Preprints.org

Article

Not peer-reviewed version

Optimization of sensor placement for modal testing using machine learning

Todd W Kelmar , Maria Chierichetti * , Fatemeh Davoudi *

Posted Date: 11 March 2024

doi: 10.20944/preprints202403.0604.v1

Keywords: modal testing; sensor placement; machine learning; finite element method; beam analysis

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Optimization of Sensor Placement for Modal Testing Using Machine Learning

Todd Kelmar [1], Maria Chierichetti [1,*] and Fatemeh Davoudi [2]

[1] San Jose' State University; tkelmar@gmail.com
[2] Machine Learning & Safety Analytics Lab; Santa Clara University; fatemeh.davoudi@scu.edu
* Correspondence: maria.chierichetti@sjsu.edu

Featured Application: This study introduces an innovative approach for optimizing sensor placement in modal testing through application of machine learning with enhanced efficiency and precision.

Abstract: Modal testing is a common step in aerostructure design, serving to validate the predicted natural frequencies and mode shapes obtained through computational methods. The strategic placement of sensors during testing is crucial to accurately measuring the intended natural frequencies. However, conventional methodologies for sensor placement are often time-consuming and involve iterative processes. This study explores the potential of machine learning techniques to enhance sensor selection methodologies. Three machine learning-based approaches are introduced and assessed, comparing their efficiency with established techniques. The evaluation of these methodologies is conducted using a numerical model of a beam to simulate real-world scenarios. The results offer insights into the efficacy of machine learning in optimizing sensor placement, presenting an innovative perspective on enhancing the efficiency and precision of modal testing procedures in aerostructure design.

Keywords: modal testing; sensor placement; machine learning; finite element method; beam analysis

1. Introduction

Mechanical structures are subject to vibrations either internal, such as engine vibration, external, such as turbulence, or a combination of both. Characterizing the behavior of a system under vibration or other dynamic forces is therefore crucial to good engineering design, and it is pivotal to the aerospace and automotive industry [1,2].

While the accessibility and speed of modern computers and FEA solvers may seem to obviate the need for physical modal testing, the results are only as accurate as the model being tested [3]. The results of modal testing can be compared to those of the theoretical model and used to establish if the model accurately describes the structure being analyzed [1]. Additional uses of modal testing include creating mathematical models of structures for integration into other analyses or developing models for structural health monitoring [1].

Modal analysis can be conducted on data acquired in laboratory conditions or from data acquired while the structure is in regular use [4–6]. In modal testing, sensors placed on the structure being tested—typically accelerometers and/or strain gauges—are measured to record their response to an excitation. The input can be provided by a modal shaker—a device which takes a signal as an input and applies that signal to the structure under test— or a modal hammer, where an impact is made against the structure to represent an instantaneous excitation [4]. In more complex tests or for large structures, multiple modal shakers may be used to induce a measurable excitation in the structure [1]. The outputs of the sensors are then post processed, with the exact methodology dependent on the excitation signal. These data may then be analyzed with a frequency response function in order to derive the natural frequencies and mode shapes of the structure under test.

© 2024 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

Placing the sensors on the structure must be done with care, as it can substantially influence the results of a modal test. Large numbers of sensors increase the cost of testing in both equipment and labor required to set up the test. As modal testing is principally concerned with structural dynamics, an ideal sensor selection would result in the least number of sensors where each sensor's individual contribution to the analysis is greatest [7]. The goal of optimizing sensor placement is to determine the most information about the structure's behavior while minimizing the required number of sensors [7].

In early modal testing, sensor placement relied on engineering judgement and institutional knowledge derived from fundamentals of vibration [8]. While this may still be used in certain situations, such as with a well understood structure or for simple geometries, novel structures present difficulties for this approach. Additionally, tight timelines due to budget constraints or limited access to testing facilities reduce the time available to refine sensor placement during testing [3]. As such, determining efficient methodology for sensor placement has real implications for increasing efficiency. As a result, several methodologies have been developed to assist engineers in determining appropriate sensor placement for modal testing and structural health monitoring.

