} \mathbf{Y}_p^\top & \mathbf{Y}_f^\top \end{bmatrix} = \begin{bmatrix} \Sigma_{pp} & \Sigma_{pf} \\ \Sigma_{fp} & \Sigma_{ff} \end{bmatrix} \quad (5)$$

where  $\Sigma_{pf}$  and  $\Sigma_{fp}$  are block cross-covariance matrices, and  $\Sigma_{ff}$ ,  $\Sigma_{pp}$  are block auto-covariance matrices respectively. The canonical correlations  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$  between the future and past are the singular values [23], obtained through the singular value decomposition (SVD) of the following matrix

$$\Sigma_{ff}^{-1/2} \Sigma_{fp} \Sigma_{pp}^{-\top/2} = \mathbf{V}_1 \Lambda \mathbf{V}_2^\top \simeq \check{\mathbf{V}}_1 \check{\Lambda} \check{\mathbf{V}}_2^\top \quad (6)$$

where  $\Sigma_{ff}^{1/2} \Sigma_{ff}^{\top/2} = \Sigma_{ff}$ , such that,

$$\Sigma_{fp} \simeq \Sigma_{ff}^{1/2} \check{\mathbf{V}}_1 \check{\Lambda} \check{\mathbf{V}}_2^\top \Sigma_{pp}^{\top/2} \quad (7)$$

where  $\mathbf{V}_1$  and  $\mathbf{V}_2$  are the left and right singular vectors, respectively and  $\check{\Lambda}$  neglects sufficiently small singular values (canonical correlations) in  $\Lambda$  such that the resultant state

vector has the dimension  $d = \dim(\check{\Lambda})$ . The cross-covariance matrix,  $\Sigma_{pf}$ , can be decomposed into the corresponding extended observability ( $\mathcal{O}$ ) and controllability ( $\mathcal{C}$ ) matrices using  $\Sigma_{fp} = \mathcal{O}\mathcal{C}$  where

$$\mathcal{O} = \Sigma_{ff}^{1/2} \check{\mathbf{V}}_1 \check{\Lambda}^{1/2}, \quad \mathcal{C} = \check{\Lambda}^{1/2} \check{\mathbf{V}}_2^T \Sigma_{pp}^{T/2} \quad (8)$$

respectively and where  $\text{rank}(\mathcal{O}) = \text{rank}(\mathcal{C}) = d$ . The extended controllability and observability matrices can be used to recover the state  $\mathbf{A}$  and output  $\mathbf{C}$  matrices,<sup>4</sup> and therefore the modal properties, in the usual manner for operational modal analysis (see Appendix C). This overall process is summarised by the schema shown in Figure 2.

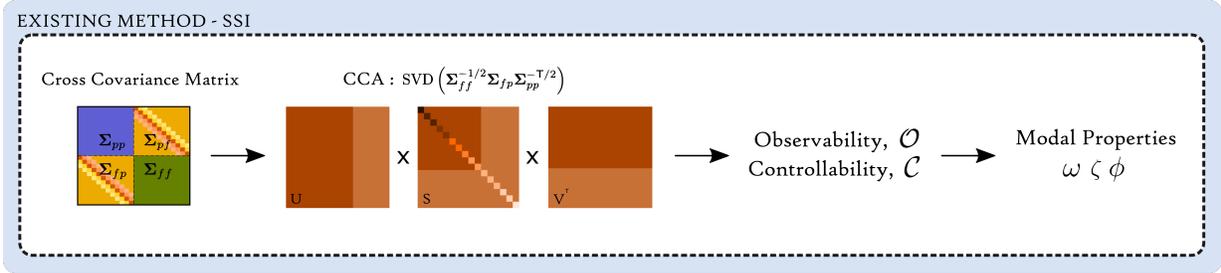


Figure 2: Schematic of the SSI procedure, highlighting the presence of CCA.

### 3. Probabilistic Stochastic Subspace Identification

#### 3.1. Probabilistic CCA (PCCA)

The probabilistic interpretation of CCA, developed by Bach and Jordan [8], uses a latent model approach, similar to that applied by Tipping and Bishop [26] in their earlier development of probabilistic principal component analysis. Probabilistic CCA is summarised here and, for the remainder of the paper, will be abbreviated to PCCA. The term CCA will solely refer to the traditional CCA method.

At the core of PCCA lies a lower dimensional, unobserved latent space described by variable  $\mathbf{z}_n \in \mathbb{R}^d$ . This latent variable is transformed through two independent linear mappings<sup>5</sup>,  $\mathbf{W}^{(1)}$  and  $\mathbf{W}^{(2)}$ , to produce a pair of observed variables  $\mathbf{x}_n^{(1)} \in \mathbb{R}^{D_1}$  and  $\mathbf{x}_n^{(2)} \in \mathbb{R}^{D_2}$ . Crucially, as the observed variables are assumed to be generated from a shared latent space, the two datasets can be considered correlated. The full set of observed samples are given by matrices  $\mathbf{X}^{(m)} = [\mathbf{x}_1^{(m)}, \dots, \mathbf{x}_N^{(m)}] \in \mathbb{R}^{D_m \times N}$  where  $m = 1, 2$  and  $N$

<sup>4</sup>The method described here for recovering the system matrices is one of many. See Stochastic Balanced Realisation Algorithm B, Chapter 8 in Katayama [23]

<sup>5</sup>Note that the transformations defined by Bach and Jordan transform variables from the latent space to the data space.

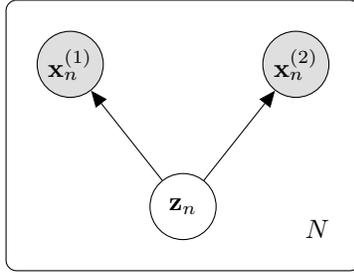


Figure 3: Graphical model for the probabilistic, latent variable interpretation of CCA (PCCA)

is the total number of observations. To simplify notation,  $\mathbf{x}_n = [\mathbf{x}_n^{(1)}; \mathbf{x}_n^{(2)}] \in \mathbb{R}^D$  and  $\mathbf{X} = [\mathbf{X}^{(1)}; \mathbf{X}^{(2)}] \in \mathbb{R}^{D \times N}$  is the feature-wise row concatenation of the two random variables, where  $D = D_1 + D_2$ . PCCA is parameterised by the model parameters  $\theta = \{\mathbf{W}, \Sigma, \boldsymbol{\mu}, \mathbf{Z}\}$  where the latent variables are denoted by  $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N]$ , one for each pair of observed samples, and the remaining variables are the independent model parameters. Given this list of assumptions, Bach and Jordan defined the following probabilistic model, shown graphically in Figure 3 and given mathematically through Equations 9 - 11.

$$\mathbf{z}_n \sim \mathcal{N}(0, \mathbb{I}) \quad (9)$$

$$\mathbf{x}_n^{(m)} | \mathbf{z}_n \sim \mathcal{N}(\mathbf{W}^{(m)} \mathbf{z}_n + \boldsymbol{\mu}^{(m)}, \Sigma^{(m)}) \quad (10)$$

$$\mathbf{x}_n | \mathbf{z}_n \sim \mathcal{N}(\mathbf{W} \mathbf{z}_n + \boldsymbol{\mu}, \Sigma) \quad (11)$$

where  $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$  corresponds to a Gaussian distribution characterised by mean and covariance,  $\mathbf{W} = [\mathbf{W}^{(1)}; \mathbf{W}^{(2)}]$ ,  $\boldsymbol{\mu} = [\boldsymbol{\mu}^{(1)}; \boldsymbol{\mu}^{(2)}]$  and  $\Sigma$  is a block-diagonal covariance matrix with  $\Sigma^{(1)}$  and  $\Sigma^{(2)}$  along the diagonal. An isotropic noise model is assumed in the latent space, which enforces independence between the variables whilst imposing a maximum correlation condition.

### 3.1.1. Maximum Likelihood Estimates

Using this model, Bach and Jordan demonstrated the equivalence of probabilistic CCA to traditional CCA by proving that the maximum likelihood estimates (MLE) for the parameters in Equation 10 lead to the results of classical CCA and therefore, contain all the necessary information traditionally obtained through the SVD. The MLE of the weights are given by

$$\hat{\mathbf{W}}^{(1)} = \Sigma_{11}^{1/2} \mathbf{V}_1 \mathbf{P}^{1/2} \mathbf{R} \quad (12)$$

$$\hat{\mathbf{W}}^{(2)} = \Sigma_{22}^{1/2} \mathbf{V}_2 \mathbf{P}^{1/2} \mathbf{R} \quad (13)$$

where  $\mathbf{V}_1$  and  $\mathbf{V}_2$  are the left and right singular vectors of  $\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2}$ , respectively,  $\mathbf{P}$  is a diagonal matrix of the largest  $d$  canonical correlations and  $\mathbf{R}$  is an arbitrary rotation matrix of size  $d$ . The MLE of the means  $\mu_1$  and  $\mu_2$  are not stated here as the assumption is zero mean data but their addition, if required, is trivial.

### 3.1.2. Expectation-Maximisation Algorithm

A convenient property of latent variable models is that model parameter estimates can also be recovered iteratively using the expectation-maximisation (EM) algorithm [27]. Despite fully tractable MLE estimates, Bach and Jordan also provide EM update equations for PCCA (see Section 4.1 in [8]).

### 3.2. Probabilistic SSI

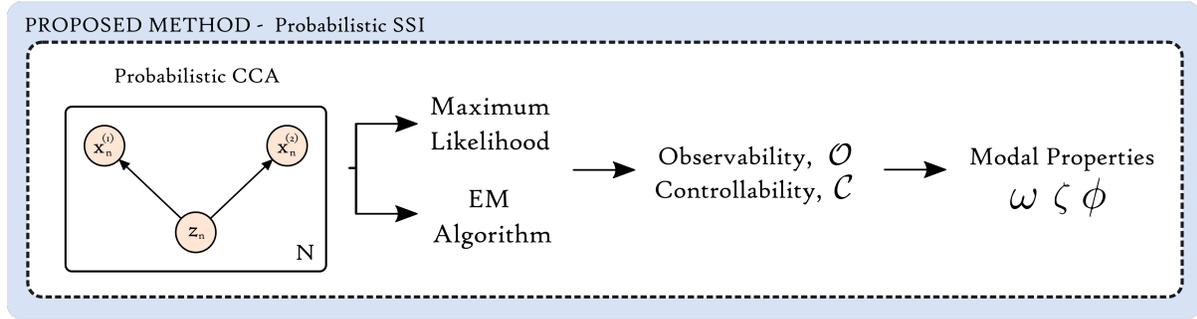


Figure 4: Schematic of the probabilistic SSI procedure, highlighting the replacement of CCA (see Figure 2) with probabilistic CCA.

Through substitution of CCA with PCCA in the SSI algorithm, where  $\mathbf{X}^{(1)} = \mathbf{Y}_f$  and  $\mathbf{X}^{(2)} = \mathbf{Y}_p$ , it can be shown that the MLE estimates for the weights  $\hat{\mathbf{W}}^{(m)}$ , when written in terms of covariance matrices of the future and past, are (as expected) equivalent to the observability matrix and controllability matrix transposed, respectively (Equations 14 - 15). The reader is referred back to Equation 8.

$$\hat{\mathbf{W}}^{(1)} = \Sigma_{ff}^{1/2} \mathbf{V}_1 \mathbf{\Lambda}^{1/2} \mathbf{R} = \mathcal{O} \quad (14)$$

$$\hat{\mathbf{W}}^{(2)} = \Sigma_{pp}^{1/2} \mathbf{V}_2 \mathbf{\Lambda}^{1/2} \mathbf{R} = \mathcal{C}^\top \quad (15)$$

Having established the relationship between the MLE result for the weight matrices and the observability and controllability matrices, this new probabilistic formulation provides the necessary mathematical foundation upon which new probabilistic and hierarchical SSI algorithms can be implemented. Such algorithms promise to enhance identification performance, aid recovery of uncertainty and/or help extract new unseen information. The first of these new probabilistic approaches to SSI can now be shown, specifically the presentation of a statistically robust method capable of automatically handling outlying data in measured timeseries.

