



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/208502/>

Version: Published Version

---

**Article:**

Haywood-Alexander, M., Mills, R.S., Champneys, M.D. et al. (2024) Full-scale modal testing of a Hawk T1A aircraft for benchmarking vibration-based methods. *Journal of Sound and Vibration*, 576. 118295. ISSN: 0022-460X

<https://doi.org/10.1016/j.jsv.2024.118295>

---

**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.

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Journal of Sound and Vibration

journal homepage: [www.elsevier.com/locate/jsv](http://www.elsevier.com/locate/jsv)

# Full-scale modal testing of a Hawk T1A aircraft for benchmarking vibration-based methods

Marcus Haywood-Alexander <sup>a,b,\*</sup>, Robin S. Mills <sup>a</sup>, Max D. Champneys <sup>a</sup>,  
Matthew R. Jones <sup>a</sup>, Matthew S. Bonney <sup>a</sup>, David Wagg <sup>a,c</sup>, Timothy J. Rogers <sup>a</sup>

<sup>a</sup> Dynamics Research Group, University of Sheffield, Mappin Street, Sheffield, S1 4ET, UK

<sup>b</sup> Chair of Structural Mechanics and Monitoring, ETH Zurich, Wolfgang-Pauli-Strasse, Zurich, 8049, Switzerland

<sup>c</sup> The Alan Turing Institute, London, NW1 2DB, UK

## ARTICLE INFO

Dataset link: <https://doi.org/10.15131/shef.data.22710040.v1>

### Keywords:

Modal testing  
Hawk T1A  
Full-scale dynamic testing  
Shaker test  
SIMO testing  
Aircraft dynamics

## ABSTRACT

Research developments for structural dynamics in the fields of design, system identification and structural health monitoring (SHM) have dramatically expanded the bounds of what can be learned from measured vibration data. However, significant challenges remain in the tasks of identification, prediction and evaluation of full-scale structures. A significant aid in the roadmap to the application of cutting-edge methods to the demands of in-service engineering structures, is the development of comprehensive benchmark datasets. With the aim of developing a useful and worthwhile benchmark dataset for structural dynamics, an extensive testing campaign is presented here. This recent campaign was performed on a decommissioned BAE system Hawk T1A aircraft at the Laboratory for Verification and Validation (LVV) in Sheffield. The aim of this paper is to present the dataset, providing details on the structure, experimental design, and data acquired. The collected data is made freely and openly available with the intention that it serve as a benchmark dataset for challenges in full-scale structural dynamics. Here, the details pertaining to two test phases (frequency and time domain) are presented. So as to ensure that the presented dataset is able to function as a benchmark, some baseline-level results are additionally presented for the tasks of identification and prediction, using standard approaches. It is envisaged that advanced methodologies will demonstrate superiority by favourable comparison with the results presented here. Finally, some dataset-specific challenges are described, with a view to form a hierarchy of tasks and frame discussion over their relative difficulty.

## 1. Introduction

How a structure behaves under dynamic loading is an imperative part of engineering design, operation, and maintenance. In both experimental and operational contexts, the measuring and monitoring of modal characteristics can provide vital information on the structure [1,2]. These modal characteristics can be used to, either directly or indirectly, infer the physical properties of the system. One such use of this information is to determine the existence of damage in important locations on the structure. When damage,

\* Corresponding author at: Dynamics Research Group, University of Sheffield, Mappin Street, Sheffield, S1 4ET, UK.

E-mail addresses: [m.haywood@ethz.ch](mailto:m.haywood@ethz.ch) (M. Haywood-Alexander), [robin.mills@sheffield.ac.uk](mailto:robin.mills@sheffield.ac.uk) (R.S. Mills), [max.champneys@sheffield.ac.uk](mailto:max.champneys@sheffield.ac.uk) (M.D. Champneys), [matthew.r.jones@sheffield.ac.uk](mailto:matthew.r.jones@sheffield.ac.uk) (M.R. Jones), [m.bonney@sheffield.ac.uk](mailto:m.bonney@sheffield.ac.uk) (M.S. Bonney), [david.wagg@sheffield.ac.uk](mailto:david.wagg@sheffield.ac.uk) (D. Wagg), [tim.rogers@sheffield.ac.uk](mailto:tim.rogers@sheffield.ac.uk) (T.J. Rogers).

<sup>1</sup> The work published here was undertaken at the University of Sheffield, but this author is now affiliated at ETH Zurich.

<https://doi.org/10.1016/j.jsv.2024.118295>

Received 8 September 2023; Received in revised form 18 January 2024; Accepted 20 January 2024

Available online 24 January 2024

0022-460X/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

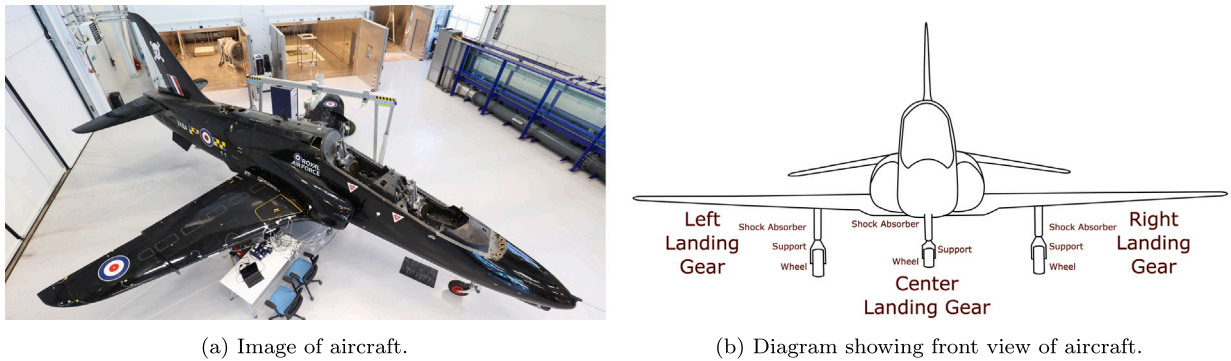


Fig. 1. Images of BAE systems Hawk T1A aircraft.

such as cracks or disbonds, occur on the structure, this corresponds to a localised loss of stiffness, which manifests as a change in the eigenfrequencies of the system. However, structural nonlinearities can also cause apparent changes to modal properties therefore causing false-positive or false-negative detection [3].

Using a well-defined model, it is possible to generate estimates of physical parameters using analytical, or purely-physics modelling (derived from first-principles analysis of elastic deformation theory). The presence of nonlinearity significantly complicates this task. The modelling of *known* nonlinearities is in itself a difficult task [4,5], however, when the nonlinearities are unknown, or ill-prescribed, in the model, this presents an even more difficult challenge. This, along with the challenge of learning model parameters, has led to the increased development of data-driven techniques to estimate or learn the models and parameters [6–8]. As these algorithms are developed, it is important that they are consistently assessed on meaningful examples, or benchmarks. The meaningfulness, as we define it here, of such benchmarks can be described by two characteristics; they present a reasonable challenge in that they that cannot be trivially solved by simple, well-established approaches, and that they provide data from a variety of loading or operating conditions.

In this work, we present a dataset with the aim to fulfil the requirements of a meaningful benchmark for structural modal analysis. In order to do so, experimental acceleration data was collected from a BAE system Hawk T1A aircraft undergoing a variety of dynamic input excitation, with simulated damage scenarios. The aim of the dataset is to provide a challenging benchmark for structural dynamics, in fields such as system identification and structural health monitoring. As well as presenting the data, an ancillary contribution of this paper is to also present a baseline standard in the tasks of system identification and structural health monitoring. The techniques presented here do not necessarily represent the state of the art in these two tasks but instead offer a first analysis; indicative of approaches that might be employed before more advanced techniques are sought. The rationale behind these choices is to establish how much insight can be gained from basic techniques and to act as a baseline against which future approaches can be assessed. A further motivation is to qualitatively establish the difficulty of selected system identification and structural health monitoring tasks on the Hawk data. This is an important undertaking as it highlights the issues presented when advanced machine learning (ML) methods are applied to case studies that can be solved by more simplistic, robust, and interpretable methods.

The Hawk dataset is freely provided with the motivation that it might be used to develop new algorithms for structural dynamics. The data is available in open *.hd5* format, and access is provided via a simple Python API, which is also provided.<sup>2</sup>

The remainder of this paper is arranged as follows. In Section 2, details of the aircraft structure are given, showing the boundary conditions and general information. This is followed by Section 3, where details on the experiment and data collection are provided, including information on the full experimental regime, data collected, control system, hardware, and acquisition. At this point the paper moves on to the baseline tasks applied to the dataset. Section 4 presents results of a series of system identification tasks applied to the dataset, both in the frequency and time domain. Following this, results from a structural health monitoring perspective are given in Section 5. Finally, in Section 6 some observations are provided on the challenges of the dataset, in order to showcase the use and meaningfulness of the data.

## 2. Structure

The structure used in this dataset is a BAE systems Hawk T1A aircraft [9]. This is a real, decommissioned aircraft, that was used for advanced training in the RAF. The aircraft was tested at the Laboratory for Verification and Validation (LVV) in Sheffield, England. For this dataset, only the starboard wing was measured, as well as supplementary measurements at the top of the aircraft and the exhaust, more details on the exact locations of the sensors will be given later in Section 3.2, Fig. 3. The aircraft was positioned resting on its wheels on the ground. Therefore, between the main body and the ground are the wheels, supports and hydraulic shock absorbers. All these substructures are joined with a variety of fixing methods and types, which are not detailed here. Fig. 1 shows both an image of the real aircraft, as well as a schematic, indicating the boundary conditions of the wing.

<sup>2</sup> All data and documentation for the API is available at <https://doi.org/10.15131/shef.data.22710040.v1>.

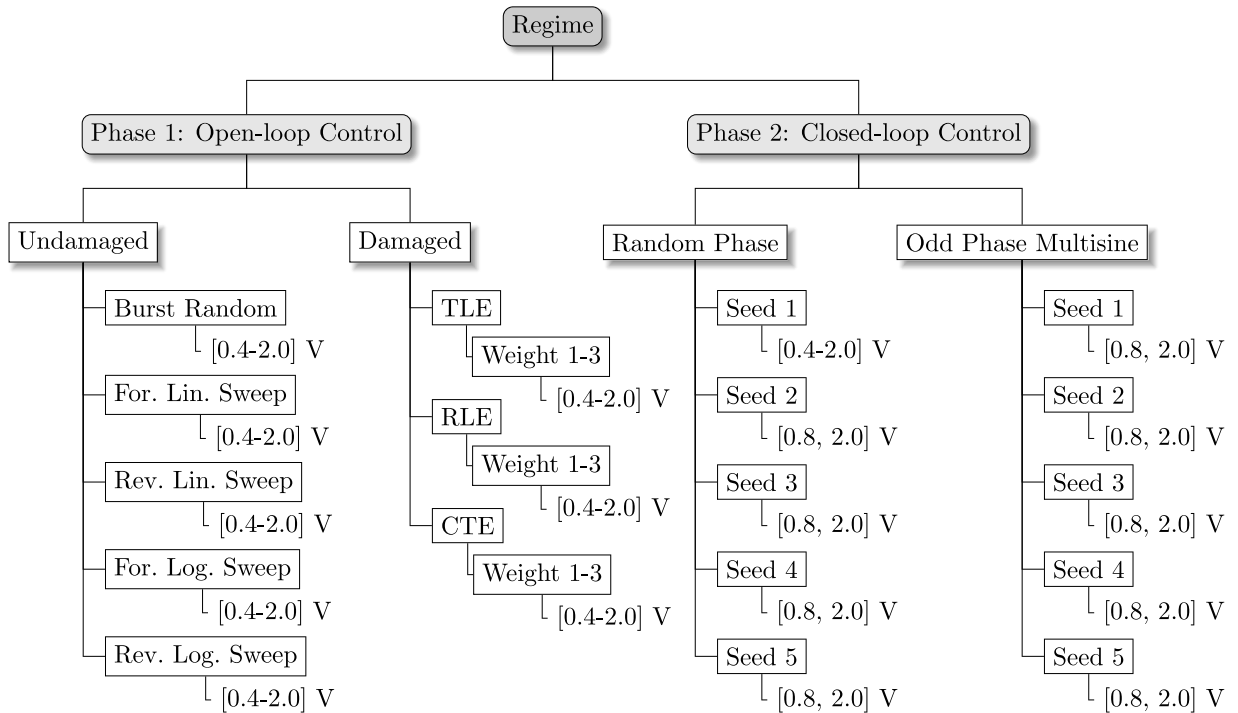


Fig. 2. Testing regime for dataset. For the damaged tests, the acronyms represent the locations of the added mass, where TLE, RLE, and CTE stand for *tip-leading-edge*, *root-leading-edge*, and *central-trailing-edge*, respectively. The weight numbers are ID-representations, the value of mass is provided in the experiment details.

### 3. Experiment

The aim of this work was to generate modal analysis data for the starboard wing of the Hawk aircraft. To do so, the wing was excited with a shaker, and the acceleration response over the upper and lower surfaces of the wing were measured using accelerometers, and processed to return the response in the frequency domain. This section will detail all the experiment details, beginning with the testing regime, the experiment setup and hardware, and then acquisition and processing details.