In general, methods for optimal sensor placement can generally be divided into two broad categories, model-based methods and data-driven methods. Model-based methodologies define the placement of sensors based on information derived from a numerical model, such as a finite element model or a multibody model. Specifically in modal testing, existing model-based methodologies used for sensor placement include the Effective Independence Method (EIM) and Iterative Residual Kinetic Energy approach (IRKE) [9]. Other techniques, such as those using information entropy, have also been developed [4]. A brief overview of these techniques is provided in section 2.

Data-driven sensor placement strategies have also been recently proposed to extract the oscillatory characteristic directly from experimental measurements without requiring the need of a numerical model [10–13]. Zhang at el. [10] proposed a sensor placement approach that relies on a repetition of in-situ trial measurements on bridges with different sensor positioning in order to avoid the need to rely on finite element data. The measured in-situ data is used to train a Recurrent Gaussian Process Regression until a sufficient number of sensors is identified. However, this approach requires multiple experimental trials of sensor configurations, a costly approach for complex and large systems such as aerospace structures. Similarly, Suryanarayana et al. [12] employs a data-driven approach for optimal sensor placement of a multi-zone building. This method requires that experiments with a large number of sensors are initially completed to collect the data necessary to apply the sensor placement approach. These data-driven methods often rely on operational modal analysis to extract the modes of the system during operations. Sashittal et al. [13] applies data-driven sensor placement to the observation of fluid flows. To the authors' knowledge, no data-driven methodologies have been proposed for modal analysis.

This paper proposes a non-iterative model-based approach for optimal sensor placement in modal testing. Finite element models are generally available for large aerospace structures and can therefore be used to identify the number and positioning of sensors in the structure. The approach is based on machine learning techniques to avoid the need for iteration.

Machine learning (ML) techniques are a promising approach for determining sensor placement in modal analysis. In supervised machine learning, an input dataset is provided consisting of both the input data and the output. In this case, the input would be data derived from a finite element model and the output would be the mode shapes and natural frequencies. Based on this information, the model is then trained to be able to predict outputs based on new input data. As many of the previously discussed methods for sensor placement are iterative approaches, the problem of solving for sensor placement seems to be one to which machine learning is well suited [14–16].

This paper presents a novel methodology for sensor placement in modal analysis using random forest techniques. An initial application of the proposed approach is discussed in Kelmar et al. [17]. The paper is organized as follows: initially, a review of traditional methodologies currently used for sensor placement in modal analysis is presented. Then, the proposed machine learning approach is

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

presented and validated for a vibrating beam. Results of the proposed approach are then compared to the results obtained using one of the traditional methodologies.

2. A Review of Traditional Sensor Placement Techniques for Modal Testing

Several methodologies have been developed to assist engineers in determining appropriate sensor placement for modal testing and structural health monitoring based on a finite element model. Some of the current existing methodologies include the Effective Independence Method (EIM), the Mass Weighted Effective Independence Method (MEIM) and the Residual Kinetic Energy approach (RKE) [16,18].

2.1. Effective Independence Method (EIM)

The Effective Independence Method, also known as the effective independence algorithm, is one of the most popular sensor placement techniques, and it bases its analysis on the sum of the diagonal terms of the Fisher information matrix. The Fisher information matrix is constructed using the modal characteristics extracted by a finite element model. It is an iterative method that evaluates the contribution of all possible sensor locations (i.e. the nodes of the finite element method) to the linear independence of the mode shapes. Sensors with small contributions to the linear independence are progressively eliminated until the desired number of sensors remains. This final set of sensors maximizes the sum of the diagonal and the condition number of the Fisher Information matrix.

The method begins with a set of target mode shapes that encompass the set of candidate sensor locations generally derived from an FE model of the structure under analysis [19,20]. The algorithm attempts to predict the independence of each node based on the expected measured mode shape, with higher values indicating increased independence [19]. In this method, it is required to know both the expected mode shape as well as the location of candidate sensors, and it is therefore well suited to be used when a finite element model is available. These candidate locations are then ranked according to the algorithm, removing the lowest ranking sensor and recalculating. As potential locations are eliminated, the relative independence of the remaining solutions increases, and the process is repeated until the required number of sensor locations is reached.