## 4. Robust Probabilistic SSI (Robust Prob-SSI)

### 4.1. Robust Probabilistic CCA

Standard probabilistic models can be converted to a more ‘robust’ form through the replacement of a Gaussian noise model to that of a Student’s t-distribution [28]. Building on the probabilistic interpretation of CCA, Archambeau, Delannay and Verleysen [29] constructed a statistically robust equivalent of PCCA. This alternative model is founded upon two key assumptions. The first is that the observed and latent variables can both be modelled by a Student’s t-distribution  $\mathcal{S}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$  with mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ . The Student’s t-distribution has heavier tails than a typical Gaussian, which are determined by the additional parameter,  $\nu$  (degrees of freedom). The presence of heavier tails is preferable as it makes the T-distribution better equipped to handle outliers as it is more likely to capture them within the distribution. The second assumption is that an outlier in the feature space must manifest as an outlier in the latent space. Given this set of assumptions, Archambeau et. al. presented the following probabilistic model (note a continuation of notation from Section 3.1)

$$\mathbf{z}_n \sim \mathcal{S}(\mathbf{z}_n | \mathbf{0}, \mathbb{I}_d, \nu) \quad (16)$$

$$\mathbf{x}_n^{(m)} | \mathbf{z}_n \sim \mathcal{S}(\mathbf{x}_n^{(m)} | \mathbf{W}^{(m)} \mathbf{z}_n + \boldsymbol{\mu}^{(m)}, \boldsymbol{\Sigma}^{(m)}, \nu) \quad (17)$$

Making the set of latent variables explicit, the model can be reformulated into the following form

$$u_n \sim \mathcal{G}\left(u_n \mid \frac{\nu}{2}, \frac{\nu}{2}\right) \quad (18a)$$

$$\mathbf{z}_n | u_n \sim \mathcal{N}(\mathbf{z}_n | \mathbf{0}, u_n^{-1} \mathbb{I}_d) \quad (18b)$$

$$\mathbf{x}_n^{(m)} | \mathbf{z}_n, u_n \sim \mathcal{N}\left(\mathbf{x}_n^{(m)} | \mathbf{W}^{(m)} \mathbf{z}_n + \boldsymbol{\mu}^{(m)}, u_n^{-1} \boldsymbol{\Sigma}^{(m)}\right) \quad (18c)$$

$$\mathbf{x}_n | \mathbf{z}_n, u_n \sim \mathcal{N}(\mathbf{x}_n | \mathbf{W} \mathbf{z}_n + \boldsymbol{\mu}, u_n^{-1} \boldsymbol{\Sigma}) \quad (18d)$$

where  $\mathcal{G}(\alpha, \beta)$  represents a Gamma distribution,  $u_n$  is an additional latent variable and  $\mathbf{W}$ ,  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$  and  $\nu$  are as previously defined. The corresponding graphical model is shown in Figure 5.

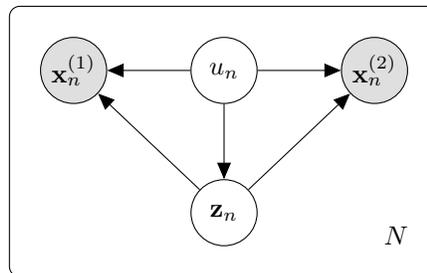


Figure 5: Graphical model for robust probabilistic canonical correlation analysis (robust-PCCA)

#### 4.1.1. Expectation-Maximisation Algorithm

In contrast to PCCA, direct maximisation of the incomplete data log-likelihood  $\sum_n \log p(\mathbf{x}_n)$  with respect to the parameters is intractable. However, using equations (18a-18d) estimates for the variables can be recovered through an iterative scheme. Archambeau et.al employ the EM approach, maximising the the expected complete-log-likelihood [29].

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{x}_n, \mathbf{z}_n, u_n) = \sum_{n=1}^N \ln p(\mathbf{x}_n, \mathbf{z}_n, u_n|\boldsymbol{\theta}) \quad (19)$$

where  $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\Sigma}, \nu)$ .

The necessary update equations in the E-step (expectation) and M-step (maximisation) of robust PCCA are given by the equations 20-27. For completeness, the robust PCCA algorithm has been re-derived<sup>6</sup> in Appendix A.

E-Step

$$\bar{u}_n = \frac{D + \nu}{u_n(\mathbf{x}_n - \boldsymbol{\mu})^\top \mathbf{A}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) + \nu} \quad (20)$$

$$\ln \tilde{u}_n = \psi\left(\frac{D + \nu}{2}\right) - \ln\left(\frac{u_n(\mathbf{x}_n - \boldsymbol{\mu})^\top \mathbf{A}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) + \nu}{2}\right) \quad (21)$$

$$\bar{\mathbf{z}}_n = \mathbf{B}^{-1} \mathbf{W}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) \quad (22)$$

$$\bar{\mathbf{S}}_n = \mathbf{B}^{-1} + \bar{u}_n \bar{\mathbf{z}}_n \bar{\mathbf{z}}_n^\top \quad (23)$$

where  $\mathbf{A} = (\boldsymbol{\Sigma} + \mathbf{W}\mathbf{W}^\top)$ ,  $\mathbf{B} = \mathbf{W}^\top \boldsymbol{\Sigma}^{-1} \mathbf{W} + \mathbb{I}_d$ ,  $\bar{u}_n = \mathbb{E}[u_n]$ ,  $\ln \tilde{u}_n = \mathbb{E}[\ln(u_n)]$ ,  $\bar{\mathbf{z}}_n = \mathbb{E}[\mathbf{z}_n]$ ,  $\bar{\mathbf{S}}_n = \mathbb{E}[u_n \mathbf{z}_n \mathbf{z}_n^\top]$  and  $\psi(\cdot)$  denotes the digamma function.

---

<sup>6</sup>The reader may notice certain equations differ from those in the original robust projections paper [29]. As is evident from the derivation, there appears to be some typographical errors in the original paper, most notably in the definition of the covariance matrix update equation (Equation 37 in [29])

Subsequently, the update equations for the parameters in the M-step are given by

$$\begin{aligned} \Sigma' = \frac{1}{N} \sum \left[ \bar{u}_n (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^\top - \bar{u}_n (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{W} \bar{\mathbf{z}}_n)^\top \right. \\ \left. - \bar{u}_n (\mathbf{W} \bar{\mathbf{z}}_n)(\mathbf{x}_n - \boldsymbol{\mu})^\top + \bar{u}_n \mathbf{W} \bar{\mathbf{S}}_n \mathbf{W}^\top \right] \end{aligned} \quad (24)$$

$$\mathbf{W}' = \frac{\sum \bar{u}_n (\mathbf{x}_n - \boldsymbol{\mu}) \bar{\mathbf{z}}_n^\top}{\sum \bar{\mathbf{S}}_n} \quad (25)$$

$$\boldsymbol{\mu}' = \frac{\sum \bar{u}_n (\mathbf{x}_n - \mathbf{W} \bar{\mathbf{z}}_n)}{\sum \bar{u}_n} \quad (26)$$

$$0 = 1 + \ln \left( \frac{\nu'}{2} \right) - 2 \psi \left( \frac{\nu'}{2} \right) + \frac{1}{N} \sum [\ln(\tilde{u}_n) - \bar{u}_n] \quad (27)$$

where  $(\cdot)'$  denotes an updated parameter and  $\nu'$  can be found by solving Equation 27 through a suitable line search. The convergence of the EM algorithm was monitored using the Q-function, derived using standard theory [28].

It is important to note that despite the presence of heavier tails, in the case of no outliers, it is expected that the Students' t-distribution will tend towards the PCCA maximum likelihood result.

#### 4.2. Robust Probabilistic SSI

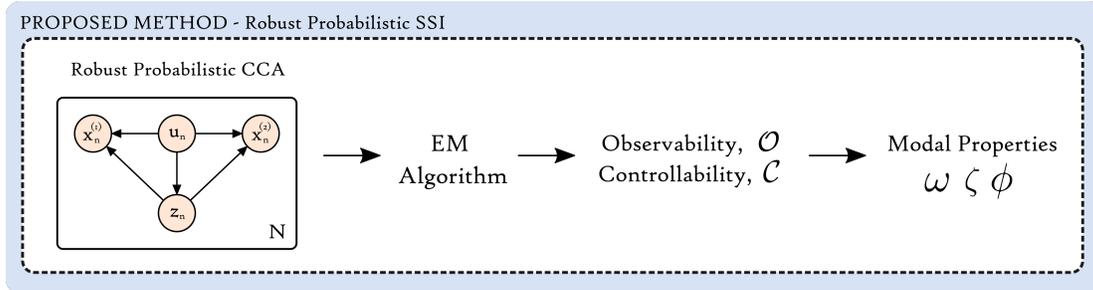


Figure 6: Schematic of the robust probabilistic SSI procedure, highlighting the replacement of PCCA (see Figure 4) with robust PCCA.

After successful convergence of the EM algorithm, final estimates for the weights can be obtained. These weights, as described by the model, are considered to be the statistically robust transformations from the latent space to the observation space. The commonality between the weights recovered by PCCA and robust PCCA, based on their ability to

transform the data to the relevant subspace, grant a similar replacement of CCA in the SSI procedure, now for robust PCCA.

$$\mathbf{W}^{(1)} = \mathcal{O} \tag{28}$$

$$\mathbf{W}^{(2)} = \mathcal{C}^\top \tag{29}$$

As noted by Archambeau et. al., the weights recovered through EM do not account for the rotational ambiguity seen previously in Section 3.1.1. Although often unnecessary<sup>7</sup>, the arbitrary rotation matrix  $\mathbf{R}$  can be recovered through a simple post-processing step, described in Appendix B. In the specific case of robust Prob-SSI, recovery of the rotation matrix is essential for formulating the stabilisation diagram in the usual way.

---

<sup>7</sup>This arbitrary rotation is often ignored as its omission still results in data being transformed into the relevant subspace.

## 5. Results and Discussion

With a mathematical framework for conducting system identification in a statistically robust and probabilistic way now established, attention can be directed towards robust identification of modal parameters using experimental data. Three separate case studies are used to demonstrate, evaluate and compare the identification performance of robust Prob-SSI to Cov-SSI. The first study (Section 5.1) uses data from a simulated linear MDOF to benchmark the general performance of robust Prob-SSI to standard Cov-SSI under ideal conditions. The variance exhibited by both methods due to different random input forcing is also briefly explored. The second case study (Section 5.2) exploits the same simulated MDOF system but now containing artificially introduced outliers. This study is used to test the robustness of identification to outliers under known conditions. The third and final case study (Section 5.3) uses data collected from the Z24 bridge benchmark. This study is used to investigate the method’s overall appropriateness to “real-world” data and assess and evaluate its overall performance against Cov-SSI.

In all tests, robust Prob-SSI was directly compared to Cov-SSI to highlight differences in identification performance. The reader will notice the omission of results from Prob-SSI. As has been shown, the MLE estimates obtained through Prob-SSI are equivalent to those found through Cov-SSI and so their addition was redundant.

Computational efficiency was not a primary factor of this work as it was deemed to be a distraction from the hypothesis being proposed. However, due to the iterative nature of the EM algorithm and the variability in convergence, robust Prob-SSI at present displays typically longer computation times, when compared to the SVD based Cov-SSI. Nevertheless, through effective computational optimisation, it is believed this could be greatly improved.

### 5.1. Benchmark: Simulated MDOF Linear System

To evaluate the identification performance of robust Prob-SSI, a suitable benchmark was created. Response data was generated using a generic three degree-of-freedom (DOF) linear dynamic system with proportional damping; shown in Figure 7 and described by the model parameters in Equation (30). The system was subsequently excited using broadband forcing, of order  $10^{-2}$  with a standard deviation of 1, mimicking ambient excitation. The system was simulated at a sample rate of  $1 \times 10^3$  Hz and generated 8192 samples. The Hankel matrix was constructed using 20 lags; i.e. 10 forwards and 10 backwards in time.