#### 3.1. Regime outline

The total experimental regime was split into two phases; open-loop control using off-the-shelf hardware, and a custom-built, closed-loop control system. Specific details on the control system and hardware will be given in Section 3.2. Each of these experimental regimes followed a hierarchical structure based on the aims of the experiment; this is shown in Fig. 2. The lowest level of this hierarchy is herein referred to as a ‘test’, and each test was performed with 10 repetitions. For clarification; a repetition is where the run is finished, allowed to reset, and then repeated. This repeat style was done to allow for quantification and mitigation of experimental error.

##### 3.1.1. Phase 1: Open-loop control

In the first phase, there were two main sub-regimes; the nominally-undamaged state, and the pseudo-damaged state. In the first case, the aircraft was left in its assumed healthy (free from significant structural damage) condition, and in the second case, pseudo-damage was ‘simulated’ by adding masses of different weights in specified locations. Additional masses were used to simulate pseudo-damage of the system, which can be viewed as an emulation of stiffness loss, or other functionally-inverse system properties [10]. Naturally, the additional mass is not a simple directly-inverse representation of damage – often manifesting as a loss in stiffness – however, it is used here for the motivation of novelty-detection and localised system-identification.

For the undamaged tests, the method of forcing input was set to either burst random, or forward/reverse sine sweeps in the linear/logarithmic scale. The burst random uses a white noise forcing input, the amplitude of which is ramped at the beginning and end of the burst. For the sine sweeps, the forcing input is a sine wave signal of which the frequency increases during the time period of the measurement. Forward/reverse refers to the direction of the frequency sweep, and linear/logarithmic refers to the pattern of the sweep. This means that for forward logarithmic sweeps, more time is spent exciting the lower frequencies, and for reverse logarithmic sweeps, more time is spent exciting the higher frequencies. For each of the excitation types, the amplitude of the forcing was increased by increasing the output voltage in increments of 0.2 V.

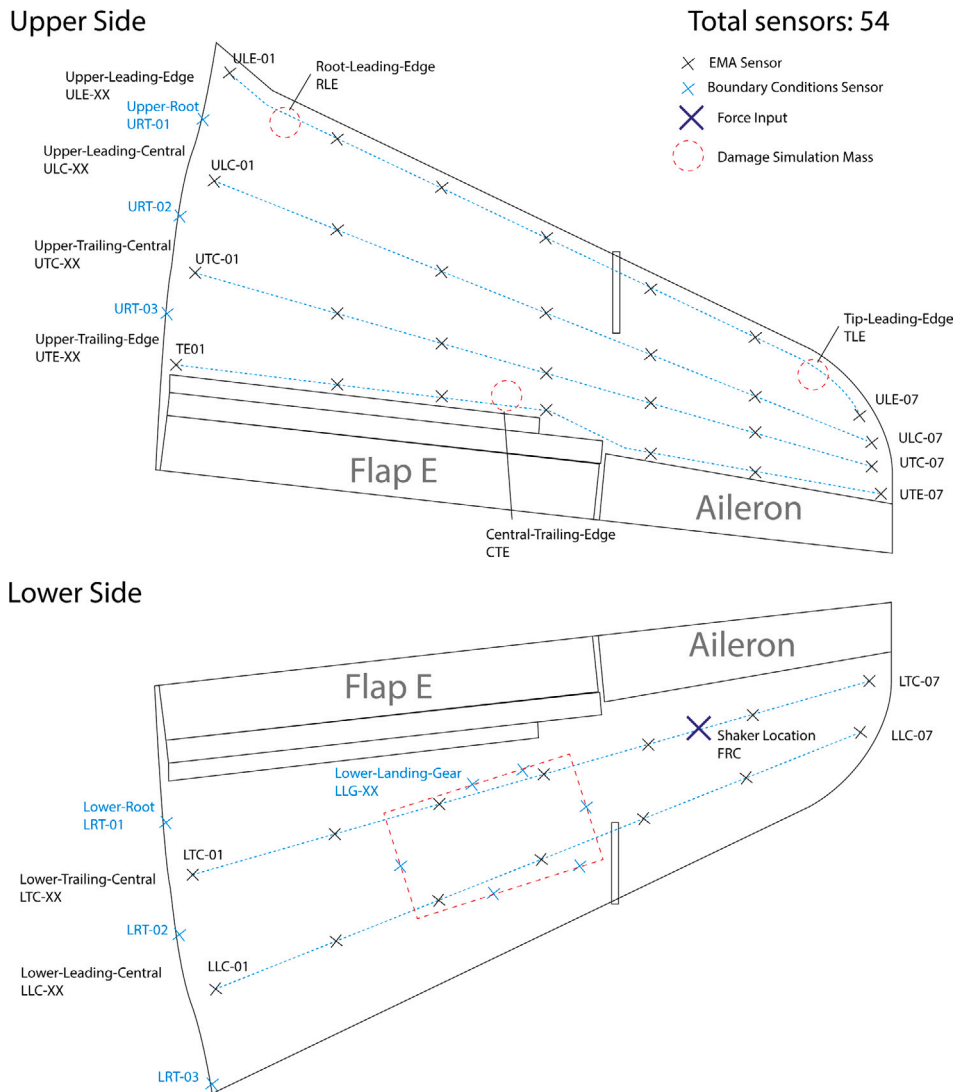


Fig. 3. Sensor layout on the wing, also indicating the locations of the shaker and additional masses.

For the damaged cases in Phase 1, only burst random excitation was used, and to simulate a range of damage cases, the weight and location of the added masses was changed. In Fig. 2, TLE, RLE, and CTE refer to these locations, and stand for *tip-leading-edge*, *root-leading-edge*, and *central-trailing-edge*, respectively. The locations on the wing are shown in Fig. 3, and exact details are given in the accompanying technical report [11]. Much like the undamaged tests, the amplitude of the forcing was increased by increasing the voltage, but this time in increments of 0.4 V.

### 3.1.2. Phase 2: Closed-loop control

The second phase of testing is split into sub-regimes based on the target signal types, by iteratively generating the signal passed to the shaker. More details on the control algorithm are given in Section 3.2.1, along with the definition of each of the actuation signal types. To determine information on the effect of randomness, each of these signal types are tested with different initial seeds. To increase the amplitude of the forcing, as this is a closed-loop control which uses the measured amplitude, changing the voltage output will simply cause the control algorithm to decrease the signal level. Therefore, instead the amplitude of the target spectra were increased to corresponding levels. It is important to note, these do not necessarily correspond exactly to the levels used in Phase 1, they are instead values which return amplitudes of a similar spectral density.

### 3.2. Setup and hardware

To measure the structural response, PCB Piezotronics accelerometers were used with a nominal sensitivity of  $10 \text{ mV g}^{-1}$ . The sensor-specific sensitivities had a mean, minimum, and maximum of 10.42, 9.77, and  $10.69 \text{ mV g}^{-1}$ , respectively. When converting

**Table 1**  
Hardware components for closed-loop control system.

Name	Quantity	Use
NI-cDAQ-9179	2	Chassis for total system
NI-9232	20	Measures voltage signals from accelerometers
NI-9260	1	Outputs voltage signal

voltage readings to acceleration values, the sensor-specific sensitivities, accurate to  $0.01 \text{ mV g}^{-1}$ , were used. A total of 54 sensors were placed on the wing, which are placed in three lines on the upper surface, and two lines on the lower surface of the wing, where a line runs from the root to the tip. As well as these 5 strips, additional sensors were placed to provide information at boundary condition locations of the root and the open panel for the landing wheel. The force was applied to the lower surface of the wing using a Tira TV 51140-MOSP shaker, attached directly to a PCB Piezotronics 208C02 force transducer with a sensitivity of  $11.11 \text{ mV N}^{-1}$ . The voltage signal between the actuation and shaker was amplified using a Brüel & Kjær LDS LPA 100, which was set throughout at 50% gain (corresponding to an  $11 \text{ V V}^{-1}$  gain). This gain is set by an analogue dial, so for the open-loop control regime, this may induce small variations on the *magnitude* of the shaker output. The gain is reset in between test repetitions so that this variation can be assessed. For all data, the signal voltage values and individual sensitivities are provided in the .hd5 files.

A diagram showing the locations of the sensors, forcing, and masses is given in Fig. 3. The locations of the sensors in this diagram are the nominal locations, but may have been slightly shifted in order to avoid components of the wing, exact coordinates are provided in the data and the technical report [11]. The black sensor locations indicate those placed for the purposes of experimental modal analysis of the general structure. The blue sensor locations indicate those placed for the collection of data relating to the boundary conditions. Also shown are the shaker attachment location, indicating where the excitation force is applied, and the locations of the added masses and their corresponding ID tags. The red dashed box indicates the area on the lower side of the wing which is open for the extension and retraction of the landing gear. Sensors were fixed to the edges of this open area, with the prefix ‘LLG’, in order to allow for benchmarking methods which incorporate boundary condition approaches.

The locations of the sensors were chosen in order to capture the longitudinal, lateral and torsional behaviour of the wing, as well as boundary condition information. To do this, they were placed in spatially-asynchronous strips between the upper and lower surfaces, as well as close to the edges of the wing on the upper surface. In order to excite torsional modes sufficiently, the shaker was placed off the central axis of the wing, and was also placed towards the wing tip so as to excite at large enough acceleration amplitudes without requiring excessive forces.

For the open-loop control, a Siemens PLM LMS SCADAS unit was used to generate the forcing signal, and simultaneously measure the acceleration responses. For the closed-loop control system, National Instruments hardware was used, along with LabView, to develop a frequency-domain control system, targeting the force signal at the location of the shaker. A table of the full hardware is shown in Table 1.

### 3.2.1. Control algorithm

In the second phase of testing, a custom algorithm was implemented to control the actuation signal in the frequency domain. This allows the use of pre-prescribed forcing types, for both amplitude and phase of the frequencies. The overall control flow is shown in Fig. 4, showing the process which covers one repeat of a test. The general aim of the algorithm is to control the output of the shaker to the structure in the frequency domain. For this work, the frequency domain information is defined and calculated as the power-spectral-density (PSD), and contains both magnitude and phase information. The target PSD of this force is denoted  $S_r(\omega)$ , and the measured PSD from the transducer is denoted as  $S_f(\omega)$ . The algorithm estimates a model of the mapping between the driving signal, in terms of voltage, to the applied force. This is achieved by assuming a linear transfer function,

$$|F(\omega)| = |H(\omega)||S_d(\omega)| \quad (1)$$

where  $F(\omega)$  is the force applied to the structure,  $H(\omega)$  is the transfer function, and  $S_d(\omega)$  is the drive signal in volts, all in the frequency domain as functions of the circular frequency  $\omega$ .

The target actuation  $S_r(\omega)$  is defined in terms of the PSD of the force applied. Details of how these target signals are generated are provided in Section 3.2.2. At the first step of the control algorithm, the drive PSD  $S_d(\omega)$  is simply set to the reference (i.e.  $|H| = 1$ ). At each iteration, the drive signal  $s_x(t)$ , is calculated using an inverse Fourier transform,

$$s_d(t) = \mathcal{F}^{-1}(|S_d(\omega)|) \quad (2)$$

When the actuation signals were generated, a linear amplitude ramp was applied to window the signal in the time domain, to reduce shock to the structure and equipment. This signal is sent out from the control system, and the PSD of the force,  $S_f(\omega)$  is measured. This measurement is then used to update the estimate of the linear transfer function using,

$$|H(\omega)| = \alpha \sqrt{\frac{S_r(\omega)}{S_f(\omega)}} \quad (3)$$

where  $0.0 \leq \alpha \leq 1.0$  is a proportional weighting factor which is included to slowly iterate the estimated transfer function to mitigate potential out-of-control issues, or overloading of the system. The value of this is chosen by the control system designer, and in this

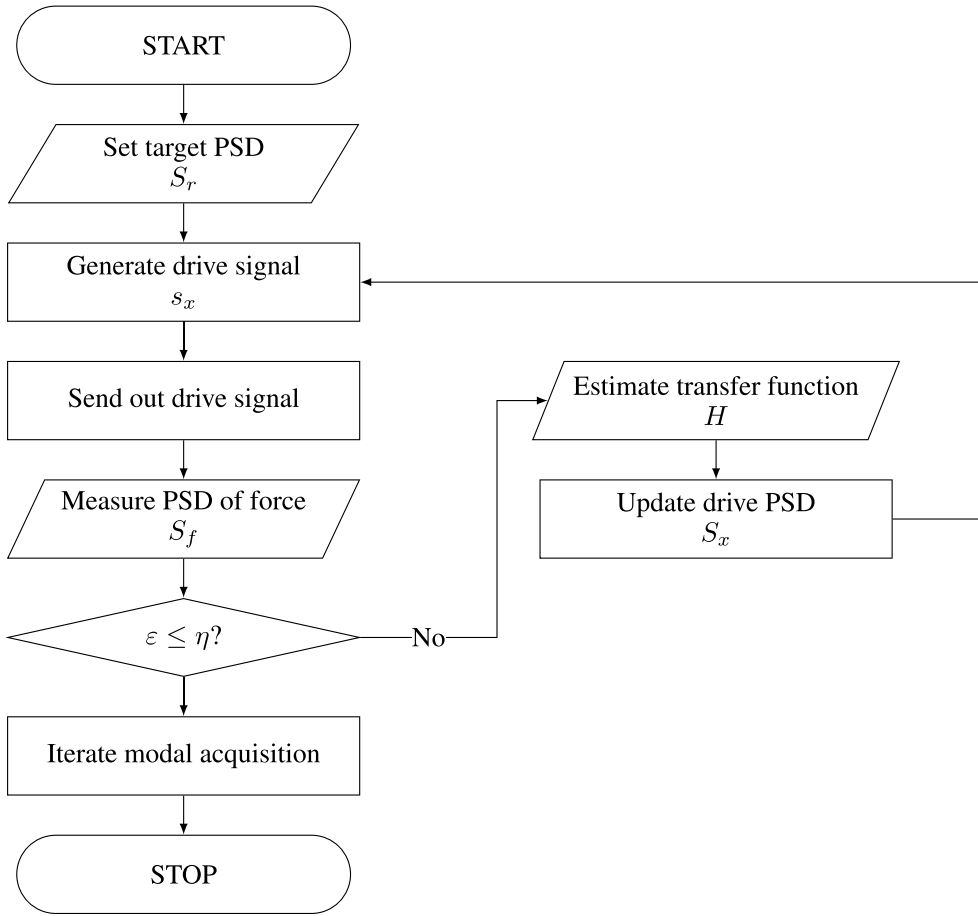


Fig. 4. Control algorithm process flow, indicating the total workflow for one *repeat* of a test.

case was chosen to be 0.8. Then, the drive PSD is updated, and subsequently the drive signal using Eq. (2),

$$|S_d(\omega)| = \frac{|S_r(\omega)|}{|H(\omega)|} \quad (4)$$

During the drive updating, the target PSD error is defined as the mean squared error,

$$\epsilon = \frac{1}{N} \sum_{i=1}^N (|S_f(\omega^{(i)})| - |S_r(\omega^{(i)})|)^2 \quad (5)$$

where  $N$  is the number of frequency points in the discretised frequency domain. Once the error is below a specified target,  $\eta$ , the latest drive PSD is used for the acquisition stage. For this work, the target error was set to 0.01.