One of the challenges of EIM is that the optimal number of sensors must be defined a priori and potential locations of the sensors must be available [21]. A very fine grid of the finite element model allows the user to analyze all the possible locations of sensors but the method then becomes very time-consuming due to the iterative nature of EIM. Some research also suggests that more optimal results may be produced compared to kinetic energy methods, although at the cost of less ability to measure unexpected modes [19,22]. Additionally, the EIM approach does not account for unknown modes that may occur in the real world but do not appear in FEA. At the same time, if there are specific modes in the FEA results that are of more interest, EIM can provide targeted sensor selection for those modes which may require less sensors than needed to capture the full behavior of the structure.

The EIM is derived by Kammer et al [22] and is based on the concept that each sensor output can be represented as a linear combination of the mode shapes of the system. An effective independence score is calculated for each possible sensor location using the modal content of the numerical model. The higher the effective independence score of a candidate sensor location, the more important that location is for calculating the independence of the mode shapes. Therefore, sensor locations with the lowest value are eliminated, and the effective independence score is then recalculated from the subset of candidate locations. The process is complete when the desired number of sensors is reached or when all remaining sensor locations have similar effective independence values [19].

2.2. Mass Weighted Effective Independence (MEIM)

A drawback to EIM is that it selects sensors only considering the contribution to the linear independence of the mode shapes and neglects their orthogonality constraints [23]. When a mass-

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

weighted approach is used, such as the Mass Weighted Effective Independence (MEIM), modes shapes that contribute the least to self-orthogonality are removed in each iteration as opposed to purely focusing on linear independence when selecting features. Cross-orthogonality checks are used to determine how analytical and empirical modal testing results correlate.

One of the drawbacks, however, is that the Mass Weighted Effective Independence requires the decomposition of the mass matrix. The use of the Guyan reduced mass matrix appears to result in the best performance computationally as well as producing the most optimized output with respect to other mass weighting techniques [23].

2.3. Residual Kinetic Energy Method (RKE)

The RKE method is a technique that provides information on the sensor location that exhibit the maximum response for each mode shape. It is commonly used by NASA to determine sensor placement for modal testing based on detailed FEA models [24]. It ensures that the residual kinetic energy is minimized in all degrees of freedom and modes under consideration. When this is computed, DOFs with high residual kinetic energy indicate that additional refinement is needed in order to measure the corresponding degree of freedom in a given mode. After another sensor is added to cover that degree of freedom, the residual kinetic energy is recomputed. This process is repeated until the solution is suitably orthogonal [25].

The RKE method ensures all recorded modes are fully orthogonal and therefore independent from each other [16]. An RKE matrix is defined in which each column represents the contribution of each degree of freedom to a specific mode. The sum of the contribution of each degree of freedom is 1. Sensors should be placed at the location of the nodes with higher RKE values. The RKE matrix column will add to much less than 1 if that mode is already appropriately instrumented [16]. By iterating through this matrix, you can determine the location where sensors should be placed as well as the minimum number of sensor locations.

This methodology works well when applied to existing analysis points to identify additional degrees of freedom that are undermeasured by the initial sensor placement, and has been adopted by NASA and others to meet NASA and Department of Defense standards for modal testing.

3. Machine Learning Approach for Sensor Selection

Machine learning (ML) techniques are a promising approach for determining sensor placement in modal analysis. Machine learning techniques are able to determine a non-deterministic relationship between an output quantity and a large number of input quantities—called "features"—on which the output depends, through an initial process called "training". A large number of features are usually defined in a machine learning database, resulting in high computational costs. In an effort to reduce the computational costs of the training procedures, and to identify the most important factors that contribute to the desired output, several approaches have been defined to identify the most important features of the database. Examples of such approaches are the SelectKBest algorithm [26,27], the Random Forest feature importance approach [28,29], and the Principal Component Analysis [30,31]. This paper focuses on the use of the Random Forest feature importance approach, as discussed in the following section.