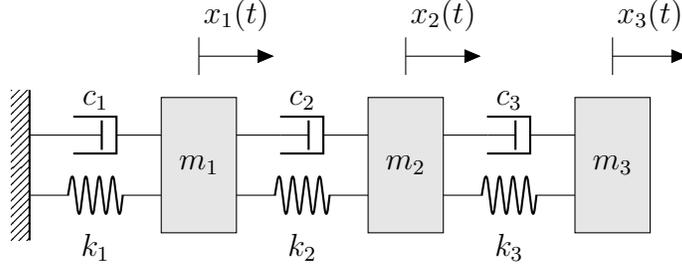


Figure 7: Simulated MDOF System

$$\mathbf{M} = \begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} 2k & -k & 0 \\ -k & 4k & -0.5k \\ 0 & -0.5k & k \end{bmatrix}, \quad \mathbf{C} = \mathbf{K} \times 10^{-4} \quad (30)$$

$$k = 1 \times 10^4 \text{ (N/m)}, \quad m = 10 \text{ (kg)}$$

Before application of robust Prob-SSI, suitable initial conditions are required for the EM algorithm. A small perturbation to the CCA (PCCA maximum likelihood) result was chosen as the initial estimate of the full weight matrix, providing a sensible prior that, one would expect, converges to a solution sooner and therefore reduces computational expense. The initial estimates for the covariance matrices were randomly sampled from an inverse Wishart distribution  $\mathcal{W}^{-1}(\mathbf{K}, \nu)$  where  $\nu = D + 2$  and  $\mathbf{K} = \mathbb{I}_D$ .

Following analysis of the data, consistency diagrams<sup>8</sup> were generated for both methods using standard techniques. These are shown in Figure 8. To ensure fair comparison, fixed definitions for the consistency criteria were applied to both methods. This ensured that any discrepancies in identified consistency arose solely from the method and not through independent changes to specific criteria. The chosen criteria were: a 2% relative change in frequency, a 5% absolute change in damping ratio, and a 98% relative correspondence in the MAC value.

---

<sup>8</sup>This often referred to as the stabilisation diagram, however the authors believe this choice of terminology may cause confusion with the ‘stability’ of poles referenced in control theory. Hence, the term consistency diagram is used. This transition has also been adopted by others in the field of dynamics [30]

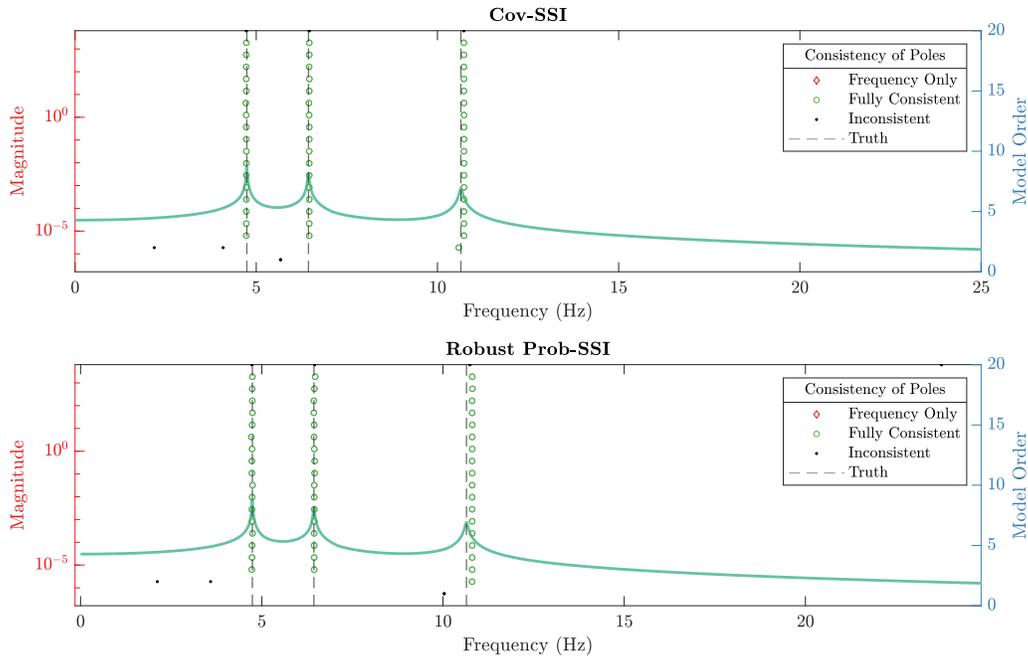


Figure 8: Consistency diagrams for the 3DOF system, recovered using Cov-SSI (top) and robust Prob-SSI (bottom).

As evident from Figure 8, the found consistency diagram for Robust Prob-SSI is very comparable with that of Cov-SSI. Clear columns of fully consistent poles are easily identifiable and, as shown in Table 1, the estimated values for the natural frequencies and damping ratios comfortably lie within an acceptable tolerance of the ground truth used in the simulation.

Table 1: A comparison of the natural frequencies and damping ratios obtained in the benchmarking study.

	Natural Frequency (Hz)			Damping Ratio		
	1	2	3	1	2	3
Truth	4.74	6.44	10.65	0.0033	0.0020	0.0015
Cov-SSI	4.73	6.46	10.73	0.0024	0.0050	0.0128
Robust Prob-SSI	4.74	6.45	10.80	0.0030	0.0047	0.0100

Due to stochastic variation in the excitation and, in the case of Robust Prob-SSI, the initialisation of the EM algorithm, the inherent variance shown by both methods was also explored. Assuming a correct number of DOF, both methods were applied to 100 datasets whereby the underlying system dynamics remain unchanged but the random seed needed to

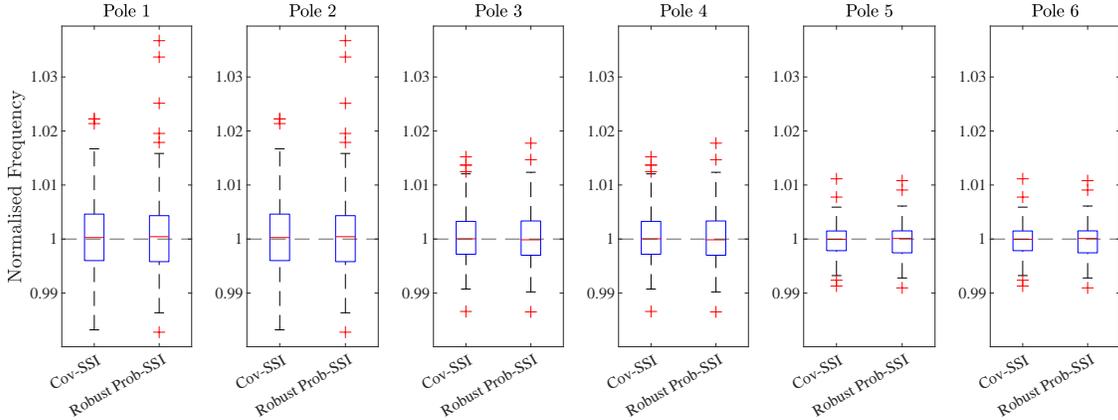


Figure 9: A demonstration of the variance seen in the natural frequency estimates using Cov-SSI and robust Prob-SSI for 100 tests on the same 3DOF dynamic system but with varying input forcing.

generate the random forcing varied between tests. The estimates of the natural frequencies from these tests are shown in Figure 9.

As expected, due to the complex nature of the system poles, the recovered estimates exist in conjugate pairs. This is evident from the clear matching of results seen in both tests between poles 1 and 2, poles 3 and 4, and poles 5 and 6. The variance exhibited by robust Prob-SSI is highly similar to that of Cov-SSI for all the recovered poles, when the correct number of degrees of freedom is assumed. Overall, this result demonstrates that robust Prob-SSI is capable of accurately replicating the results of Cov-SSI.

## 5.2. Corrupted Dataset: Simulated MDOF Linear System

Having established that robust Prob-SSI is capable of replicating the results of Cov-SSI for a simulated, ‘clean’ case, we now turn our attention to the proposed robustness of this method to corrupted datasets. In the context of this work, a ‘corrupted’ dataset will refer to any dataset containing either artificially generated or naturally occurring, atypical observations (outliers<sup>9</sup>). Furthermore, the terms ‘atypical observations’ and ‘outliers’ will be used interchangeably.

The corrupted dataset used in this study is based on the same linear dynamic MDOF

<sup>9</sup>An important detail worth noting is the definition of outliers. Archambeau et.al. assume that an outlier in the latent space will manifest as an outlier in the dataspace. This is a relatively sensible assumption given the linear transformation to the data. However, recall that Cov-SSI relies on components of the cross-covariance matrix. Therefore, if the cross-covariance is insensitive to atypical observations present in the timeseries, there will be no discernible effect on the identification procedure. In the simulated case, this can be explored, but for real datasets, the ‘clean’ case is unobtainable.

system used in the benchmarking case but now with the inclusion of artificial outliers of a chosen type and dispersion. In the example shown here, outliers were introduced at random locations in each sensor channel and set to a specified value, plus some small amount of noise. This was intended to mimic random sensor dropout where the measured value is pinned to the lower supply rail of the DAQ unit with some noise. As one might expect, there are many different types and patterns of outliers that could manifest in experimental data and it would be extremely difficult to generate and analyse all such cases. Nevertheless, for completeness, a selection of corrupted datasets with varying outlier types were also generated and analysed. The results of these series of tests are presented in Appendix D.

In the case shown, the number of outliers was set to 0.1% per channel which, for a dataset with 3 sensor channels and 8192 data points, equated to approximately 24 outliers in total. The response data and outliers are shown in Figure 10. To ensure a fair comparison of the two approaches, definitions for the consistency criteria were once again fixed. The consistency diagrams recovered from this scenario are shown in Figure 11.

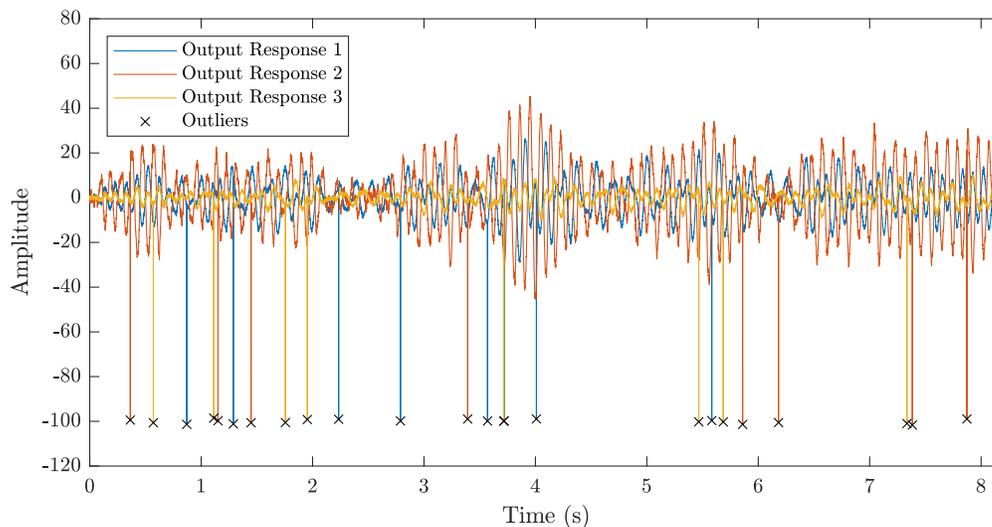


Figure 10: Response data from a simulated 3DOF ‘corrupted’ dataset, containing 0.1% artificially introduced, random outliers in each signal channel. The value of the outliers was set to a specified value plus some small amount of noise, intended to mimic random sensor dropout where the measured value is pinned to the lower supply rail of the DAQ unit with some noise

It is immediately apparent that robust Prob-SSI displays a significant improvement in identification performance over Cov-SSI when confronted with the corrupted dataset. Whilst robust Prob-SSI comfortably identifies all three correct natural frequencies of the system, traditional Cov-SSI fails to find the first, struggles to accurately identify the second and only successfully identifies the third natural frequency. This failure in identification