### 3.2.2. Frequency domain actuation types

This subsection aims to explain the difference between the actuation types in Phase 2, which are defined in the frequency domain. Both actuation types generate the time domain signal as a sum of the discrete frequencies represented by their definitions in terms of power-spectral-density. The actuation is defined using a vector of frequencies,  $\omega$ , with corresponding vectors of amplitude,  $\mathbf{A}$ , and phase,  $\phi$ . Some examples of these for random phase and odd-phase frequency domain definitions are shown in Fig. 5. Where  $N$  is the length of these vectors, and is the same as the PSD definitions in the above subsection, the discrete time signal,  $s(t)$  is defined as scalar function of the discrete time points,

$$s(t) = \sum_{i=1}^N A_i \sin(\omega_i t + \phi_i) \quad (6)$$

The vector of frequencies is chosen based on the bandwidth of modes of interest, and its discretisation is aligned with the acquisition parameters. Generally, the vector  $\mathbf{A}$  is designed to ‘ramp up and ramp down’ the amplitude with respect to the frequencies. This is done in order to maintain periodicity in the time domain signal to avoid shock in the physical system. The

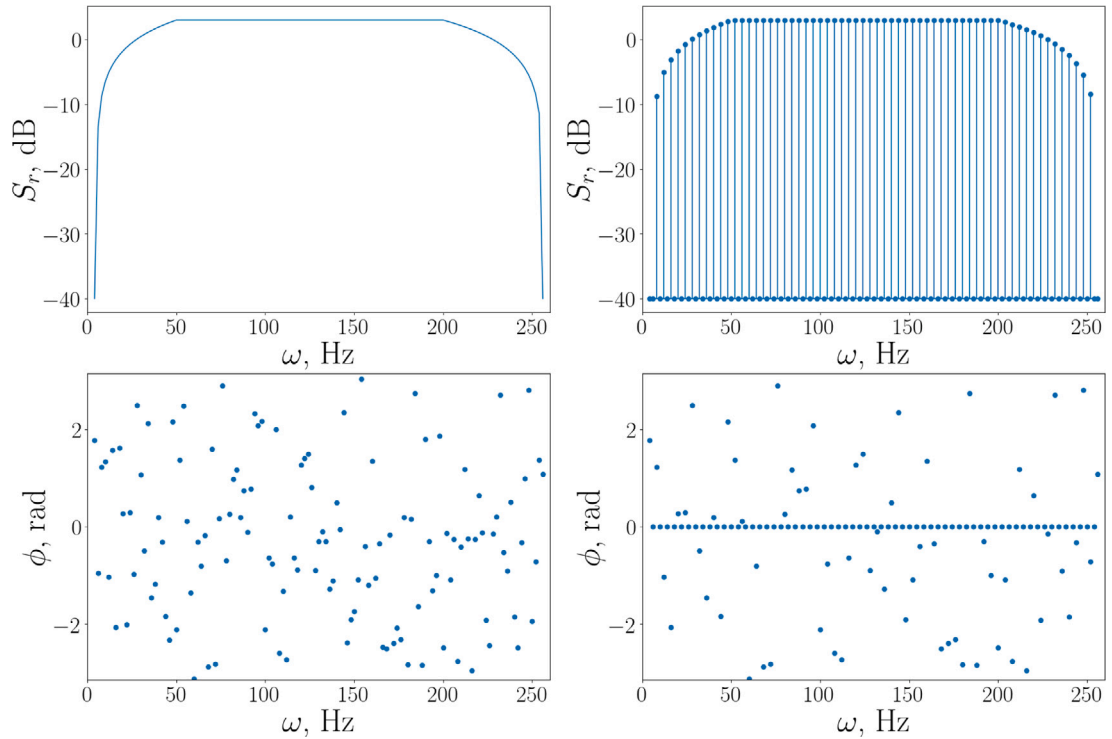


Fig. 5. Examples of the target driver PSDs for (left) random phase and (right) odd phase multisine types. The top plots show the target PSD power in terms of dB, and the bottom plots show the corresponding phase.

**Table 2**  
Acquisition parameters for each test phase.

Phase 1		Phase 2	
Parameter	Value	Parameter	Value
Sample rate	512 Hz	Sample rate	2048 Hz
Number of samples	16,384	Number of samples	32,768
Burst/sweep time	90%	Burst time	90%
Amplifier level	50%	Amplifier level	50%
Repetitions	10	Repetitions	10

phase is generated by randomly sampling from a uniform distribution between 0 and  $2\pi$ . Using these generation methods with no modification forms the **random phase** actuation type. However, for the **odd phase** actuation, every *other* element in the vectors  $A$  and  $\phi$  are set to zero,

$$A_i = 0, \quad \phi_i = 0, \quad \text{for } i = 2n - 1 \quad \text{for } 0 \leq n \leq N/2 \quad (7)$$

### 3.3. Acquisition and processing

For all the tests, a number of acquisition parameters (listed in Table 2) were kept constant. The number of samples indicates the number of sampling points taken in the time domain. The burst/sweep time represents the portion of the measured signal in which the random burst or the sine sweep takes place; i.e. for 90% burst time, the first and last 5% of the acquisition time has no actuation.

The data is then saved with identifiers which relate to the test series as shown in Fig. 2. The sub-regime identifiers are shown in Table 3, where the placeholder 'X' in the identifiers are the test numbers identifiers. These are integer values which correspond to the voltage level, added mass weight, and/or seed used. Further details of the exact labels for each individual test can be found in the technical report [11].

### 3.4. Dataset structure

In order to keep this paper self-contained, a brief description of the dataset structure is included here. All data is stored in the open-source *hierarchical data format* (hd5). The hd5 format can be accessed using several open-source tools, but the authors also

**Table 3**  
Test identifiers used for naming data in acquisition and processing.

Phase 1		Phase 2	
Undamaged	Damaged	Random phase	Odd phase multisine
BR_AR_X	DS_TLE_X	RPH_AR_X	ORP_VS_X
FSS_AR_X	DS_RLE_X	FLS_AR_X	
RSS_AR_X	DS_CTE_X	RLS_AR_X	
FLS_AR_X			
RLS_AR_X			

provide a python interface.<sup>3</sup> The hd5 format essentially mimics the tree-like file structure of modern computers. Data can be accessed by specifying the path to the desired sensor channel and data series. The general form of the path for the Hawk SBW test data is:

```
/SBW/{test phase}/{test series}/{repetition}/{sensor}/{signal}
```

where SBW (starboard wing test) is the overall test campaign the bracketed quantities refer to the following optional fields:

- **test phase**  $\in \{\text{LMS}, \text{NI}\}$ : Refers to the frequency domain and time domain data respectively, as per Table 2 and Fig. 2.
- **test series**  $\in \{\text{BR\_AR}, \text{DS\_TLE}, \dots\}$  (Phase 1) and  $\in \{\text{RPH\_AR}, \text{ORP\_VS}\}$  (Phase 2): Test series identifier (see Fig. 2 for definitions).
- **repetition**  $\in \{01, 02, \dots\}$ : Test repetition number. For each test series there are nominally 10 repetitions, although for some tests fewer repetitions are available.
- **sensor**  $\in \{\text{FRC}, \text{LLE-01}, \dots, \text{LLC-07}\}$ : Sensor location (see Fig. 3 for definitions).
- **signal**  $\in \{\text{coherenceSpectrum}, \text{spectra}, \text{frf}\}$  (Phase 1) and  $\in \{\text{acc}, \text{force}\}$  (Phase 2): Desired signal. Note that frequency-response-functions (FRFs) are only available for output channels in the phase 1 tests.

The self-describing nature of the .hd5 format allows metadata to be stored alongside the collected series. Test data related to sensor position, sensitivities, units and test parameters (added mass, excitation level etc.) can all be accessed by querying the metadata fields at each level of the .hd5 hierarchy. Additional detail regarding the dataset structure and interface can be found at <https://doi.org/10.15131/shef.data.22710040.v1>.

#### 4. System identification

With the description of the experimental campaign complete, attention can now be turned to establishing a benchmark for the Hawk data in two common tasks in structural dynamics analysis *system identification* and *structural health monitoring*. It is envisaged that the results here will act as a baseline level against which more advanced methods might be compared in future work.

One of the key challenges in structural dynamics is obtaining accurate models of system dynamics. For structures that can be considered predominantly linear, the default framework for modelling dynamics is that of *modal analysis* [12]. Modal analysis is familiar to many engineers as a method of decomposing system dynamics into a number of eigenfrequencies and eigenfunctions that encode the response of the system to any input. These properties are invariant to any excitation and can be conveniently expressed in terms of a discrete number of interpretable properties; the natural frequencies  $\omega_r$ , damping ratios  $\zeta_r$  and modal matrix  $\Phi$ .<sup>4</sup>

Modal analysis [12] is a general term for methods that attempt to specify the modal properties from measured structural dynamics data. Broadly, these methods can be categorised by whether the input signal is available<sup>5</sup> and whether the identification operates in the time and frequency domains. In this paper, two methods are presented on the proposed Hawk dataset. In the frequency domain, the EMA method of rational fractional polynomials (RFP) [13] is employed. In the time domain, an output-only stochastic subspace identification (SSI) is shown.

Given the ubiquity of modal analysis in structural identification, many methods have been proposed for performing both EMA and OMA each with their own advantages and disadvantages. A common choice for performing EMA is the commercially-available PolyMAX algorithm as part of the Siemens PLM LMS system. Although well-known, the software is proprietary and, therefore, not directly reproducible without access to the DAQ system and software used in testing. In this paper, the similar method of rational fractional polynomials (RFP) [13] is used instead to give researchers transparency on the analysis and allow the same processing of the data not collected through this DAQ system if a user so desired. For OMA, one of the most popular commodity algorithms for identification is stochastic subspace identification (SSI) [14]. As such, the authors feel that SSI provides a natural baseline identification result for the category of OMA methods.

<sup>3</sup> A python API for the hawk data is available at <https://github.com/MDCHAMP/hawk-data>.

<sup>4</sup> In this section, the  $r$ th vector of the modal matrix (commonly known as the  $r$ th modeshape) will be denoted by  $\phi^{(r)}$ .

<sup>5</sup> Methods where the input to the system is known are typically referred to as experimental modal analysis (EMA) and identification procedures that do not use the input signal are known as operational modal analysis (OMA).

#### 4.1. Modal analysis: Frequency domain

Frequency domain methods for EMA rely on approximating the form of the frequency response function (FRF), often employing techniques from curve fitting. Assuming that the dynamics are linear, the FRF is a linear transfer function from the Fourier-transformed input force (N) to the output acceleration (g) that has a generalised form that depends only on the modal properties. In order to keep the notation in this section general (and in keeping with the system identification literature), the input forcing spectra will be denoted  $X(\omega)$  and the output acceleration  $Y(\omega)$ , where  $\omega$  is the vector of frequencies for which data were recorded. The accelerance FRF can be expressed as,

$$H(\omega) = \frac{Y(\omega)}{X(\omega)} \quad (8)$$

The expression for a general accelerance FRF in terms of the modal parameters is,

$$H_{il}(\omega) = \sum_{r=1}^n \frac{-\omega^2 \phi_i^{(r)} \phi_j^{(r)}}{-\omega^2 + 2j\zeta_r \omega_r \omega + \omega_r^2} \quad (9)$$

where  $\omega_r$ ,  $\zeta_r$  and  $\phi^{(r)}$  are the natural frequency, damping ratio and modeshape corresponding to the  $r$ th vibration mode and  $i$  and  $l$  are the indices of the input and output location respectively.