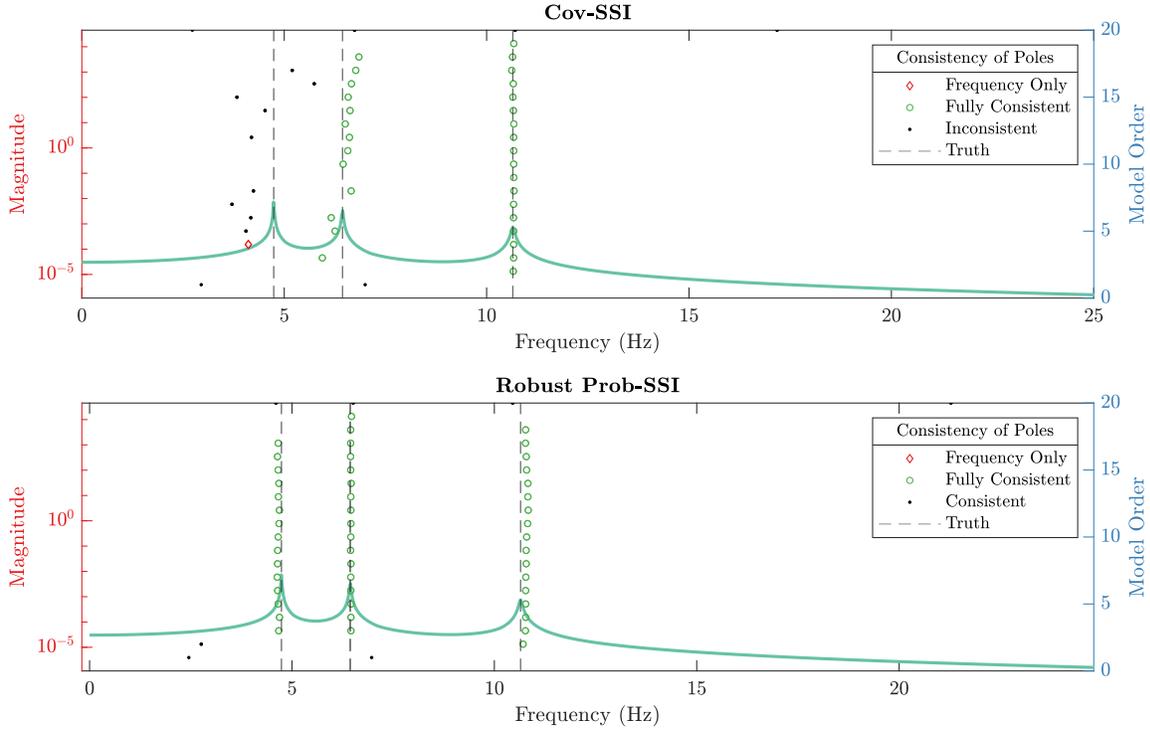


Figure 11: Consistency diagrams recovered using Cov-SSI (top) and robust Prob-SSI (bottom) using response data from a simulated 3DOF ‘corrupted’ dataset, containing 0.1% artificially introduced, randomly located outliers (in each channel), set to a specified value, plus some small amount of noise.

can be clearly observed as the lack of a column of fully consistent poles in the upper frame of Figure 11.

However, it is expected that there will be several instances where, despite the presence of outliers, SSI will continue to perform as expected. Similarly, there will also be cases where robust Prob-SSI will also fail to identify the system. The replacement of SSI with robust Prob-SSI is not a straightforward one. In Appendix E, it is illustrated how increasing the percentage of outliers inevitably leads to a deterioration in performance for both methods. Clearly, although robust Prob-SSI offers some protection, in the presence of many outliers it will be impossible to identify the dynamics.

Similar to the benchmark case, the variance of the estimates from both tests was also explored and similarly the correct number of DOF was assumed. The variance in the resulting natural frequency estimates is shown in Figure 12. As is clearly shown, the variance in Cov-SSI is significantly greater than that of robust Prob-SSI,

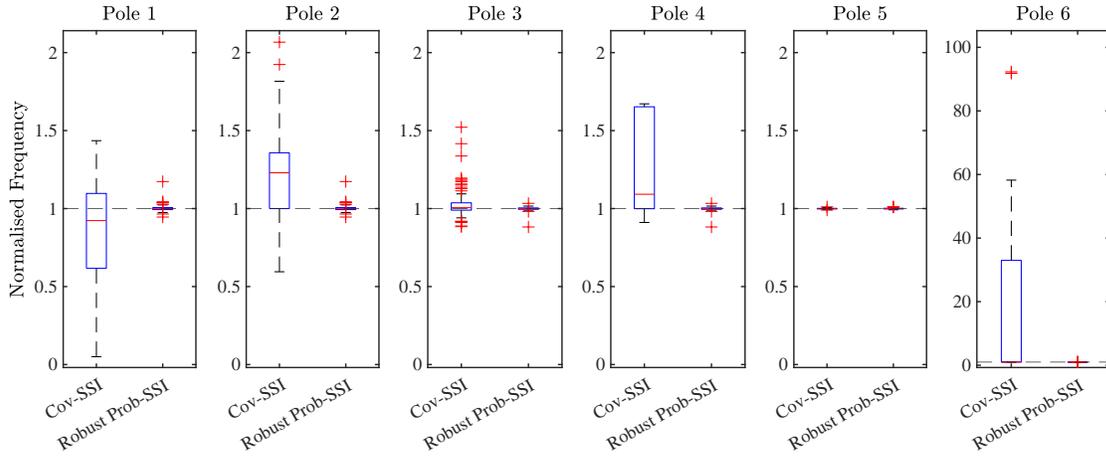


Figure 12: A demonstration of the variance seen in the natural frequency estimates when using Cov-SSI and robust Prob-SSI to analyse response data from 100 datasets of the same 3DOF dynamic system with varying input forcing but with artificially induced random outliers (0.1% per channel).

### 5.3. Case Study: Z24 Bridge

This final study is used to investigate and evaluate the performance of robust Prob-SSI when confronted with a real world example; a subset of the ever dependable Z24 bridge dataset. More information on the Z24 bridge, the available datasets and references to the original work, can be found here ([bwk.kuleuven.be/bwm/z24/z24](http://bwk.kuleuven.be/bwm/z24/z24)) [31–33]. Data collected from the Z24 bridge is often employed for new SHM tasks concerned with temporal changes to the modal properties, induced by damage and/or temperature. Its frequent application across structural dynamic research for benchmarking new system identification approaches to large scale structures, in SHM and automatic pole selection, make it a sensible choice for this study.

It is unknown if the Z24 data features any outlying observations or signal corruption. However, multiple output acceleration signals, across multiple tests in the initial stages of testing, appear to demonstrate clipping (see Figure 13). It is undetermined in previous literature if this clipping affects the identification and fundamentally this will remain unknown. However, this interesting feature of the Z24 data makes it highly suitable for our investigation into robust Prob-SSI as a robust alternative to SSI given possible atypical observations.

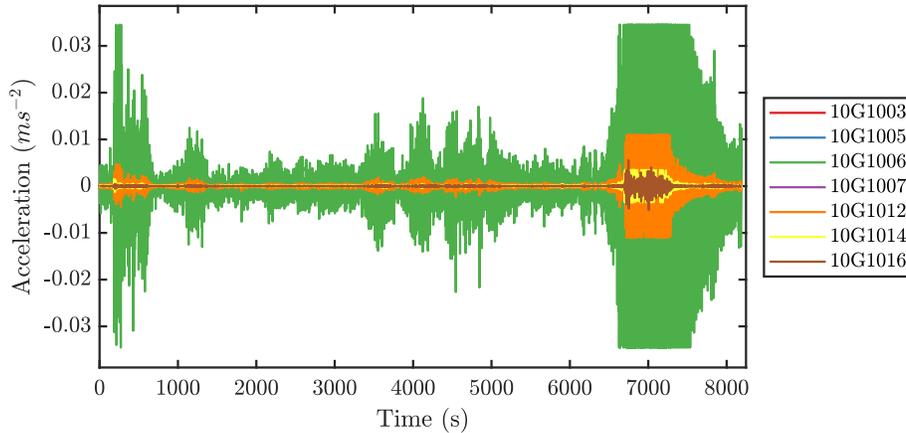


Figure 13: A sample of output time series acceleration measurements (the first 8192 points) taken from the Z24 bridge dataset 10G10 (corresponding to data recorded on ‘17-Jan-1998’ at ‘10:00:00’), that appear to demonstrate over-ranging of the sensors and therefore clipping of the measured signals.

A subset of the Z24 dataset, specifically data folder 07E01 containing acceleration measurements observed on the ‘06-Dec-1997’ at ‘10:00:00’, prior to the induced damage, was initially selected at random for processing. Only the initial segment, the first 8192 points, was used in the analysis. The resulting consistency diagrams generated using both SSI methods are shown in Figure 14. The frequency range is limited to show the first 4 natural frequencies, as identified in the original works, as these are often the main frequencies used when analysing this dataset.

The first two identified modes are easily observable in both diagrams, given the presence of highly consistent poles. The same can also be said of the third and fourth modes, with robust Prob-SSI demonstrating better consistency when compared to Cov-SSI over increasing model order. Across all the identified modes, one could argue that robust Prob-SSI identifies consistent poles at all of these modes at lower model orders than Cov-SSI and thus, would provide more confidence to the practitioner. There is some speculation in the SHM community regarding the existence of a mode at 7Hz, however, it is believed to be illusive. Interestingly, the robust approach appears to largely ignore the presence of this mode suggesting it may be the result of noise, aptly accounted for in this case by the Student’s t noise model.

This promising initial evidence suggests that robust Prob-SSI exhibits comparatively better identification performance over Cov-SSI. However, a single dataset alone is insufficient to make such a claim. To better assess the overall identification performance to a larger range of cases, Cov-SSI and robust Prob-SSI were applied to the first segment (first 8192 points) of multiple Z24 datafiles, specifically those dated 21st November 1997 14:00:00 through to 25th April 1998 14:00:00, and the resultant poles extracted.

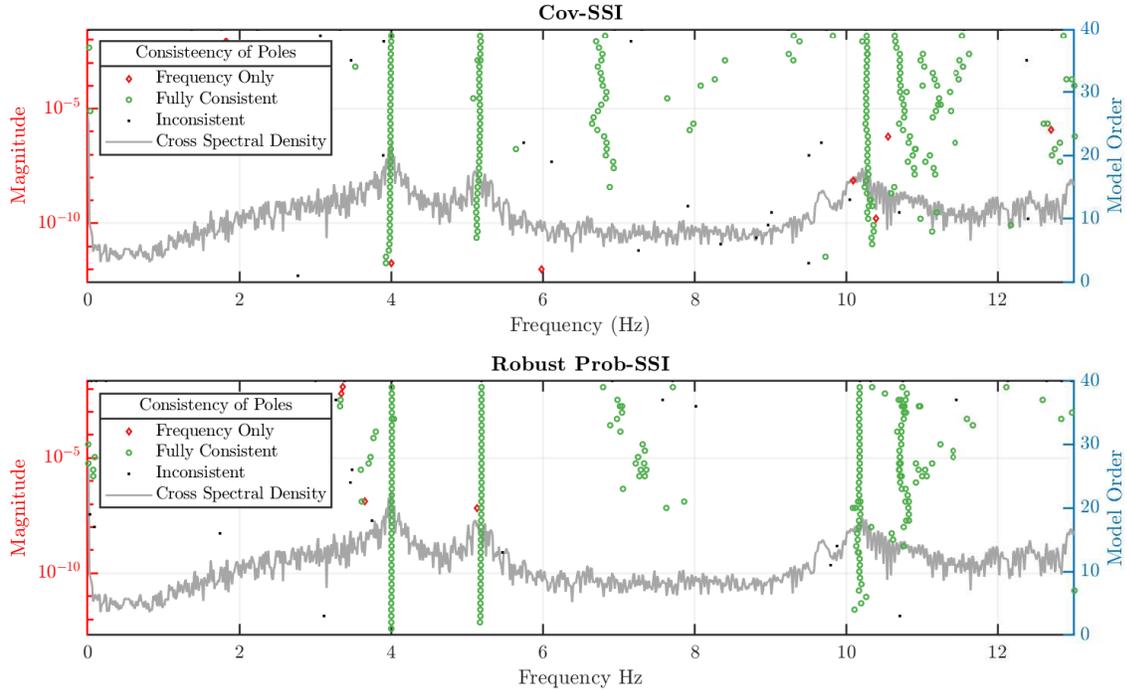


Figure 14: Consistency diagrams generated using results from Cov-SSI (top) and robust Prob-SSI (bottom) after application to a single subset of the Z24 bridge dataset (07E01); corresponding to data recorded on ‘06-Dec-1997’ at ‘01:00:00’

Rather than manually selecting consistent poles in each consistency diagram, a very time consuming activity, automatic selection of the consistent poles was achieved using an implementation of the clustering algorithm DBSCAN. The application of DBSCAN (or derivative thereof, OPTICS) to pole selection in stabilisation diagrams is not a novel one, existing work can be found here [34]. As the performance of any given clustering technique was not the focus of this study, the decision to use DBSCAN clustering over another was arbitrary, however as a very common mode of clustering and one frequently cited in literature, it may seem the most sensible. For information on the theory and implementation of DBSCAN, the reader is directed here [35].

Akin to the approach used by Boroschek and Bilbao [34], the distance metric in DBSCAN was chosen to take the following form:

$$dist(p_i, p_j) = \frac{|\omega_i - \omega_j|}{\max(\omega_i, \omega_j)} + (1 - \text{MAC}(\phi_i, \phi_j)) \quad (31)$$

where the natural frequencies  $\omega_i, \omega_j$  and mode shapes  $\phi_i, \phi_j$  correspond to poles  $p_i$  and  $p_j$  respectively, and the MAC is the modal assurance criterion. The reachability distance was

chosen to be 0.005%, with the minimum number of objects set to 25.

Following application of DBSCAN to the Z24 data, centres of the recovered pole clusters were plotted against the corresponding date and time of collection. Figure 15 displays the temporal changes to the natural frequency estimates obtained using Cov-SSI and robust Prob-SSI. The first noticeable difference between the two plots in Figure 15 is variation in the number of consistent poles situated around 7 Hz, corresponding to the supposedly illusive mode. Congruent to the assessment of the single consistency diagram, robust Prob-SSI (bottom) appears to find a lower number of consistent poles at this frequency across the full range of tests, when compared to Cov-SSI (top). Again, this is likely due to any noise generating this nuisance mode, being accounted for by the Student's  $t$  noise model of robust Prob-SSI.

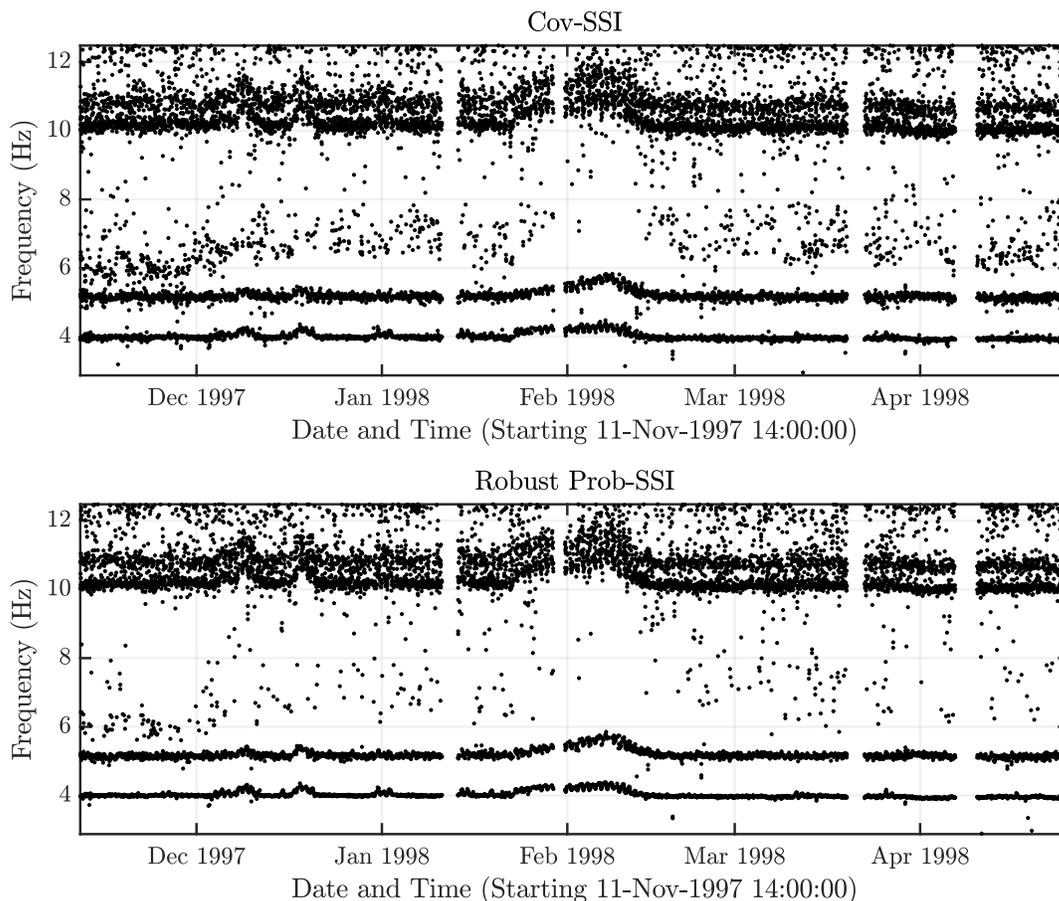


Figure 15: Temporal changes to the natural frequency estimates of the Z24 bridge, identified using Cov-SSI (top) and Robust Prob-SSI (below). DBSCAN clustering was used to extract the non-spurious, consistent poles from each individual dataset. Both methods were applied to the first segment (first 8192 points) of all datafiles ranging from 21st November 1997 14:00:00 to 25th April 1998 14:00:00.

Turning attention to the first two modes, much lower variance is observed in the frequency estimates recovered using robust Prob-SSI across the entire test series compared to Cov-SSI. This reinforces preliminary conclusions drawn from Figure 14 regarding the increased consistency of recovered poles using robust Prob-SSI. This reduction in variance is also apparent, albeit less obviously, in modes three and four, suggesting robust Prob-SSI is capable of recovering less noisy pole estimates than Cov-SSI. This has several benefits, not least in providing the experienced practitioner with more confidence when selecting suitable poles from the consistency diagram. This new technique may also prove to be a preferred method of modal parameter recovery for SHM schemes tasked with accurately monitoring temporal changes to modal properties and inferring possible damage. Reducing variance in modal property estimates is of value, in the SHM setting, as it offers a route to reducing false positives and increasing sensitivity in anomaly detection.

## 6. Conclusions

This paper presented a new formulation of the stochastic subspace identification (SSI) algorithm, recasting it as a problem in probabilistic inference. Such an approach was made possible through close alignment of SSI with the theory of probabilistic projections [8]. Mathematically, an equivalence between the maximum likelihood estimates of the weights (linear transformations) of the latent projections, and the observability and controllability matrix transposed, was established. This unique perspective of SSI lays the requisite mathematical foundation required to enable a multitude of new SSI-based OMA algorithms, likely to be constructed using more complex probabilistic and hierarchical techniques. Such algorithms could add more value to the inclusion of SSI as a component in broader probabilistic frameworks, although the detail of which not discussed here.

Based on this new probabilistic form, this paper then introduced a second contribution - a statistically robust SSI algorithm, capable of providing a principled and automatic way of handling atypical observations in multi-output time-series responses, such as those observed in field measurements. Consequently, the proposed method was evaluated against three primary case studies. The first study benchmarked robust Prob-SSI against conventional SSI using a familiar MDOF dataset and displayed highly comparable results in the case where data are not corrupted. The second study sought to evaluate the methods ability to resist misidentification when presented with outliers. Robust Prob-SSI was exposed to a 'corrupted' dataset containing artificially introduced outliers; designed to mimic a typical form of anomalous data observed in experimental testing. Subsequent analysis showed robust Prob-SSI outperform conventional SSI, with the former exhibiting better modal identification performance (consistency) and higher confidence in the found poles. The last study investigated data taken from the Z24 Bridge benchmark, chosen with the view of better demonstrating identification performance on data more indicative of a real system. Results from this analysis showed that robust Prob-SSI, enabled by the

novel probabilistic interpretation of the SSI algorithm, gives an automated and consistent approach to managing data collected in OMA settings.

This initial body work highlighted the first of many new formulations of SSI, made possible through the novel viewpoint of SSI as a problem in probabilistic inference. The authors envisage future work pursuing new algorithms, constructed upon Prob-SSI, for example, adopting a Bayesian methodology through the inclusion of priors over the model parameters or alternative hierarchical structures.

### **CRedit Authorship Contribution Statement**

**Brandon O’Connell:** Conceptualization, Methodology, Software, Validation, Visualization, Writing - Original Draft **Timothy Rogers:** Conceptualization, Software, Writing - Review & Editing, Supervision, Funding acquisition

### **Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### **Acknowledgements**

The authors gratefully acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC), UK through grant number EP/W002140/1. The authors also thank Elizabeth Cross and Ulf Tygesen for their insight and many helpful discussions. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising.

## Appendix A. Robust Probabilistic CCA Derivation

$$\langle \mathcal{L} \rangle = \sum_{n=1}^N \log p(x_n, z_n) \quad (\text{A.1})$$

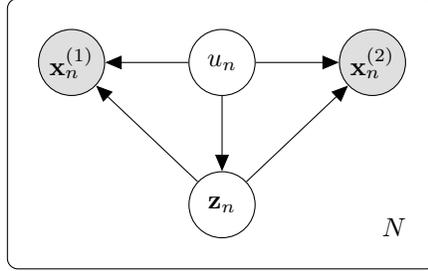


Figure A.16: Graphical latent model for Robust Canonical Correlation Analysis

$$p(u_n) = \mathcal{G}\left(u_n \mid \frac{\nu}{2}, \frac{\nu}{2}\right) \quad (\text{A.2a})$$

$$p(\mathbf{z}_n | u_n) = \mathcal{N}(\mathbf{z}_n | \mathbf{0}, u_n^{-1} \mathbb{I}_d) \quad (\text{A.2b})$$

$$p(\mathbf{x}_n^{(m)} | \mathbf{z}_n, u_n) = \mathcal{N}(\mathbf{x}_n^{(m)} | \mathbf{W}^{(m)} \mathbf{z}_n + \boldsymbol{\mu}^{(m)}, u_n^{-1} \boldsymbol{\Sigma}^{(m)}) \quad (\text{A.2c})$$

$$p(\mathbf{x}_n | \mathbf{z}_n, u_n) = \mathcal{N}(\mathbf{x}_n | \mathbf{W} \mathbf{z}_n + \boldsymbol{\mu}, u_n^{-1} \boldsymbol{\Sigma}) \quad (\text{A.2d})$$

where  $\mathbf{x}_n = [\mathbf{x}_n^{(1)}; \mathbf{x}_n^{(2)}]$ ,  $\mathbf{X}^{(m)} = [\mathbf{x}_1^{(m)}, \dots, \mathbf{x}_N^{(m)}]$  with  $m = 1, 2$ ,  $\mathbf{X} = [\mathbf{X}^{(1)}; \mathbf{X}^{(2)}]$ ,  $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N]$  and  $\boldsymbol{\Sigma}$  is a block-diagonal covariance matrix with matrices  $\boldsymbol{\Sigma}^{(1)}$  and  $\boldsymbol{\Sigma}^{(2)}$  as the diagonal elements.

In the original Robust Projections paper, the normal distributions are defined using the precision matrix, rather than covariance matrix. Note that here only covariance matrices are used.

To derive the EM update equations for the parameters and an appropriate Q-function to monitor convergence of the EM algorithm, the log-likelihood function must be defined:

$$\mathcal{L}(\boldsymbol{\theta} | \mathbf{x}_n, \mathbf{z}_n, u_n) = \sum_{n=1}^N \ln p(\mathbf{x}_n, \mathbf{z}_n, u_n | \boldsymbol{\theta}) \quad (\text{A.3})$$

where  $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\Sigma}, \nu)$ .