The method of RFP works by approximating (9) as a ratio of two polynomials. Considering here a single-input, single output (SISO) system comprised of  $n$  modes, one has,

$$H(\omega) = \frac{\sum_{k=0}^m a_k (j\omega)^k}{\sum_{k=0}^n b_k (j\omega)^k} \quad (10)$$

where the coefficients  $a_k$  and  $b_k$  depend on the modal properties. Note that the number of terms in the numerator polynomial does not necessarily have to be the same as the denominator. Additional terms in the numerator polynomial can be used to compensate for out-of-band effects (i.e modes from outside the frequency range of interest).

RFP learns the  $a_k$  and  $b_k$  coefficients from the measured  $H$  and then infers the modal properties. Rather than learn the coefficients by optimisation, it is preferable to construct a least squares scheme. In order for the coefficients to be identifiable we must include a constraint that fixes one of the polynomial coefficients (or else there would be a constant factor invariance). Here, after [13], the lead coefficient of the denominator is arbitrarily set to unity  $b_n = 1$ . In order to simplify notation, the vectors  $\mathbf{a}$ ,  $\mathbf{b}$  are defined as the unknown coefficients of the numerator (up to order  $m$ ) and the denominator (up to order  $n - 1$ ) polynomials respectively.

Representing the rational FRF form above in matrix form,

$$H = (\Phi_a \mathbf{a})(b_n (j\omega)^n + \Phi_b \mathbf{b})^{-1} \quad (11)$$

where the matrices  $\Phi_a$  and  $\Phi_b$  are polynomial bases of order  $n$  and  $m$  respectively for the vector of frequency lines  $\omega$ , and  $(j\omega)^n$  is the vector  $j\omega$  raised to the power  $n$  element-wise. In order to construct the least squares scheme, the error between the measured  $\hat{H}(\omega)$  and predicted FRFs can be expressed as,

$$\mathbf{e} = \Phi_a \mathbf{a} - ((j\omega)^n + \mathbf{b}\Phi_b) \hat{H}(\omega) \quad (12)$$

In order to simplify the forthcoming notation (and for compatibility with the notation of [13]), the following notation is introduced here,

$$\mathbf{P} = \Phi_a \quad (13)$$

$$\mathbf{T} = \Phi_b \hat{H}(\omega) \quad (14)$$

$$\mathbf{w} = b_n (j\omega)^n \hat{H}(\omega) \quad (15)$$

Thus the objective function becomes,

$$J(\mathbf{a}, \mathbf{b} | \hat{H}) = \mathbf{e}^\dagger \mathbf{e} = (\mathbf{P}\mathbf{a} - (\mathbf{w} + \mathbf{T}\mathbf{b}))^\dagger (\mathbf{P}\mathbf{a} - (\mathbf{w} + \mathbf{T}\mathbf{b})) \quad (16)$$

where  $\dagger$  represents the Hermitian transpose. The above is a positive quadratic in both  $\mathbf{a}$  and  $\mathbf{b}$  and therefore has a single optimum  $(\mathbf{a}^*, \mathbf{b}^*)$ , obtained exactly by setting the partial derivatives in  $\mathbf{a}$  and  $\mathbf{b}$  to zero,

$$\left. \frac{\partial J(\mathbf{a}, \mathbf{b} | \hat{H})}{\partial \mathbf{a}} \right|_{\mathbf{a}^*, \mathbf{b}^*} = 0 \quad (17)$$

and,

$$\left. \frac{\partial J(\mathbf{a}, \mathbf{b} | \hat{H})}{\partial \mathbf{b}} \right|_{\mathbf{a}^*, \mathbf{b}^*} = 0 \quad (18)$$

Now evaluating these derivatives starting with the former condition,

$$\frac{\partial J}{\partial \mathbf{a}} = 2\Re(\mathbf{P}^\dagger \mathbf{P})\mathbf{a} - 2\Re(\mathbf{P}^\dagger \mathbf{T})\mathbf{b} - 2\Re(\mathbf{P}^\dagger \mathbf{w}) \quad (19)$$

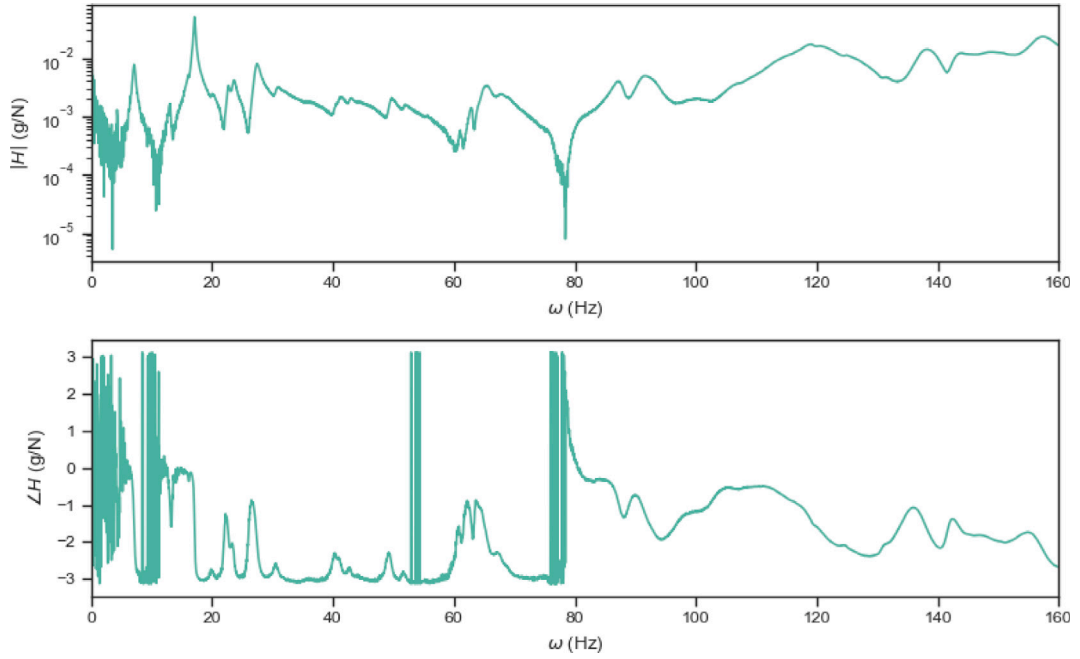


Fig. 6. FRF from sensor LTC-05 subject to burst-random excitation at 0.4 V.

where  $\Re$  denotes the real part of the complex vector. Next, for the second condition,

$$\frac{\partial J}{\partial \mathbf{b}} = -2\Re(T^\dagger P)\mathbf{a} + 2\Re(T^\dagger T)\mathbf{b} + 2\Re(T^\dagger \mathbf{w}) \quad (20)$$

Finally, applying the condition of zero gradient at the optimum, one has the block-linear system that can be solved for the polynomial coefficients.

$$\Re \left( \begin{bmatrix} P^\dagger P & -P^\dagger T \\ -T^\dagger P & T^\dagger T \end{bmatrix} \right) \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \Re \left( \begin{bmatrix} P^\dagger \mathbf{w} \\ -T^\dagger \mathbf{w} \end{bmatrix} \right) \quad (21)$$

It is worth noting that this expression differs in several respects from the expression in [13] due to the presence of several errors in the original manuscript. Finally the modal properties (natural frequencies and damping ratios) can be recovered from the coefficients of the denominator polynomial as,

$$\omega_r = |R_r| \quad (22)$$

$$\zeta_r = \frac{-\Re(R_r)}{|R_r|} \quad (23)$$

where  $R_r$  is the  $r$  th complex-conjugate root of the denominator polynomial in (10). Extraction of the modeshapes is also possible, requiring both the coefficients of the numerator and the denominator polynomials. However it is not considered here.

In Fig. 6, the FRF of the hawk at sensor LTC-05 (Lower trailing edge centre, burst-random excitation, 0.4 V LMS input voltage, averaged over 10 repeats) is plotted. In order to demonstrate the RFP approach, some modal properties will be extracted from this signal.

As a first analysis, all frequency lines are included in the computation of the modal properties with the number of modes set to 10 (corresponding to a 20th order polynomial). The resulting FRF is plotted in Fig. 7 with dashed lines corresponding to the natural frequencies. As can be seen, the higher frequencies are well approximated but there is significant errors in the lower frequencies and many of the dominant modes have been missed entirely. This is due to an inherent bias towards higher frequencies in the RFP method that can be attributed to the choice of the error criterion  $e$ .<sup>6</sup>

A more practical approach (indicative of how frequency-domain RFP is often implemented) is to instead perform the curve fit on a number of pre-selected bounds. In order to compensate for the effects of ‘out-of-bound’ modes, additional terms can be added to the numerator polynomial. Here, for each range, 6 additional terms are added to the numerator polynomial. The predicted FRF and natural frequencies are plotted in Fig. 8 for identification bounds given in Table 4. In the figure, the quality of the FRF fit is

<sup>6</sup> This bias can be seen by noting that the choice of  $e$  does not correspond directly to the difference between the measured and predicted FRF, but a difference that is scaled by the frequency.

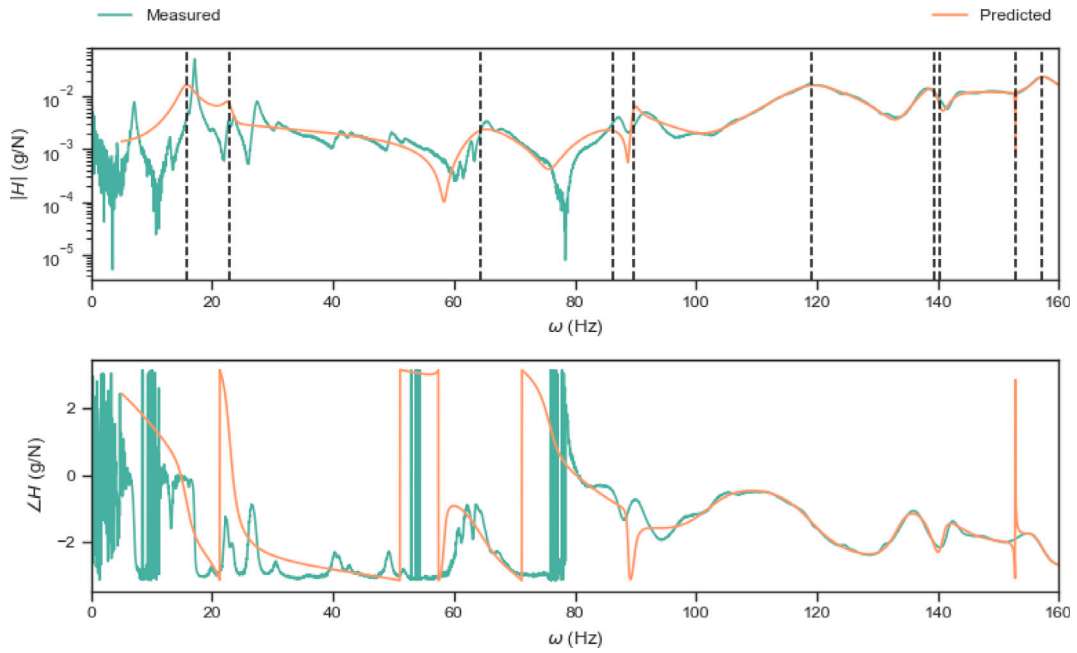


Fig. 7. Measured vs. identified RFP FRF for the sensor at LTC-05, vertical lines are the identified natural frequencies.

**Table 4**  
Bands and identified modal parameters from the banded RFP identification.

Lower bound (Hz)	Upper bound (Hz)	Modes in band	$\omega_n$ (Hz)	$\zeta_n$
5	9	1	7.094	0.02434
12	14	1	13.05	0.01552
15	19	1	17.1	0.005918
22	24	2	22.66, 23.56	0.008452, 0.009352
26	30	1	27.35	0.01584
30.5	31	1	30.77	0.0004578
35	37	1	36.06	0.00005626
40.5	44	2	41.16, 42.85	0.008827, 0.007161
48.5	54	2	49.61, 52.07	0.009950, 0.00469
60	62	1	60.81	0.003912
62.5	63.5	1	63.01	0.004542
63.8	70	2	67.39, 65.23	0.004978, 0.01048
86	90	1	87.41	0.01138
92	100	1	92.54	0.02798
112	118.5	1	116.9	0.01307
119.5	122	1	120	0.003427
122	125	1	124.6	0.008256
135	138	1	124.6	0.008256
154	162	1	157.1	0.01505

visibly superior to the global fit and the peak locations are well identified within the selected bands. Note that the total FRF can be reconstructed by first estimating the damping ratios and then using the general form of the modal transfer function although this is not considered here. It is clear that such an approach cannot be considered to be ‘automatic’ and that the performance comes to depend strongly on the choice of the identification ranges. However it is important to consider such approaches as a baseline to which more advanced methods might be compared.

#### 4.2. Modal analysis: Time domain

Several approaches to EMA consider the identification problem in the time-domain. Of these, a common approach is stochastic subspace identification (SSI) [15]. The SSI approach outlined here has the added advantage that it operates in an output-only setting (subject to a broadband excitation assumption). This is especially attractive for applications whereby measurement of system input in-service (such as in an aircraft during flight) is challenging.

The collection of time-series data from the Hawk allows the effectiveness of SSI to be tested. Here, a classical approach of “covariance-driven” SSI (Cov-SSI) is used. In this work the terms SSI and Cov-SSI will now be used interchangeably. A full

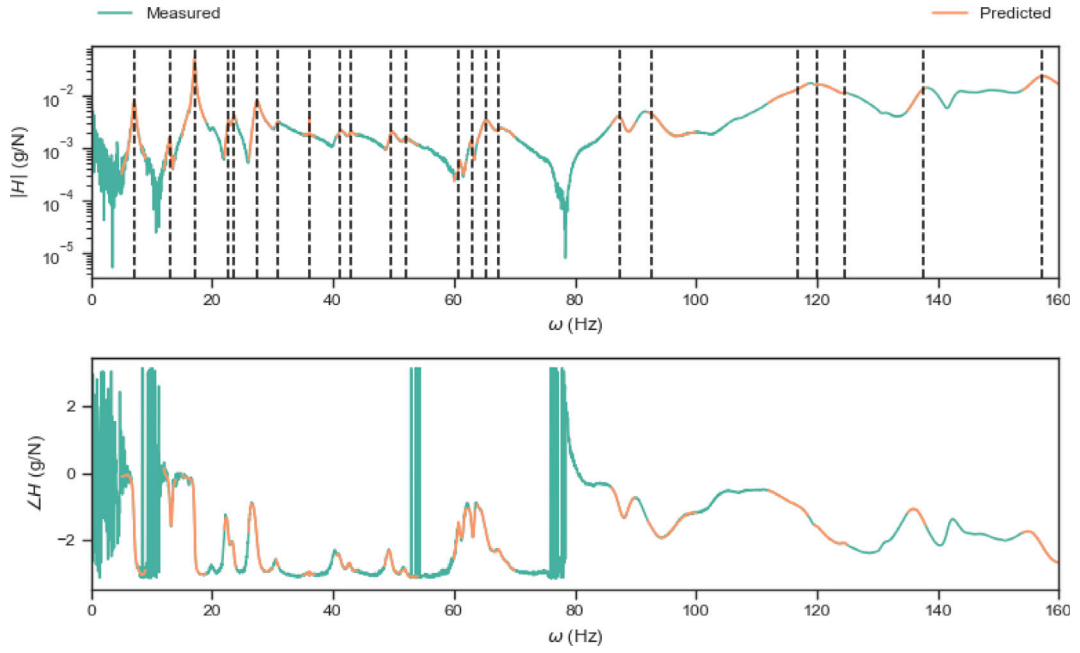


Fig. 8. Measured vs. identified banded RFP FRF for the sensor at LTC-05, vertical lines are the identified natural frequencies.

background to SSI will not be presented in this paper in the name of brevity, instead the reader is referred to [14,15] for general introductions to the family of methods or towards [16] in reference to the output-only *operational modal analysis* (OMA) setting or more recently with respect to uncertainty quantification [17] (although the subject of uncertainty is not addressed in this work, see Section 6).

Briefly, the application of SSI requires formulation of the (block) Hankel matrices of the “past” and “future”. These are constructed through stacking time-delayed (lagged) versions of the measured time-series responses of the system. The number of lags chosen in the Hankel matrices defines the maximum order of the system which can be identified, where this maximum order is twice the maximum number of modes. After constructing these matrices,  $Y_p$  and  $Y_f$ , for the past and future respectively, the cross covariance can be computed for  $N$  datapoints as,

$$\Sigma = \begin{bmatrix} \Sigma_{pp} & \Sigma_{pf} \\ \Sigma_{fp} & \Sigma_{ff} \end{bmatrix} = \frac{1}{N} \begin{bmatrix} Y_p \\ Y_f \end{bmatrix} \begin{bmatrix} Y_p^T & Y_f^T \end{bmatrix} \quad (24)$$

This cross covariance is partitioned to consider the auto-covariance within the  $Y_p$  and  $Y_f$ , and the cross covariance between them. To perform SSI, a canonical correlation analysis between these two sets of variables is performed based on the covariances. Mathematically, a singular value decomposition (SVD) is taken of the normalised covariance,

$$USV^T = \Sigma_{pp}^{1/2} \Sigma_{pf} \Sigma_{ff}^{1/2}$$

The singular values (along the diagonal of  $S$ ) are sorted in descending order and, to select an appropriate model order, a number of the smallest of these values is discarded. If one wishes to construct the consistency (stabilisation) diagram, sequentially more and more values are discarded such that the model order is reducing and identification is performed at each of these orders. Since the computationally intensive stage of SSI is the formation of the Hankel matrices, covariances and SVD, this procedure for generating the consistency diagram is quite efficient. The truncated set of singular values can be denoted as  $\hat{S}$ , the associated columns of  $U$  and  $V$  are also removed with these truncated matrices as  $\hat{U}$ ,  $\hat{V}$ . Given that it can be shown that the cross covariance is the product of the observability and controlability matrices,  $\Sigma_{fp} = \mathcal{O}C$ , these can be recovered by:

$$\mathcal{O} = \Sigma_{ff}^{1/2} \hat{U} \hat{S}^{1/2} \quad (25a)$$

$$C^T = \Sigma_{pp}^{1/2} \hat{V} \hat{S}^{1/2} \quad (25b)$$

From  $\mathcal{O}$  the system matrices  $A$  and  $C$  of a linear dynamic state space model with states  $\vec{x}$ :

$$\begin{aligned} \vec{x}_{t+1} &= A\vec{x}_t + \vec{v}_t \\ \vec{y}_t &= C\vec{x}_t + \vec{w}_t \end{aligned} \quad (26)$$

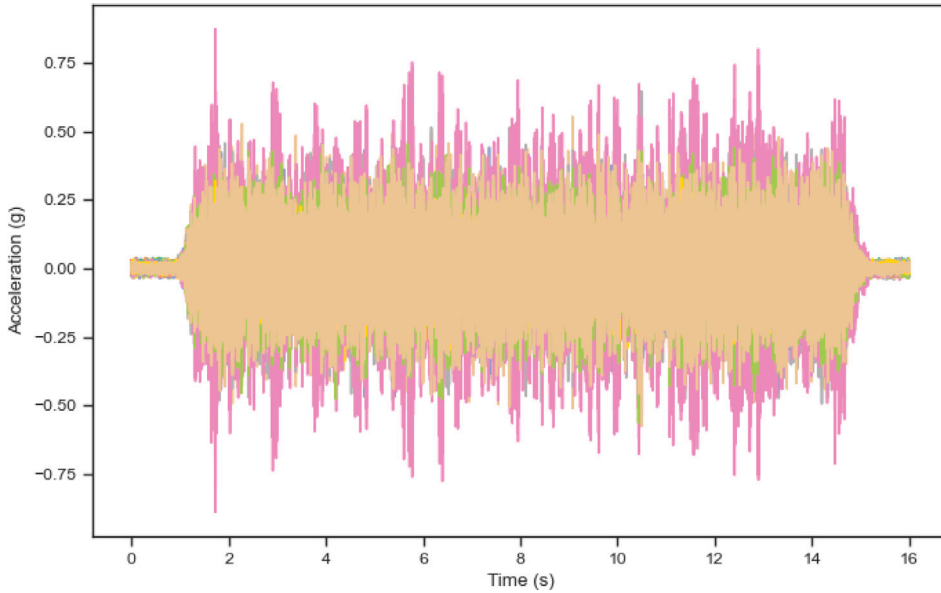


Fig. 9. Time series collected at each of the 55 sensor locations subject to closed-loop random phase excitation at 0.4 V.

where  $\bar{v}_t$  and  $\bar{w}_t$  are samples from multivariate Gaussian distributions with covariances  $Q$  and  $R$  respectively. Given that the observability matrix is constructed as,

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^K \end{bmatrix}$$

where  $K$  is the maximum model order considered, the  $C$  matrix of the system can be recovered as the first block row of this matrix and then the  $A$  matrix can be recovered via a least squares solution. The  $A$  and  $C$  matrices fully define the dynamics of the system and it is therefore possible to recover the modal properties. A common step in the SSI approach is to vary the size of the  $A$  and  $C$  matrices of the system and hence the model order by reducing the level of truncation of the singular values (usually in steps of two since the underlying system is assumed to be of a set of second order modes). Then, comparing the recovered modal properties a consistency<sup>7</sup> (also “stabilisation”) diagram is constructed based on the changes in these modal properties as the model order varies.

In Fig. 9, the time series from all accelerometers in the random phase amplitude test (excitation level 0.4 V) are plotted. Each burst signal represents 32,768 datapoints captured over a 16 s second interval. Unlike the RFP approach above, SSI is able to simultaneously incorporate data from all of the sensors at once. In order to speed up the computation of the SSI, the time series are decimated by a factor of 5 prior to the calculation of the Hankel matrices.

In order to evaluate the performance of the SSI, the singular-valued spectrum (SVS) [18] is calculated. It is useful to consider the SVS as a convenient visualisation as it is expected that all modes should be evident in only the first few singular values, whereas any given sensor may not have all modes visible (for example if a sensor is close to a nodal point in one of the modeshapes). In Fig. 10, the modal consistency diagram is plotted over the SVS. This diagram plots the predicted natural frequencies at each model order up to a maximum of 50. A standard approach in EMA is to interpret the ‘true’ modes as those that are ‘consistent’ in the diagram (that is they are present at each model order).

Given the somewhat arbitrary choice of which poles are considered ‘consistent’, the authors feel that a numerical comparison with the poles computed in the RFP method is not appropriate for these baseline identifications. However, as can be seen visually in Fig. 10, the SSI approach has identified the major resonances well (both in terms of correspondence to the SVS and the RFP results) although there are a significant number of non-physical modes especially in the higher order-models. Once again it is useful to interpret this result as a baseline against which more advanced output-only time domain methods might be considered.

## 5. Structural health monitoring

To equip engineering asset management teams with the information to make more optimal operation, maintenance and repair decisions for their high value infrastructure, the last several decades have seen a rise in developments of structural health monitoring

<sup>7</sup> Here the term *consistency* is preferred to *stabilisation* since there is not any link between the pole stability in terms of a requirement for the system to possess positive damping and the similarity of the modal properties between each model order. Therefore, the terminology could be ambiguous and give some implication about the system which may be misinterpreted.

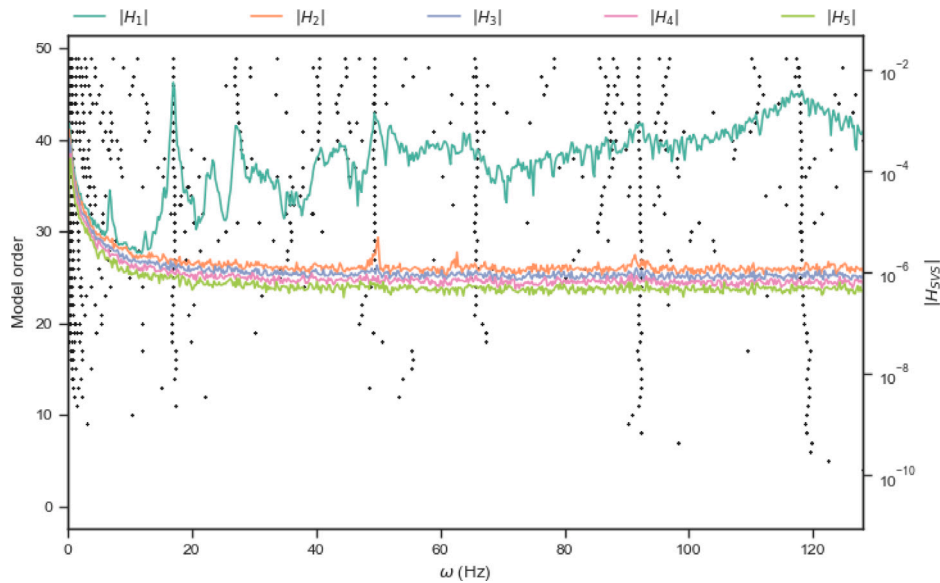


Fig. 10. SSI consistency diagram (black dots) compared to the first 5 components of the SVS (lines), subject to closed-loop random phase excitation at 0.4 V.

strategies to provide insight into the condition of structures using remote sensing [19]. A health monitoring system is defined as the implementation of a strategy that facilitates the real time assessment of the condition of a structure – often for civil infrastructure or high value assets such as aircraft – for ideally the span of the structure’s operational life. Broadly speaking, installing an SHM strategy first involves acquiring data relating to the dynamic behaviour of the structure, such as accelerometer measurements. Features sensitive to damage are then extracted, such as the natural frequencies of the structure, which are subsequently used as inputs to a statistical model that relate the features to some information pertaining to the damage state of the system.

The necessary complexity and sophistication of the model used can vary significantly, and is generally dependent on which level of the SHM hierarchy one is interested, of which there are five stages [20]:

1. Detection - is damage present?
2. Localisation - where is the damage?
3. Classification - what type of damage?
4. Assessment - how severe is the damage?
5. Prognosis - what is the remaining useful life of the structure?

As progress is made up the hierarchy, the complexity of the task increases, with classification, assessment and prognosis being particularly challenging. Regardless of the complexity of the SHM task that is being attempted, to transition monitoring algorithms from research to operational implementation, it is vital that the algorithms are validated on real world data to ensure that they are robust and reliable such that the risk of malfunction is inline with the guidelines of the relevant industry.