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = & \frac{1}{N} \sum \left[ -\frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln |u_n^{-1} \mathbb{I}_d| - \frac{1}{2} (\mathbf{z}_n - \mathbf{0})^\top (u_n^{-1} \mathbb{I}_d)^{-1} (\mathbf{z}_n - \mathbf{0}) \right. \\
& \cdots + \frac{\nu}{2} \ln \left( \frac{\nu}{2} \right) - \ln \left( \Gamma \left( \frac{\nu}{2} \right) \right) + \frac{\nu}{2} \ln(u_n) - \frac{\nu}{2} u_n - \frac{D}{2} \ln(2\pi) \\
& \left. \cdots - \frac{1}{2} \ln |u_n^{-1} \boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{x}_n - (\mathbf{W} \mathbf{z}_n + \boldsymbol{\mu}))^\top (u_n^{-1} \boldsymbol{\Sigma})^{-1} (\mathbf{x}_n - (\mathbf{W} \mathbf{z}_n + \boldsymbol{\mu})) \right]
\end{aligned} \tag{A.4}$$

Through the simplification of terms, Eq.A.4 can be rearranged and simplified to

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = & \frac{1}{N} \sum \left[ -D \ln(2\pi) - \frac{1}{2} \ln |u_n^{-1} \mathbb{I}_d| - \frac{1}{2} u_n \mathbf{z}_n^\top \mathbf{z}_n + \frac{\nu}{2} \ln \left( \frac{\nu}{2} \right) \right. \\
& \cdots - \ln \left( \Gamma \left( \frac{\nu}{2} \right) \right) + \frac{\nu}{2} \ln(u_n) - \frac{\nu}{2} u_n - \frac{1}{2} \ln |u_n^{-1} \boldsymbol{\Sigma}| \\
& \left. \cdots - \frac{1}{2} u_n ((\mathbf{x}_n - \boldsymbol{\mu}) - (\mathbf{W} \mathbf{z}_n))^\top \boldsymbol{\Sigma}^{-1} ((\mathbf{x}_n - \boldsymbol{\mu}) - (\mathbf{W} \mathbf{z}_n)) \right]
\end{aligned} \tag{A.5}$$

The log matrix identity  $\ln |aX| = D \ln(a) + \ln |X|$  can be used to expand and simplify the terms  $\ln |u_n \mathbb{I}_d|$  and  $\ln |u_n \boldsymbol{\Sigma}^{-1}|$ , giving

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) = & \frac{1}{N} \sum \left[ -D \ln(2\pi) + D \ln(u_n) - \frac{1}{2} u_n \mathbf{z}_n^\top \mathbf{z}_n + \frac{\nu}{2} \ln \left( \frac{\nu}{2} \right) \right. \\
& \cdots - \ln \left( \Gamma \left( \frac{\nu}{2} \right) \right) + \frac{\nu}{2} \ln(u_n) - \frac{\nu}{2} u_n - \frac{1}{2} \ln |\boldsymbol{\Sigma}| \\
& \cdots - \frac{1}{2} u_n (\mathbf{x}_n - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}) + \frac{1}{2} u_n (\mathbf{x}_n - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{W} \mathbf{z}_n) \\
& \left. \cdots + \frac{1}{2} u_n (\mathbf{W} \mathbf{z}_n)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}) - \frac{1}{2} u_n (\mathbf{W} \mathbf{z}_n)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{W} \mathbf{z}_n) \right]
\end{aligned} \tag{A.6}$$

### *Q FUNCTION DERIVATION*

Given the definition of the log likelihood in Equation A.6, the Q-function and subsequently the EM update equations, can be derived.

$$\mathcal{Q} = \mathbb{E}_{\mathbf{z}, u | \mathbf{x}, \boldsymbol{\theta}} [\mathcal{L}(\boldsymbol{\theta})] = \iint \mathcal{L}(\boldsymbol{\theta}) p(\mathbf{z}, u) \, d\mathbf{z} du \tag{A.7}$$

$$\begin{aligned}
\mathcal{Q} = & \frac{1}{N} \sum \left[ -D \ln(2\pi) + D \mathbb{E} [\ln(u_n)] - \frac{1}{2} \mathbb{E} [u_n] \mathbb{E} [\mathbf{z}_n^\top \mathbf{z}_n] + \frac{\nu}{2} \ln \left( \frac{\nu}{2} \right) \right. \\
& \dots - \ln \left( \Gamma \left( \frac{\nu}{2} \right) \right) + \frac{\nu}{2} \mathbb{E} [\ln(u_n)] - \frac{\nu}{2} \mathbb{E} [u_n] - \frac{1}{2} \ln |\boldsymbol{\Sigma}| \\
& \dots - \frac{1}{2} \mathbb{E} [u_n] (\mathbf{x}_n - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}) + \frac{1}{2} \mathbb{E} [u_n] (\mathbf{x}_n - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{W} \mathbb{E} [\mathbf{z}_n]) \\
& \left. \dots + \frac{1}{2} \mathbb{E} [u_n] (\mathbf{W} \mathbb{E} [\mathbf{z}_n])^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}) - \frac{1}{2} \mathbb{E} [u_n (\mathbf{W} \mathbf{z}_n)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{W} \mathbf{z}_n)] \right]
\end{aligned} \tag{A.8}$$

Let the expectations of the latent parameters be denoted by

$$\mathbb{E} [u_n] = \bar{u}_n \tag{A.9a}$$

$$\mathbb{E} [\ln(u_n)] = \ln \tilde{u}_n \tag{A.9b}$$

$$\mathbb{E} [\mathbf{z}_n] = \bar{\mathbf{z}}_n \tag{A.9c}$$

$$\mathbb{E} [u_n \mathbf{z}_n \mathbf{z}_n^\top] = \bar{\mathbf{S}}_n \tag{A.9d}$$

Using the conditional distributions of the variables, exact solutions for the expectations can be recovered.

$$p(\mathbf{x}_n | u_n) = \int p(\mathbf{x}_n | \mathbf{z}_n, u_n) p(\mathbf{z}_n | u_n) d\mathbf{z} \tag{A.10}$$

$$= \int \mathcal{N}(\mathbf{W} \mathbf{z}_n + \boldsymbol{\mu}, u_n^{-1} \boldsymbol{\Sigma}) \mathcal{N}(\mathbf{0}, u_n^{-1} \mathbb{I}_d) d\mathbf{z} \tag{A.11}$$

$$= \mathcal{N}(\boldsymbol{\mu}, u_n^{-1} (\boldsymbol{\Sigma} + \mathbf{W} \mathbf{W}^\top)) \tag{A.12}$$

where  $\mathbf{A} = (\boldsymbol{\Sigma} + \mathbf{W} \mathbf{W}^\top)$ , and

$$p(u_n | \mathbf{x}_n) \propto p(\mathbf{x}_n | u_n) p(u_n) \tag{A.13}$$

$$p(u_n | \mathbf{x}_n) = \mathcal{G} \left( u_n \left| \frac{D + \nu}{2}, \frac{u_n (\mathbf{x}_n - \boldsymbol{\mu})^\top \mathbf{A}^{-1} (\mathbf{x}_n - \boldsymbol{\mu})}{2} + \frac{\nu}{2} \right. \right) \tag{A.14}$$

Using the conditional distributions, values for the expectation of these variables, with respect to the other variables.

$$\bar{u}_n = \frac{\alpha}{\beta} = \frac{D + \nu}{u_n(\mathbf{x}_n - \boldsymbol{\mu})^\top \mathbf{A}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) + \nu} \quad (\text{A.15})$$

is found using the equation for the mean of a Gamma distribution, and

$$\ln \tilde{u}_n = \psi \left( \frac{D + \nu}{2} \right) - \ln \left( \frac{u_n(\mathbf{x}_n - \boldsymbol{\mu})^\top \mathbf{A}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) + \nu}{2} \right) \quad (\text{A.16})$$

is the expectation of the log Gamma distribution.

$$\bar{\mathbf{z}}_n = \mathbf{B}^{-1} \mathbf{W}^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}) \quad (\text{A.17})$$

where  $\mathbf{B} = \mathbf{W}^\top \boldsymbol{\Sigma}^{-1} \mathbf{W} + \mathbb{I}_d$

$$\bar{\mathbf{S}}_n = \mathbf{B}^{-1} + \mathbb{E}[u_n] \mathbb{E}[\mathbf{z}_n] \mathbb{E}[\mathbf{z}_n^\top] \quad (\text{A.18})$$

$$\bar{\mathbf{S}}_n = \mathbf{B}^{-1} + \bar{u}_n \bar{\mathbf{z}}_n \bar{\mathbf{z}}_n^\top \quad (\text{A.19})$$

Expectation-maximisation update equations can be analytically found from the Q-function using

$$\boldsymbol{\theta}^{\text{new}} = \arg \max_{\boldsymbol{\theta}} (Q(\boldsymbol{\theta}^{\text{old}}, \mathbf{X})) \quad (\text{A.20})$$

The EM update equations can be recovered by taking the derivative of the Q-function, with respect to the various parameters  $\boldsymbol{\theta}$ , and setting the resulting equations to zero and solving for the parameter in question.

### *Covariance Update*

$$\begin{aligned} \frac{\partial Q}{\partial \boldsymbol{\Sigma}^{-1}} = \frac{1}{N} \sum \left[ \frac{1}{2} \boldsymbol{\Sigma}^\top - \frac{1}{2} \bar{u}_n (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^\top + \frac{1}{2} \bar{u}_n (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{W} \bar{\mathbf{z}}_n)^\top \right. \\ \left. \dots + \frac{1}{2} \bar{u}_n (\mathbf{W} \bar{\mathbf{z}}_n)(\mathbf{x}_n - \boldsymbol{\mu})^\top + \frac{1}{2} \bar{u}_n \mathbf{W} \bar{\mathbf{S}}_n \mathbf{W}^\top \right] \end{aligned} \quad (\text{A.21})$$

setting  $\frac{\partial \mathcal{Q}}{\partial \Sigma^{-1}} = 0$  and rearranging, note that due to the symmetry of  $\Sigma$ ,  $\Sigma^\top = \Sigma$ , the following update equation for the covariance can be defined.

$$\begin{aligned} \Sigma = \frac{1}{N} \sum \left[ \bar{u}_n (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^\top - \bar{u}_n (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{W} \bar{\mathbf{z}}_n)^\top \right. \\ \left. \dots - \bar{u}_n (\mathbf{W} \bar{\mathbf{z}}_n)(\mathbf{x}_n - \boldsymbol{\mu})^\top + \bar{u}_n \mathbf{W} \bar{\mathbf{S}}_n \mathbf{W}^\top \right] \end{aligned} \quad (\text{A.22})$$

*Weight Update*

$$\frac{\partial \mathcal{Q}}{\partial \mathbf{W}} = \frac{1}{N} \sum \left[ \bar{u}_n \Sigma^{-1} (\mathbf{x}_n - \boldsymbol{\mu}) \bar{\mathbf{z}}_n^\top - \Sigma^{-1} \mathbf{W} \bar{\mathbf{S}}_n \right] \quad (\text{A.23})$$

letting  $\frac{\partial \mathcal{Q}}{\partial \mathbf{W}} = 0$  and rearranging,

$$\mathbf{W} = \left( \sum^N \bar{u}_n (\mathbf{x}_n - \boldsymbol{\mu}) \bar{\mathbf{z}}_n^\top \right) \left( \sum^N \bar{\mathbf{S}}_n \right)^{-1} \quad (\text{A.24})$$

*Mean Update*

$$\frac{\partial \mathcal{Q}}{\partial \boldsymbol{\mu}} = \frac{1}{N} \sum \left[ \bar{u}_n \Sigma^{-1} (\mathbf{x}_n - \boldsymbol{\mu}) - \bar{u}_n \Sigma^{-1} \mathbf{W} \bar{\mathbf{z}}_n \right] \quad (\text{A.25})$$

letting  $\frac{\partial \mathcal{Q}}{\partial \boldsymbol{\mu}} = 0$  and expanding,

$$0 = \frac{1}{N} \sum \left[ \bar{u}_n \Sigma^{-1} \mathbf{x}_n - \bar{u}_n \Sigma^{-1} \boldsymbol{\mu} - \bar{u}_n \Sigma^{-1} \mathbf{W} \bar{\mathbf{z}}_n \right] \quad (\text{A.26})$$

$$\sum^N \left[ \bar{u}_n \Sigma^{-1} \boldsymbol{\mu} \right] = \sum^N \left[ \bar{u}_n \Sigma^{-1} \mathbf{x}_n - \bar{u}_n \Sigma^{-1} \mathbf{W} \bar{\mathbf{z}}_n \right] \quad (\text{A.27})$$

premultiplying by  $\Sigma$  and rearranging for  $\boldsymbol{\mu}$  gives

$$\sum^N \left[ \bar{u}_n \boldsymbol{\mu} \right] = \sum^N \left[ \bar{u}_n (\mathbf{x}_n - \mathbf{W} \bar{\mathbf{z}}_n) \right] \quad (\text{A.28})$$

$$\boldsymbol{\mu} = \frac{\sum^N \bar{u}_n (\mathbf{x}_n - \mathbf{W} \bar{\mathbf{z}}_n)}{\sum^N \bar{u}_n} \quad (\text{A.29})$$

$\nu$  Update

$$\frac{\partial \mathcal{Q}}{\partial \nu} = \frac{1}{N} \sum^N \left[ \frac{1}{2} \ln \left( \frac{\nu}{2} \right) + \frac{1}{2} \cdot \frac{\nu}{2} \cdot \frac{2}{\nu} - \psi \left( \frac{\nu}{2} \right) + \frac{1}{2} \ln(\tilde{u}_n) - \frac{1}{2} \bar{u}_n \right] \quad (\text{A.30})$$

$$\frac{\partial \mathcal{Q}}{\partial \nu} = \frac{1}{N} \sum^N \left[ \frac{1}{2} + \frac{1}{2} \ln \left( \frac{\nu}{2} \right) - \psi \left( \frac{\nu}{2} \right) + \frac{1}{2} \ln(\tilde{u}_n) - \frac{1}{2} \bar{u}_n \right] \quad (\text{A.31})$$

letting  $\frac{\partial \mathcal{Q}}{\partial \nu} = 0$ , multiplying by 2 and rearranging the sum,

$$0 = 1 + \ln \left( \frac{\nu}{2} \right) - 2 \psi \left( \frac{\nu}{2} \right) + \frac{1}{N} \sum^N [\ln(\tilde{u}_n) - \bar{u}_n] \quad (\text{A.32})$$

The maximum likelihood of  $\nu$  can be found through solving the above equation using line search.

## Appendix B. Recovery of Rotation Matrix $\mathbf{R}$

The following theory for the recovery of the rotation matrix in Robust PCCA summarises the work presented by Archambeau et.al. in Appendix A of [29] and the accompanying errata of Appendix A. for the same paper.

Having defined

$$\mathbf{B}_1 = \check{\mathbf{W}}^{(1)\top} \boldsymbol{\Sigma}^{(1)-1} \check{\mathbf{W}}^{(1)} + \mathbb{I}_d \quad (\text{B.1})$$

$$\mathbf{B}_2 = \check{\mathbf{W}}^{(2)\top} \boldsymbol{\Sigma}^{(2)-1} \check{\mathbf{W}}^{(2)} + \mathbb{I}_d \quad (\text{B.2})$$

The matrix  $\mathbf{R}_1$  contains the eigenvectors of

$$\mathbf{J}_1 = (\mathbb{I}_d - \mathbf{B}_1^{-1})^{\frac{1}{2}} (\mathbb{I}_d - \mathbf{B}_2^{-1}) (\mathbb{I}_d - \mathbf{B}_1^{-1})^{\frac{1}{2}} \quad (\text{B.3})$$

with  $\tilde{\Upsilon}^2$  corresponding eigenvalues. Similarly  $\mathbf{R}_2$  contains the eigenvectors of

$$\mathbf{J}_2 = (\mathbb{I}_d - \mathbf{B}_2^{-1})^{\frac{1}{2}} (\mathbb{I}_d - \mathbf{B}_1^{-1}) (\mathbb{I}_d - \mathbf{B}_2^{-1})^{\frac{1}{2}} \quad (\text{B.4})$$

with the same eigenvalues  $\tilde{\Upsilon}^2$ .

Given the above, it can finally be shown that the canonical directions are

$$\mathbf{U}_d^{(1)} = \boldsymbol{\Sigma}_{11}^{-1} \check{\mathbf{W}}^{(1)} (\mathbb{I}_d - \mathbf{B}_1^{-1})^{-\frac{1}{2}} \mathbf{R}_1 \quad (\text{B.5})$$

$$\mathbf{U}_d^{(2)} = \boldsymbol{\Sigma}_{22}^{-1} \check{\mathbf{W}}^{(2)} (\mathbb{I}_d - \mathbf{B}_2^{-1})^{-\frac{1}{2}} \mathbf{R}_2 \quad (\text{B.6})$$

### Appendix C. Recovery of the modal properties from the state and output matrices.

Given a state matrix  $\mathbf{A}$ , the following eigenvalue problem

$$\mathbf{A}\boldsymbol{\psi} = \lambda\boldsymbol{\psi} \quad (\text{C.1})$$

can be solved to recover a set of eigenvalues  $\lambda$  and corresponding eigenvectors  $\boldsymbol{\psi}$ . The respective eigenvalues correspond to a set of discrete time system poles. The continuous system poles can be recovered easily using

$$\mu = \log(\Lambda) \quad (\text{C.2})$$

where  $\Lambda$  is a diagonal matrix containing eigenvalues  $\lambda$ . Once recovered, the poles can be used to trivially calculate the natural frequencies in  $\text{rad } s^{-1}$

$$\omega_n = \frac{|\mu|}{\Delta t} \quad (\text{rads}^{-1}) \quad (\text{C.3})$$

where  $\Delta t$  is the sample time, and the damping ratios can be found using

$$\zeta_n = -\text{Re}\{\mu\} \oslash |\mu| \quad (\text{C.4})$$

where  $\oslash$  denotes a Hadamard (elementwise) division.

Finally, the mode shape matrix  $\Phi$  can also be recovered using the found output matrix  $\mathbf{C}$  and the eigenvectors  $\boldsymbol{\Psi}$ , such that

$$\Phi = \text{Re}\{\mathbf{C}\boldsymbol{\Psi}\} \quad (\text{C.5})$$

## Appendix D. Varying Outlier Type Case Studies

Employing the same underlying linear dynamic system used in the outlier study shown in Section 5.2, a range of alternative outlier cases were explored to observe the effects of varying outlier type on identification performance across both methods.

### *Appendix D.1. Periodic Block Dropout to Electrical Noise Floor*

The outliers in this example are designed to mimic periodic dropout of a single sensor channel to a supposed ‘electrical noise floor’, repeating in regular blocks with a duration of 0.01 seconds. Output response data and outliers are shown in Figure D.1, whilst the consistency diagrams for Cov-SSI and robust Prob-SSI are shown in Figure D.2.

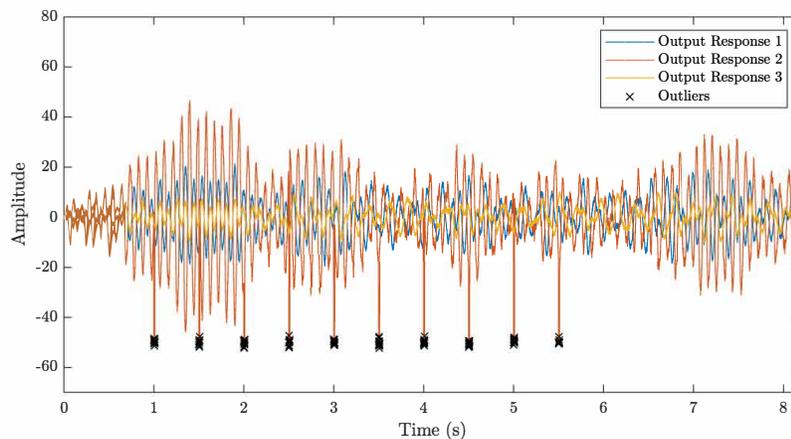


Figure D.1: Response data from a simulated 3DOF ‘corrupted’ dataset, containing 0.1% of artificially introduced, periodic blocks of outliers in one channel. The value of the outliers was set to a specified value, plus some small amount of noise.

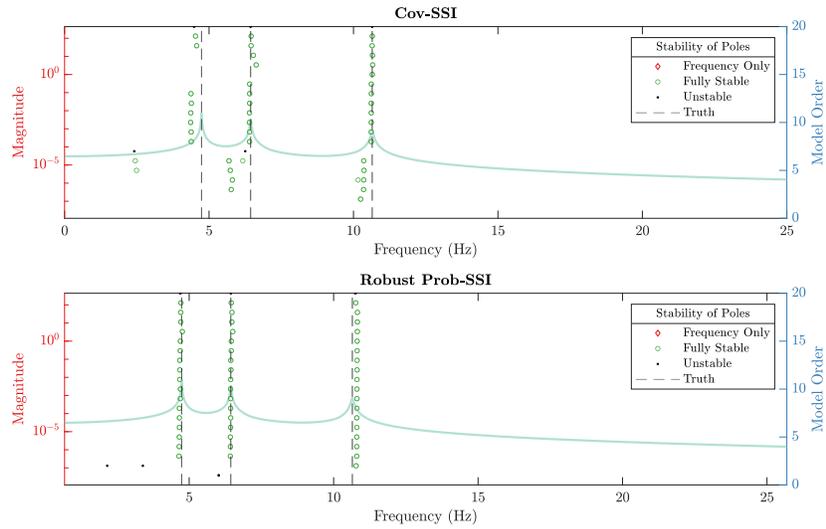


Figure D.2: Consistency diagrams recovered using Cov-SSI (top) and robust Prob-SSI (bottom) for the periodic block dropout case.

## Appendix D.2. Clipping

The outliers in this example are designed to mimic the clipping of all sensor channels to 80% of the maximum amplitude of the individual signals. Output response data and the outliers are shown in Figure D.3, whilst the consistency diagrams for Cov-SSI and robust Prob-SSI are shown in Figure D.4.

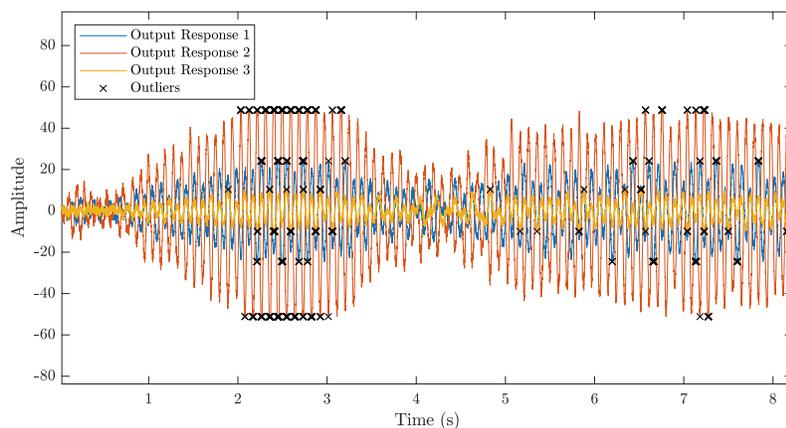


Figure D.3: Response data from a simulated 3DOF ‘corrupted’ dataset, where all channels were clipped to 80% of the maximum value of each channel.

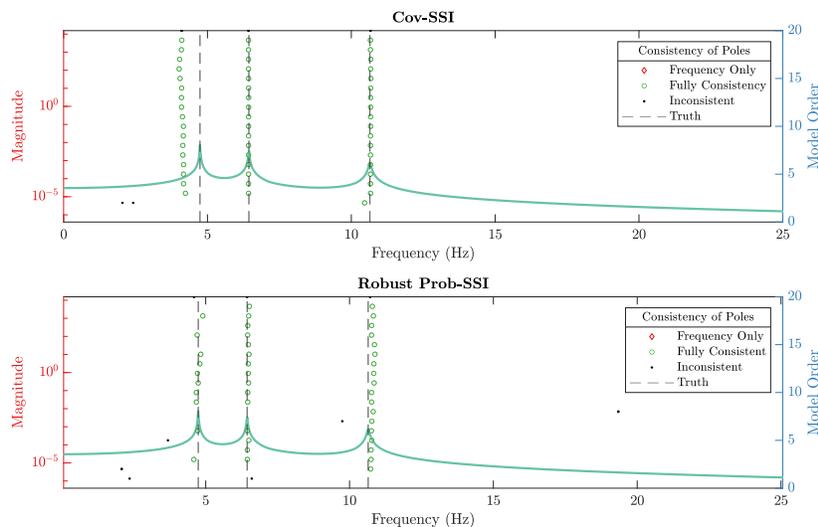


Figure D.4: Consistency diagrams recovered using Cov-SSI (top) and robust Prob-SSI (bottom) using response data from a simulated 3DOF ‘corrupted’ dataset, where all channels are clipped to 80% of the maximum value of each channel.