As such, a core area of research that this dataset can find value in is the testing of SHM methods. To that end, this section will demonstrate an overview of some basic SHM approaches and the challenges that arise in their application to the presented dataset. It should be reiterated that the purpose of the section is not to demonstrate state-of-the-art machine learning techniques, but instead apply what the authors would describe as “first principles machine learning for SHM” so that challenges in the dataset from an SHM perspective (damage detection, localisation and severity) may be clearly identified, as well as establishing a baseline performance from which SHM practitioner’s can look to improve.

In this paper, two SHM approaches are presented. As a supervised learning benchmark, a Gaussian mixture model (GMM) is fitted to the principal components of the natural frequencies. In the unsupervised case, a novelty detector based on the Mahalanobis distance is employed with damage sensitive spectral lines used as features. In total, 27 test runs are taken at amplitude levels between 0.4 V and 2.0 V for three levels of damage severity (added mass acting as pseudo-damage) and at three different locations along the wing. For undamaged condition data, the tests are taken from the burst-random amplitude ramp test series with excitation values between 0.4 V and 2.0 V. In total, this corresponds to 32 test series and 320 FRFs at each sensor.

### 5.1. Damage detection: Supervised learning

In order to demonstrate a first analysis in supervised SHM, a GMM is fitted to features arising from a modal identification as per the previous section. Natural frequencies are extracted using an RFP approach with ranges defined in Table 4. Overall, 18 natural

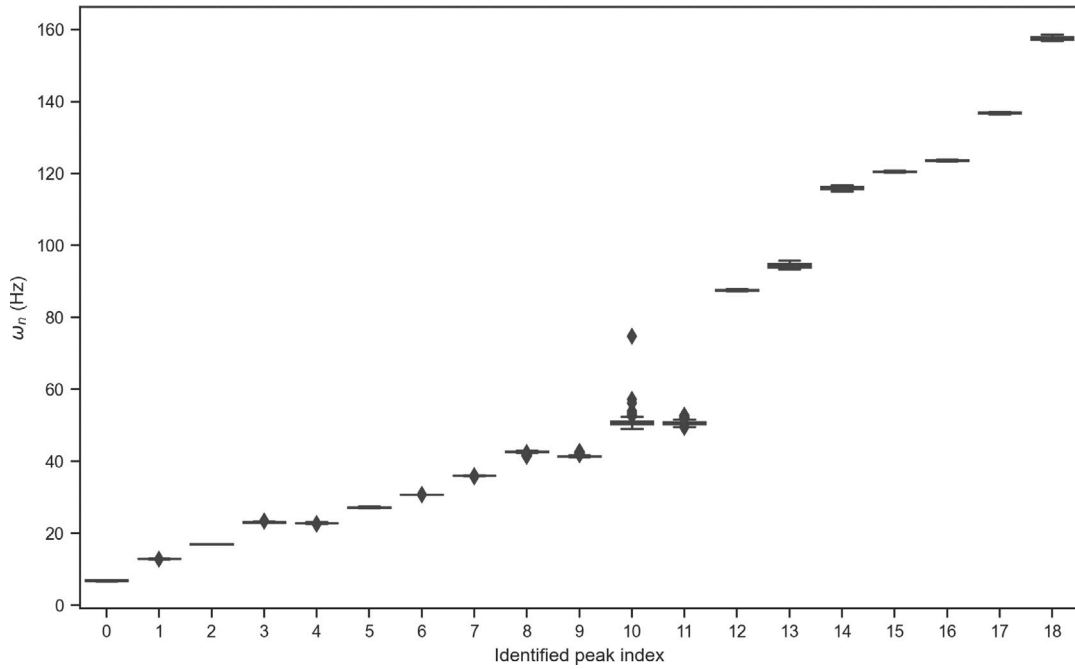


Fig. 11. Boxplot of the natural frequencies from the RFP analysis for each of the testing scenarios.

frequencies are derived for each sensor for each of the test runs. In order to reduce the variance in the identification process, the natural frequencies are averaged over each sensor location. Fig. 11 plots a boxplot of the resulting features. As can be seen in the figure, there are several outliers in the identification of the 10th peak, indicating that this peak has not been well identified. To avoid the variance in this mode dominating the features, the 10th natural frequency is removed from the feature set for an overall feature dimension of 18.

In order to further reduce the feature dimension, a principal component analysis (PCA) decomposition is taken on the natural frequencies. The first two principal components of the resulting features are plotted in Fig. 12. In the plot the samples are coloured by damage extent and the relative sizes of the markers represent the excitation level. Considering the data in Fig. 12, it is clear that the classes are not linearly separable and therefore the GMM classifier might not be expected to perform well. However it is still a useful result to consider the GMM as baseline for clustering performance such that advanced approaches might be compared to first analyses.

In order to classify the data, a GMM is trained. For training data, a subset of half of the training examples is used with balanced membership of each class in the training and testing sets. The training data are then used to empirically fit Gaussian distributions to each of the classes by evaluating the sample mean and covariances. The overall probability that an unseen example  $x^*$  has a label  $y^*$  corresponding to the  $k$ th class can be written,

$$p(y^*|X) = \sum_k \pi_k \mathcal{N}(\mu_k, \Sigma_k) \tag{27}$$

where  $X$  are the training data and,

$$\pi_k = \frac{n_k}{N} \tag{28}$$

the proportion of training samples in the  $k$ th class and,

$$\mu_k = \mathbb{E}[x_i] \quad \forall i \text{ s.t. } y_i = k \tag{29}$$

$$\Sigma_k = \text{cov}[x_i] \quad \forall i \text{ s.t. } y_i = k \tag{30}$$

are the maximum likelihood estimators of the class mean and covariances. The classification performance of the GMM in several classification tasks is plotted in Fig. 13, and confusion matrices are depicted in Fig. 14. In the figures, datum points used during training are marked with a grey cross. The predicted label is denoted by the centre of each marker and the true label is denoted by the colour of the border.

As can be seen in the figures, the performance of the classification is strong in the simplest damage detection case (i.e. binary classification) but the performance in other scenarios is poor. It appears that the simplistic model and feature set used here is insufficiently rich to separate the classes.

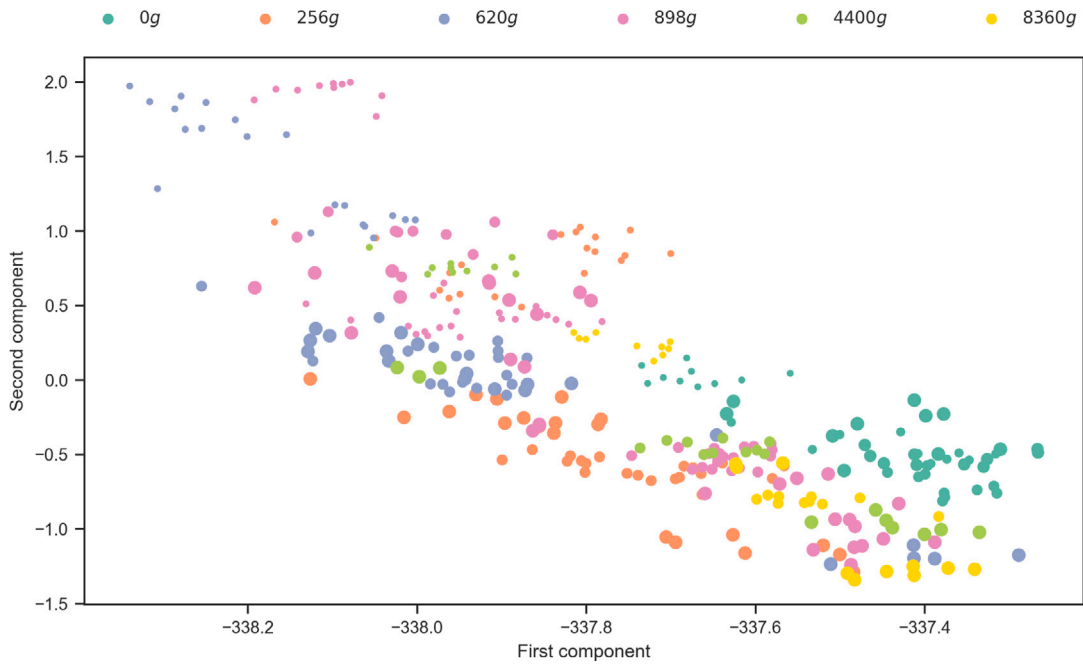


Fig. 12. First two principal components of the natural frequency features. In the plot colour denotes damage severity (added mass) and marker size denotes excitation level (larger markers for higher excitation levels).

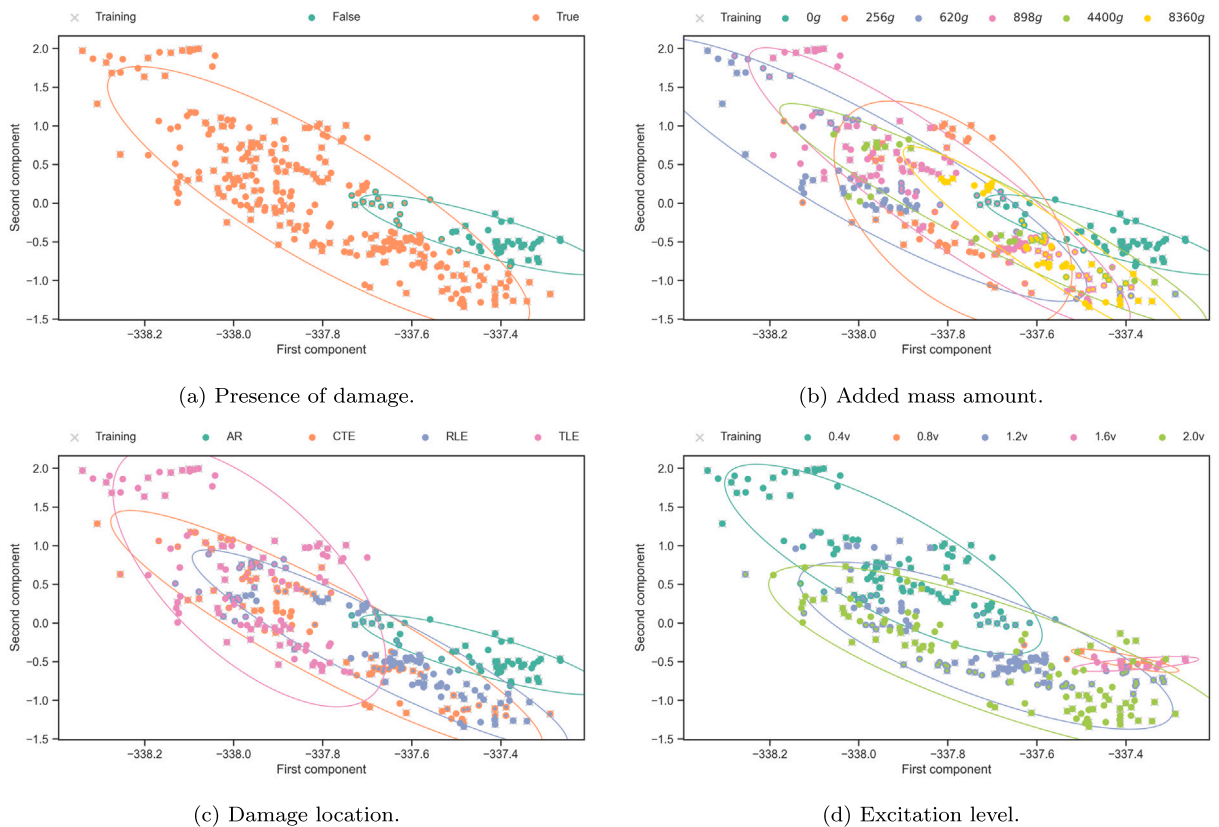


Fig. 13. Supervised GMM clustering results on the first two principal components of the natural frequency features, grouped by common characteristics of (a) damage presence, (b) added mass amount, (c) location of damage, and (d) excitation level.