### Appendix D.3. Block Dropout to Zero

The outliers in this example are designed to mimic a block dropout of a single sensor channel to zero. The dropout starts at 3 seconds and has a duration of 1 second. Output response data and outliers are shown in Figure D.5, whilst the consistency diagrams for Cov-SSI and robust Prob-SSI are shown in Figure D.6.

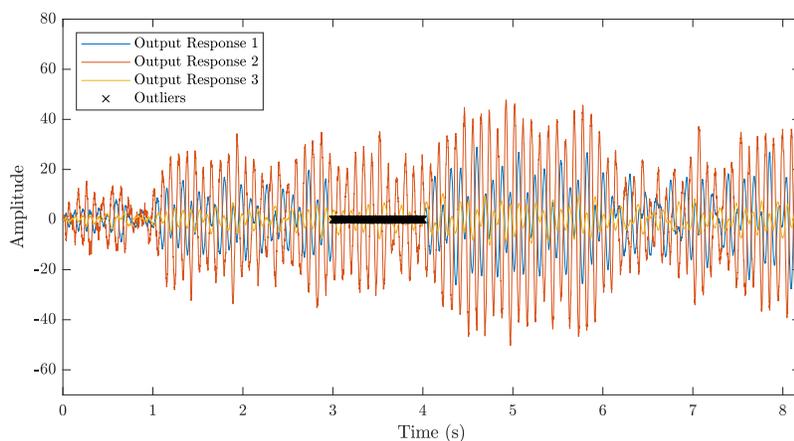


Figure D.5: Response data from a simulated 3DOF ‘corrupted’ dataset, where values in a single channel were set to zero amplitude for a 1000 point block in the 8192 point long data series.

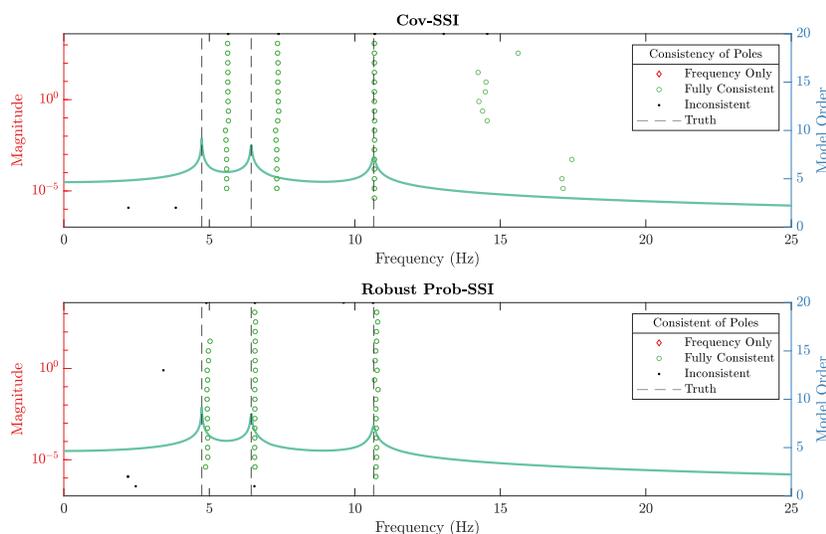


Figure D.6: Consistency diagrams recovered using Cov-SSI (top) and robust Prob-SSI (bottom) using response data from a simulated 3DOF ‘corrupted’ dataset, where values in a single channel were set to zero amplitude for a 1000 point block in the 8192 point long data series.

## Appendix E. Percentage Outlier Study

Employing the same parameters used in the outlier study shown in Section 5.2, the effect of changing the percentage of outliers was explored to observe the effects on identification performance across both methods.

### Appendix E.1. Random Outliers - 0.5%

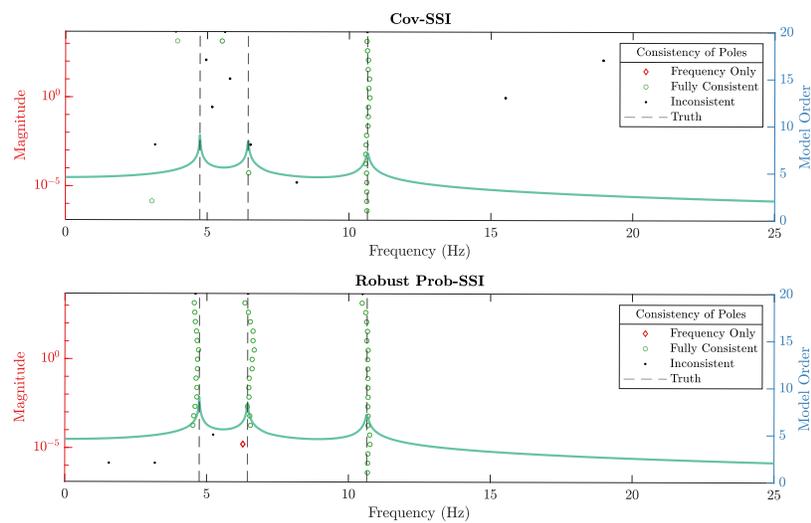


Figure E.7: Consistency diagrams recovered using Cov-SSI (top) and robust Prob-SSI (bottom) using response data from a simulated 3DOF ‘corrupted’ dataset, containing 0.5% artificially introduced, randomly located outliers (in each channel), set to a specified value, plus some small amount of noise.

## References

- [1] Bart Peeters and Guido De Roeck. Stochastic system identification for operational modal analysis: A Review. *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, 123(4):659–667, 2001.
- [2] Bart Peeters and Guido De Roeck. Reference-based stochastic subspace identification for output-only modal analysis. *Mechanical Systems and Signal Processing*, 13(6): 855–878, 1999.
- [3] G H James III, T G Carne, Lauffer, J P., and J P Lauffer. The Natural Excitation Technique (NExT) for Modal Parameter Extraction From Operating Wind Turbines. *The International Journal of Analytical and Experimental Modal Analysis*, 10(4):260–277, 1993.
- [4] Edwin Reynders. System Identification Methods for (Operational) Modal Analysis: Review and Comparison. *Archives of Computational Methods in Engineering*, 19(1): 51–124, 2012.
- [5] Chin Hsiung Loh, Jian Huang Weng, Yi Cheng Liu, Pei Yang Lin, and Shieh Kung Huang. Structural damage diagnosis based on on-line recursive stochastic subspace identification. *Smart Materials and Structures*, 20(5), 2011.
- [6] Torodd S. Nord, Øyvind W. Petersen, and Hayo Hendrikse. Stochastic subspace identification of modal parameters during ice-structure interaction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 377(2155), 2019.
- [7] Nan Jin, Y. B. Yang, Elias G. Dimitrakopoulos, Themelina S. Paraskeva, and Lambros S. Katafygiotis. Application of short-time stochastic subspace identification to estimate bridge frequencies from a traversing vehicle. *Engineering Structures*, 230(June 2020): 111688, 2021.
- [8] Francis R Bach and Michael I. Jordan. A probabilistic interpretation of canonical correlation analysis. Technical report, University of California, Berkeley, 2005.
- [9] Charles R Farrar and Keith Worden. *Structural Health Monitoring*. John Wiley and Sons, Ltd, Chichester, UK, nov 2012.
- [10] K. Worden, E. J. Cross, R. J. Barthorpe, D. J. Wagg, and P. Gardner. On Digital Twins, Mirrors, and Virtualizations: Frameworks for Model Verification and Validation. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 6(3), 2020.
- [11] Timothy J Rogers. *Towards Bayesian System Identification: With Application to SHM of Offshore Structures*. PhD thesis, University of Sheffield, 2018.

- [12] Shiqiang Qin, Juntao Kang, and Qiuping Wang. Operational Modal Analysis Based on Subspace Algorithm with an Improved Stabilization Diagram Method. *Shock and Vibration*, 2016, 2016.
- [13] Johann Priou, Szymon Gres, Matthieu Perrault, Laurent Guerineau, Johann Priou, Szymon Gres, Matthieu Perrault, Laurent Guerineau, and Michael Döhler Automated. Automated uncertainty-based extraction of modal parameters from stabilization diagrams. In *9th IOMAC - International Operational Modal Analysis Conference, 2022*.
- [14] Edwin Reynders, Rik Pintelon, and Guido De Roeck. Uncertainty bounds on modal parameters obtained from Stochastic Subspace Identificatio. *Mechanical Systems and Signal Processing*, 22(4):948–969, 2008.
- [15] Edwin Reynders, Kristof Maes, Geert Lombaert, and Guido De Roeck. Uncertainty Quantification in Operational Modal Analysis with Stochastic Subspace Identification: Validation and Applications. *Mechanical Systems and Signal Processing*, 66-67:13–30, 2016.
- [16] Edwin P.B. Reynders. Uncertainty quantification in data-driven stochastic subspace identification. *Mechanical Systems and Signal Processing*, 151:107338, 2021.
- [17] R Pintelon, P Guillaume, and J Schoukens. Uncertainty calculation in (operational) modal analysis. *Mechanical Systems and Signal Processing*, 21:2359–2373, 2007.
- [18] Sifeng Bi, Morvan Ouisse, and Emmanuel Foltete. A probabilistic approach for damping identification considering uncertainty in experimental modal analysis. *AIAA Journal*, 56:4953 – 4964, 2018.
- [19] P. Cheema, M. Makki Alamdari, G. A. Vio, F. L. Zhang, and C. W. Kim. Infinite mixture models for operational modal analysis: An automated and principled approach. *Journal of Sound and Vibration*, 491:115757, 2021.
- [20] Fulong Liu, Jigang Wu, Fengshou Gu, and Andrew D. Ball. An Introduction of a Robust OMA Method: CoS-SSI and Its Performance Evaluation through the Simulation and a Case Study. *Shock and Vibration*, 2019, 2019.
- [21] M. Goursat, M. Döhler, L. Mevel, and P. Andersen. Crystal clear ssi for operational modal analysis of aerospace vehicles. In Tom Proulx, editor, *Structural Dynamics, Volume 3*, pages 1421–1430, New York, NY, 2011. Springer New York.
- [22] Harold Hotelling. Relations Between Two Sets of Variates. *Biometrika*, 28(3/4):321, 1936.
- [23] Tohru Katayama. *Subspace Methods for System Identification: A Realisation Approach*. Springer, 2005.

- [24] Peter Van Overschee and Bart De Moor. *Subspace Identification for Linear Systems*. Springer, 1996.
- [25] Michael Döhler and Laurent Mevel. Efficient multi-order uncertainty computation for stochastic subspace identification. *Mechanical Systems and Signal Processing*, 38(2): 346–366, 2013.
- [26] Michael E. Tipping and Christopher M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.
- [27] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, sep 1977.
- [28] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, Cambridge, 2006.
- [29] Cédric Archambeau, Nicolas Delannay, and Michel Verleysen. Robust probabilistic projections. In *Proceedings of the 23rd International Conference on Machine Learning*, volume 148, pages 33–40, Pittsburgh, 2006.
- [30] RJ Allemang, DL Brown, and AW Phillips. Survey of modal techniques applicable to autonomous/semi-autonomous parameter identification. 2010.
- [31] J. Maeck and G. De Roeck. Description of Z24 benchmark. *Mechanical Systems and Signal Processing*, 17(1):127–131, 2003.
- [32] J. Maeck, B. Peeters, and G. De Roeck. Damage identification on the Z24 bridge using vibration monitoring. *Smart Materials and Structures*, 10(3):512–517, 2001.
- [33] Bart Peeters and Guido De Roeck. One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake Engineering & Structural Dynamics*, 30(2):149–171, 2001.
- [34] Ruben L. Boroschek and Joaquin A. Bilbao. Interpretation of stabilization diagrams using density-based clustering algorithm. *Engineering Structures*, 178(September 2018): 245–257, 2019.
- [35] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.