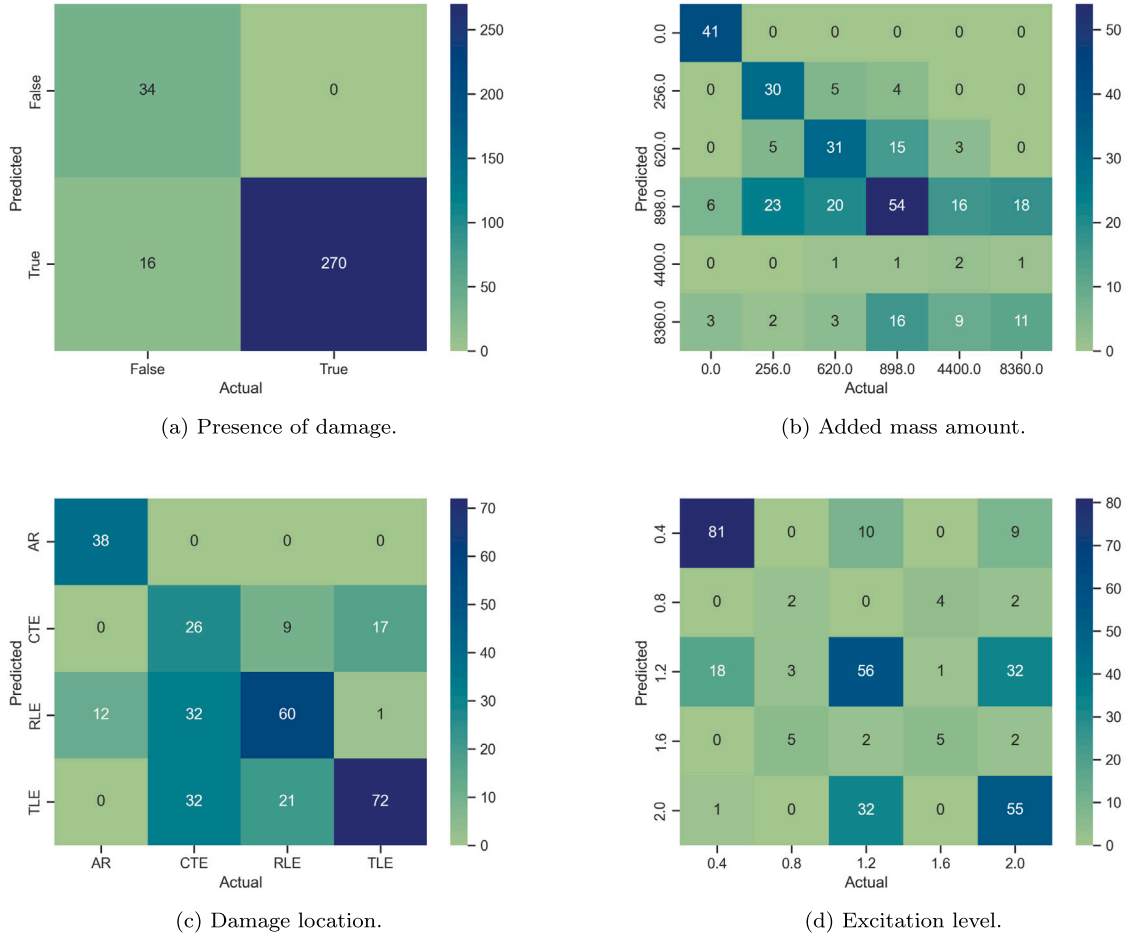


Fig. 14. Confusion matrices from the supervised GMM clustering, grouped by common characteristics of (a) damage presence, (b) added mass amount, (c) location of damage, and (d) excitation level.

### 5.2. Damage detection: Unsupervised learning

In SHM tasks, it is common for data to lack labels pertaining to the damage condition of the structure that the data are representative of, often leading to SHM practitioners undertaking unsupervised learning approaches. In this setting, it is common to perform outlier analysis, where incoming data from the structure in an unknown condition are compared by some distance metric to data generated by a baseline “normal” condition. If outliers are observed in the incoming data, then it is believed that damage may have developed in the structure.<sup>8</sup>

The principle of outlier detection is the computation of some distance metric between a test data point  $\mathbf{x}_i$  and the centre of the distribution that we believe the normal condition data to be drawn from. Assuming that the data in the baseline condition follow a normal distribution, a standard distance or *discordancy measure* is that of the Mahalanobis distance, defined as,

$$D = (\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \tag{31}$$

where  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are the mean and covariance of the normal condition observations. Monte Carlo sampling can then be used to determine a threshold for  $D$  over which a test point  $\mathbf{x}_i$  is believed to be an outlier.

Following the definition of a discordancy measure, all that remains to construct the outlier analysis is the selection of a feature sensitive to damage. As mentioned previously, natural frequencies are a common damage sensitive feature to use given that they will shift inline with local stiffness and mass changes that will be caused as a consequence of structural damage. To demonstrate, Fig. 15 plots the FRF extracted from the undamaged and damaged (898 g added mass at the tip) condition from sensor ULC-03

<sup>8</sup> Whether an outlier is truly representative of damage or not is strongly dependent on the quality of the feature extraction, and how sensitive to damage whilst insensitive to external influences such as operational and environmental fluctuations the chosen feature is.

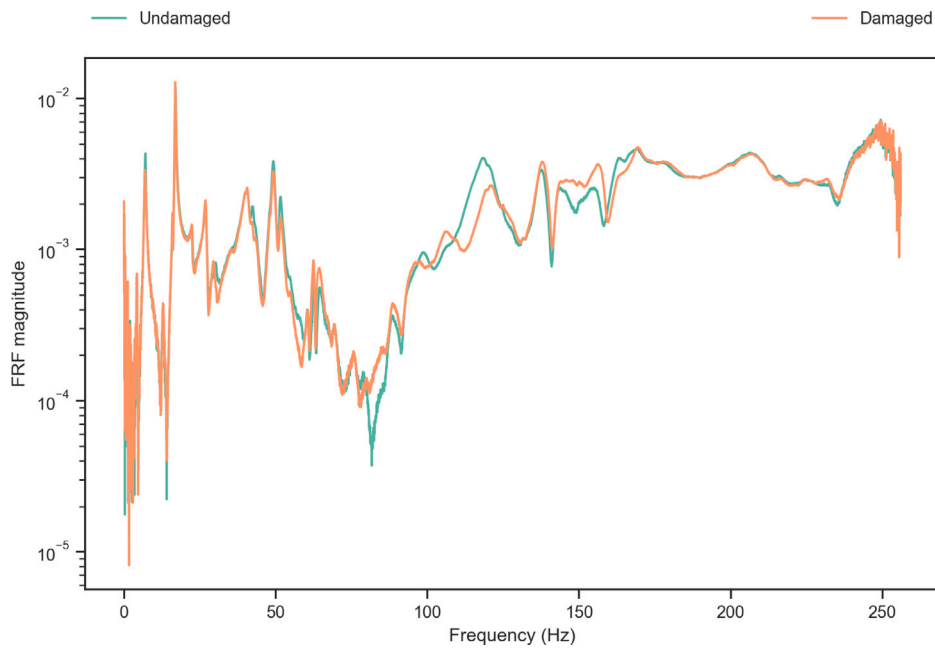


Fig. 15. Comparison between undamaged and damaged (898 g added mass at the tip) FRFs for sensor ULC-03, subject to burst-random excitation at 2.0 V.

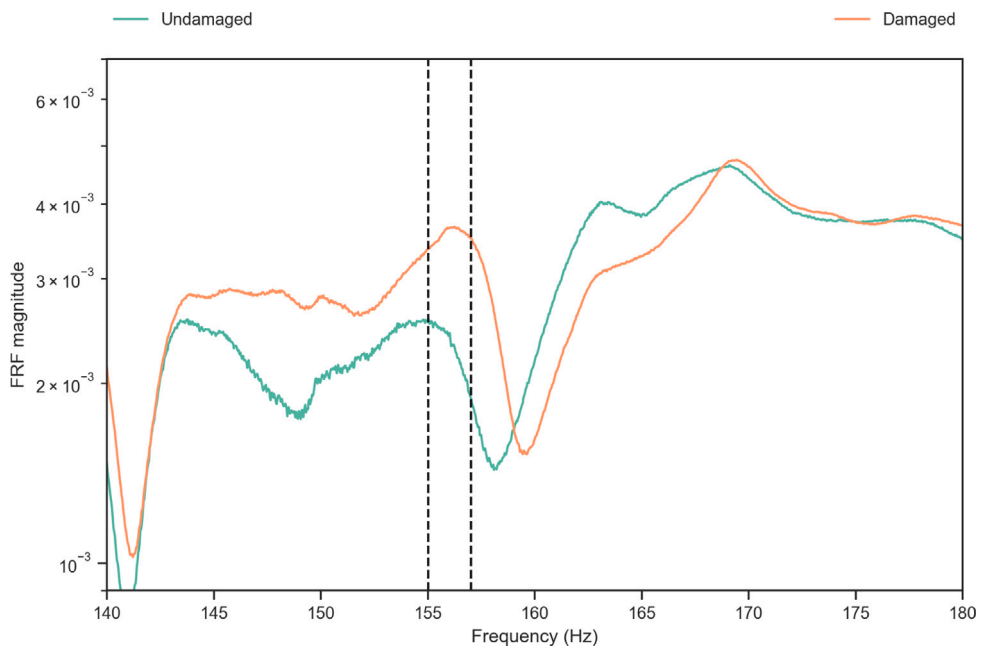
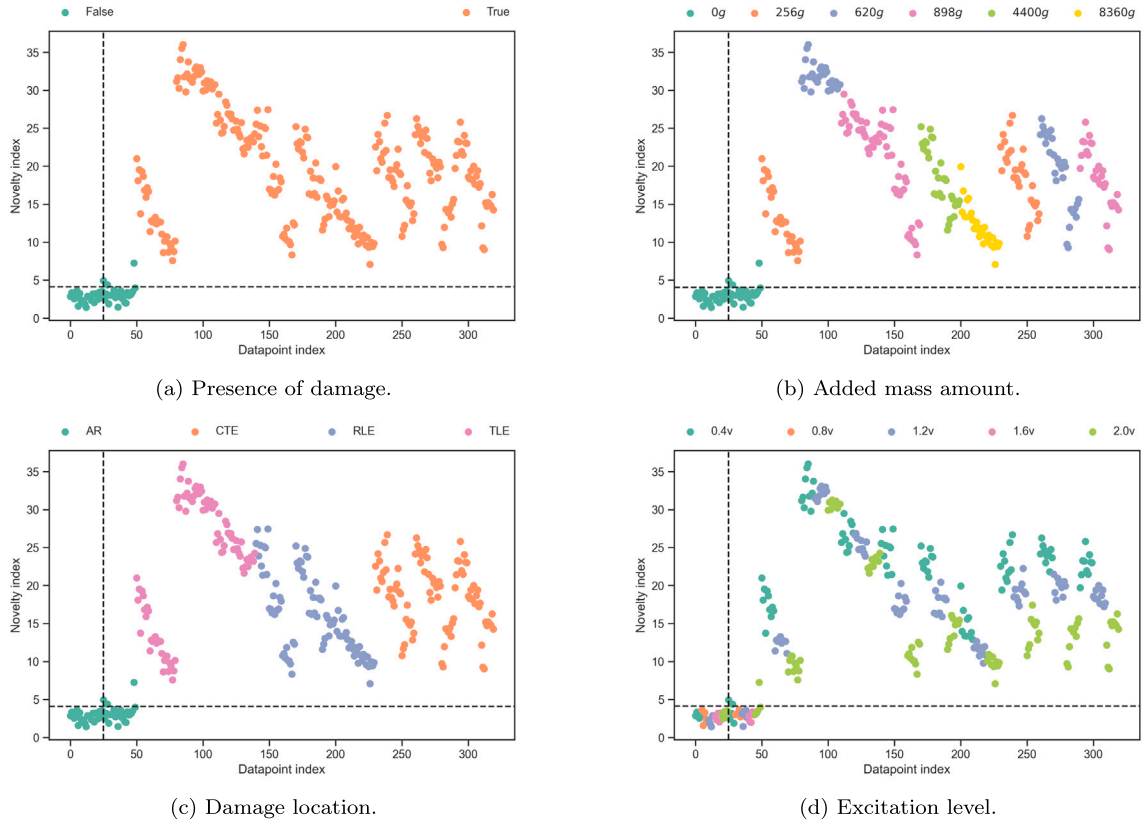


Fig. 16. Zoom into damage sensitive region of the FRFs. Vertical lines indicate the frequency range used as features for the novelty detection algorithm.

(sensor positions are in reference to Fig. 3) subject to burst random excitation at 2.0 V. It can be seen that there is a shift in both the peak frequencies and amplitudes in the damage case relative to the undamaged, confirming a suitable damage sensitive feature.

To mitigate the curse of dimensionality and avoiding taking the entire FRF as the feature vector, it is sensible to focus on just one peak of the FRF (e.g. around a natural frequency). Zooming in to the FRF around the natural frequency at 156 Hz, shown in Fig. 16, it can be seen that there is still a clear shift between the two damage cases. The feature vector is then taken to be 8 spectral lines between 155 and 157 Hz.

Having identified a suitable input feature for the outlier analysis, it is possible to use data that is known to have been collected from the structure in of an undamaged case to form a baseline for the novelty index, which sets the threshold for whether a datum



**Fig. 17.** Results from the novelty detection algorithm, trained on half of the normal condition data, grouped by common characteristics of (a) damage presence, (b) added mass amount, (c) location of damage, and (d) excitation level. Features are spectral lines between 155 and 157 Hz.

point is believed to be an outlier, and thus indicative of damage, or not. Beginning with just a comparison between data from an undamaged and damaged condition, Fig. 17(a) shows that the novelty detection is easily able to distinguish between the two damage classes, which includes varying levels of mass (pseudo-damage level), mass location and excitation, as shown in Figs. 17(b)–17(d). With outlier analysis, however, one is limited to a binary state classification, which in this case is whether the data is representative of a healthy or damaged (anomalous) structural condition. Attempts to attempt to classify on the basis of the novelty measure online will generally be futile, as shown in Figs. 17(b) and 17(c), where there exists no discernible separation in the novelty measure space for data across differences in mass amount and location.

## 6. Dataset challenges

The aim of this publication has been to layout a benchmark dataset for practitioners interested in structural dynamics, SHM and digital twins. To this end, a set of baseline analysis has been conducted to establish what is possible with well known techniques within the field. As such, it seems fitting to describe here what the remaining open challenges are when considering this dataset. It is the hope of the authors that this (admittedly non-exhaustive) list may prompt other members of the research community to apply their state-of-the-art methods.

### 6.1. Identification

The first category of challenges in this dataset is related to the system identification of a large and complex structure. From the perspective of the authors, the purpose of this system identification task is to provide a model which can:

1. Replicate the dynamic behaviour of the physical system through a mathematical model.
2. Provide some level of insight into the physical mechanisms which have led to the observed response of the system.
3. Quantify the uncertainty/confidence of the modeller, perhaps considering if the observed uncertainty is both epistemic and aleatoric.

Any given approach may target one or more of these objectives and, ideally, all three. It is also worth noting that the analysis carried out should be usable in such a way that it informs further decisions, e.g. the results of a linear modal system identification may be used as features within an SHM scheme.

Considering the methodologies shown so far in this publication which are intended as a baseline, it is worth noting some of the major assumptions made thus far which (the authors would argue) should and can be challenged in future work. The analysis thus far has:

- Considered only the *low* frequency regime, this has neglected *local modes*, e.g. the response of individual panels on the wing which may behave as thin plates.
- Relied on there being a *modal* decomposition, such that the observed transfer function of the system is a superposition of linear modes.
- Extending this modal assumption, all analysis has considered the system linear, despite some observed sensitivity in the FRF to excitation amplitude.
- Considered the system to be completely time-invariant across a testing campaign of several weeks.
- The aircraft is treated directly in a reduced order form as a relatively small number of ordinary differential equations (in the decoupled modal space), compared to the overall complexity of the structure.
- All analysis shown in this paper has avoided a formal treatment or quantification of uncertainty, instead only point estimates are given.

Removing or replacing these assumptions and simplifications would be valuable topics for further research. Some potentially interesting avenues of investigation include: consideration of the best manner in which to build the low order model of the system; verification of the identification results by other algorithms; deeper investigation into the potential nonlinearity in the system and its impact.

## 6.2. Prediction

In addition to the tasks listed above which consider the identification of the system, it is also useful to imagine scenarios in which the practitioner may wish to predict certain aspects of the system behaviour. The extension of the identification procedures to prediction has not been shown in this work but should be an area of investigation in the future. The value of such predictions to the engineer would be to allow replication of (as yet) unseen behaviours; to explore the response of the system in extreme events or critical scenarios; the ability to infer the response of the system at locations which are not measurable in operation (i.e. virtual sensing); to allow validation of the identified models against other collected dataset; to provide an objective means for comparison between models (especially if looking at the relative performance of physics-based and data-driven models).

It is instructive to further consider two of these tasks in greater detail. For the problem of virtual sensing which can be considered to reproducing data which would have been measured at some dense (by some measure) set of sensors on the basis of observations at a sparse set of sensors. There exist several methodologies in the literature which provide promising approaches for the virtual sensing problem. Applying these to the Hawk data presented here in future work would form a further valuable benchmark on this system. In addition to this there are several outstanding questions which are relevant when considering the problem of virtual sensing. Firstly, one could explore what would be the minimal set of sensors required to provide some bounded uncertainty/error in the locations which are being reconstructed. Secondly, it is important to explore which sensors are (in some sense) optimal to perform the reconstruction and how the user may choose this subset. Thirdly, the verification and validation of virtual sensing systems especially in terms of their reconstruction under extrapolation, e.g. when operating in a larger amplitude excitation regime.

Another prediction task of interest, for which this data could serve as a benchmark, is that of load estimation. In many settings of interest, unlike the scenario encountered in this laboratory test, it may be impractical to acquire measurements of the loads applied to the structure. In aerospace applications, it is very challenging to acquire the aerodynamic loads on the wings during flight. Similar scenarios are encountered in wind turbine or bridge applications. In all such cases, the fluid structure interaction forces are typically unavailable. Owing to this difficulty, in recent years, there has been some attention given to estimation of the loads alongside other properties of interest in the system. This is referred to as either load estimation/reconstruction or as joint input-state(-parameter) estimation [21]. In the context of using in service measurements for asset management, knowledge regarding the loads is highly valuable as the loads are intrinsically linked to the fatigue life of the structure. In practice the estimation of these loads is complicated by noise on the measurements, uncertainty regarding the dynamic system itself and a lack of full state information/observation, i.e. only sparse measurements are available.

## 6.3. Evaluation

In addition to the understanding and reconstruction of the dynamic system, it is of interest to consider the follow-on tasks which this form of testing/monitoring will enable. One example of a task for which dynamics might be studied (and demonstrated in this paper) is that of SHM which enables automated asset management through monitoring. This desire for more tightly integrated modelling, incorporation of data and asset management may also be characterised, more recently, using the language of digital twins [22,23]. In this paper tasks related to the lower levels of Rytter's hierarchy have been shown. For problems of damage detection, it has been seen in this work how (given insight about the form of the damage or some prior knowledge of sensitive features) it is possible to build a good quality model. However, doing this in the more representative context of only having

access to the outputs or without good prior knowledge of the features may yet pose a challenge. What has been clear is that the more challenging tasks in the hierarchy of localisation, classification and extent remain very challenging based on the features extracted. This difficulty is present even when considering the high level of prior knowledge used in this study. Additionally, the interaction of the SHM procedure with, for example, the nonlinearity or other confounding influences remains to be explored. The topic prognosis has not been approached in this work and, although different extents of (proxy) damage are used, this would not easily be accomplished with this benchmark dataset. Some additional tasks which may add high value to the SHM community would be to attempt to develop a generative model of damage which would allow classifiers to be trained without examples of the previous damage cases. For example, this could be to reproduce the effect of damage at the trailing edge of the wing from the example data collected at the tip.

This section has attempted to provide a non-exhaustive list of future avenues of investigation for which this dataset could prove a useful benchmark. It is the belief of the authors that the presented data represents a challenging full-scale example of some of the open questions within the field of structural dynamics and SHM. Hopefully, this discussion will stimulate further investigation of this dataset and subsequent research.

## 7. Conclusion

In this work, a new, large-scale structural dynamics benchmark dataset is introduced. Details of the structure, experimental campaign, acquisition and control systems are presented and the data is made freely available. As well as the supplied data, this work also presents a first analysis of several dataset challenges using standard approaches in structural dynamics. For the task of system identification, a *rational fraction decomposition* is applied in the frequency domain, and this is compared to an output only time domain identification in *stochastic subspace identification*.

For the task of damage identification, a *supervised* machine learning approach based on *Gaussian mixture models* on the principal coordinates of the natural frequencies is considered, with an *outlier analysis* presented as a first benchmark as an *unsupervised* method.

The analyses and results presented here are not groundbreaking and are intended to serve as a baseline, against which, more advanced methods might be compared. Some challenges which the dataset presents have been discussed, with the view to inspire readers with motivation for new approaches to structural dynamics analysis.

## CRedit authorship contribution statement

**Marcus Haywood-Alexander:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Robin S. Mills:** Software, Resources, Methodology, Investigation, Data curation, Conceptualization. **Max D. Champneys:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Data curation. **Matthew R. Jones:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Data curation. **Matthew S. Bonney:** Data curation. **David Wagg:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Timothy J. Rogers:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

## 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.

## Data availability

All data and documentation as well as a python API for accessing and interacting with the data are openly available at <https://doi.org/10.15131/shef.data.22710040.v1>.

## Acknowledgements

The authors of the paper are thankful for the support of The Alan Turing Institute under the Data-Centric Engineering programme, UK, specifically through the project Digital twins for high-value engineering applications (DTHIVE), as well as the EPSRC through grants EP/R006768/1, EP/W002140/1 and EP/S001565/1. The experimental work made use of The Laboratory for Verification and Validation (LVV), which was funded by the EPSRC, UK (grant numbers EP/J013714/1 and EP/N010884/1),

The authors are also extremely grateful for the stimulating and insightful discussions during an Isaac Newton Institute for Mathematical Sciences (UK) data-driven deep-dive “From physics-based to data-driven assessment of structures” in February 2023, which was supported through grant no EP/R014604/1. Finally, the authors extend their thanks to Aidan Hughes, Brandon O’Connell, Daniel Brennan, Daniel Pitchforth, George Tsialiamanis, Jack Poole, Jacques Mclean, Joe Longbottom, Tristan Gowdridge for assistance in collecting the data.

## References

- [1] R.R. Craig Jr., A.J. Kurdila, *Fundamentals of Structural Dynamics*, John Wiley & Sons, 2006.
- [2] D.H. Hodges, G.A. Pierce, *Introduction to Structural Dynamics and Aeroelasticity*, vol. 15, Cambridge University Press, 2011.
- [3] F.M. Hemez, S.W. Doebbling, Review and assessment of model updating for non-linear, transient dynamics, *Mech. Syst. Signal Process.* 15 (1) (2001) 45–74.
- [4] G. Kerschen, K. Worden, A.F. Vakakis, J.-C. Golinval, Past, present and future of nonlinear system identification in structural dynamics, *Mech. Syst. Signal Process.* 20 (3) (2006) 505–592.
- [5] K. Worden, *Nonlinearity in Structural Dynamics: Detection, Identification and Modelling*, CRC Press, 2019.
- [6] T.J. Rogers, G.R. Holmes, E. Cross, K. Worden, On a grey box modelling framework for nonlinear system identification, in: *Special Topics in Structural Dynamics, Volume 6: Proceedings of the 35th IMAC, A Conference and Exposition on Structural Dynamics 2017*, Springer, 2017, pp. 167–178.
- [7] K. Worden, R. Barthorpe, E. Cross, N. Dervilis, G. Holmes, G. Manson, T. Rogers, On evolutionary system identification with applications to nonlinear benchmarks, *Mech. Syst. Signal Process.* 112 (2018) 194–232.
- [8] R. Fuentes, R. Nayek, P. Gardner, N. Dervilis, T. Rogers, K. Worden, E. Cross, Equation discovery for nonlinear dynamical systems: A Bayesian viewpoint, *Mech. Syst. Signal Process.* 154 (2021) 107528.
- [9] H. Fraser-Mitchell, The Hawk story, *J. Aeronaut. Hist.* (2011).
- [10] E. Papatheou, G. Manson, R.J. Barthorpe, K. Worden, The use of pseudo-faults for novelty detection in SHM, *J. Sound Vib.* 329 (12) (2010) 2349–2366.
- [11] M. Haywood-Alexander, R. Mills, M. Champneys, M. Bonney, D. Wagg, T. Rogers, BAE T1A Hawk Starboard Wing Modal Test, 2023, <http://dx.doi.org/10.15131/shef.data.22710040.v1>, URL [https://figshare.shef.ac.uk/articles/dataset/BAE\\_T1A\\_Hawk\\_Starboard\\_Wing\\_Modal\\_Test/22710040](https://figshare.shef.ac.uk/articles/dataset/BAE_T1A_Hawk_Starboard_Wing_Modal_Test/22710040).
- [12] D.J. Ewins, *Modal Testing: Theory, Practice and Application*, John Wiley & Sons, 2009.
- [13] M. Richardson, D. Formenti, Global curve fitting of frequency response measurements using the rational fraction polynomial method, in: *Proceeding of 3rd IMAC*, 1985, pp. 390–397.
- [14] T. Katayama, *Subspace Methods for System Identification*, vol. 1, Springer, London, 2005.
- [15] P. Van Overschee, B. De Moor, N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems, *Automatica* 30 (1) (1994) 75–93, [http://dx.doi.org/10.1016/0005-1098\(94\)90230-5](http://dx.doi.org/10.1016/0005-1098(94)90230-5).
- [16] B. Peeters, G. De Roeck, Stochastic system identification for operational modal analysis: a review, *J. Dyn. Syst. Meas. Control* 123 (4) (2001) 659–667.
- [17] E. Reynnders, K. Maes, G. Lombaert, G. De Roeck, Uncertainty quantification in operational modal analysis with stochastic subspace identification: Validation and applications, *Mech. Syst. Signal Process.* 66 (2016) 13–30.
- [18] H. Hassani, Singular spectrum analysis: Methodology and comparison, *Data Sci. J.* 5 (2) (2007) 239–257, [http://dx.doi.org/10.6339/jds.2007.05\(2\).396](http://dx.doi.org/10.6339/jds.2007.05(2).396).
- [19] C.R. Farrar, K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*, John Wiley & Sons, 2012.
- [20] K. Worden, J.M. Dulieu-Barton, An overview of intelligent fault detection in systems and structures, *Struct. Health Monit.* 3 (1) (2004) 85–98.
- [21] S. Vettori, Development of Virtual Sensing Techniques for Dynamic Virtualization (Doctoral thesis), ETH Zurich, Zurich, 2023, <http://dx.doi.org/10.3929/ethz-b-000623754>.
- [22] D.J. Wagg, K. Worden, R.J. Barthorpe, P. Gardner, Digital Twins: State-of-the-Art and Future Directions for Modeling and Simulation in Engineering Dynamics Applications, *ASCE-ASME J. Risk Uncertain. Eng. Syst.* B 6 (3) (2020) 030901, <http://dx.doi.org/10.1115/1.4046739>, arXiv:[https://asmedigitalcollection.asme.org/risk/article-pdf/6/3/030901/6534130/risk\\_006\\_03\\_030901.pdf](https://asmedigitalcollection.asme.org/risk/article-pdf/6/3/030901/6534130/risk_006_03_030901.pdf).
- [23] P. Gardner, M. Dal Borgo, V. Ruffini, A.J. Hughes, Y. Zhu, D.J. Wagg, Towards the development of an operational digital twin, *Vibration* 3 (3) (2020) 235–265.