3.1. Random Forest Feature Selection Approach for Sensor Placement

The random forest (RF) feature selection approach was selected for sensor placement based on its promising performance in existing sensor selection applications [32]. The RF is a learning method for classification and regression that belongs to the CART family (CART: Classification and Regression Trees). It is considered an averaging ensemble method because it combines the results from multiple estimators and averages the predicted results to reduce variance.

During the training process, the random forest algorithm constructs a multitude of decision trees with a predicted estimation of the output variables; then, the outputs of all trees are aggregated and the algorithm returns the average prediction of the individual decision trees. This aggregation

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

process is called bagging method, and highly reduces the variance and the prediction bias—either underestimation or overestimation—of the output target, reducing overfitting [28]. In addition to the randomness introduced by varying the input data for each DT, random perturbations in the DTs are also introduced. Each decision tree constructed by the algorithm is composed by internal nodes and leaves; in each internal node, all the features are used to make decision on how to binary split the data set further based on a defined criterion, such as the Gini impurity or variance reduction parameters. This criterion measures how each feature decreases the impurity of the split at each node. For each feature, it is then possible to determine how on average it decreases the impurity of all trees in the forest, which becomes a measure of the feature importance [33]. In this paper, the Gini index will be used to determine the feature importance. Once the algorithm calculates the feature importance for each input variable, a bar graph can be obtained to determine the most important features. Additionally, the R squared value (R2) and mean squared error (MSE) can be used to evaluate performance of the approach.

Initially, a Random Forest model is constructed using all the features available in the dataset. The Random Forest algorithm computes the feature importance of each input variable as it maps to the output variable. The regressor attributes a feature importance value that ranges between 0 and 1 to each input variable in the model. This value represents how much variance in the output is represented by each input variable. The sum of all feature importance of the model is 1. By selecting the inputs with the largest values of feature importance we can determine which inputs are most valuable in representing the output. A number of inputs should be selected such that a sufficient percentage of the variance of the data is represented.

Conceptually, if a model has 4 input variables called *a, b, c, d,* the RF regressor attributes a feature importance value to each input. In Figure 1, input *b* has a feature importance of 0.6, input *d* has a feature importance of 0.2, input *a* has a feature importance of 0.15, and input *c* has a feature importance of 0.05. Input *b* therefore is the most important feature and represents 60% of the variance in the data. Input *d* is the second most important feature and represents 20% of the variance in the data. Inputs *b* and *d* therefore together represent 80% of the variance in the data and could be used as a reduced model of the system.

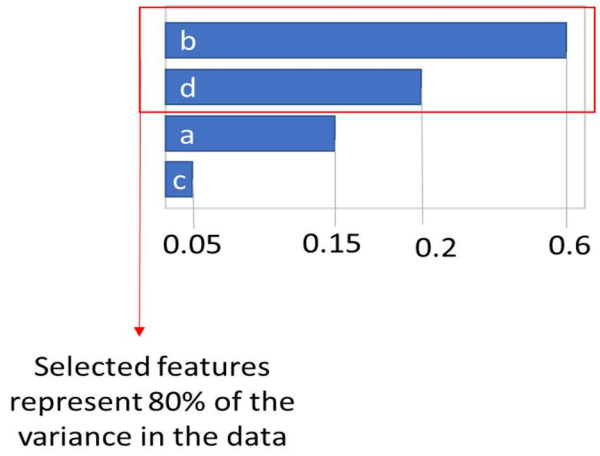


Figure 1. Conceptual representation of Random Forest feature selection approach.

This concept can be applied to sensor placement and selection by defining as input variables all the possible locations and types of sensors in the system [34]. The optimal number, location and type of sensors is determined based on the most important features selected by the random forest regressor.

3.2. Approach

The flow diagram of the approach can be seen in Figure 2.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

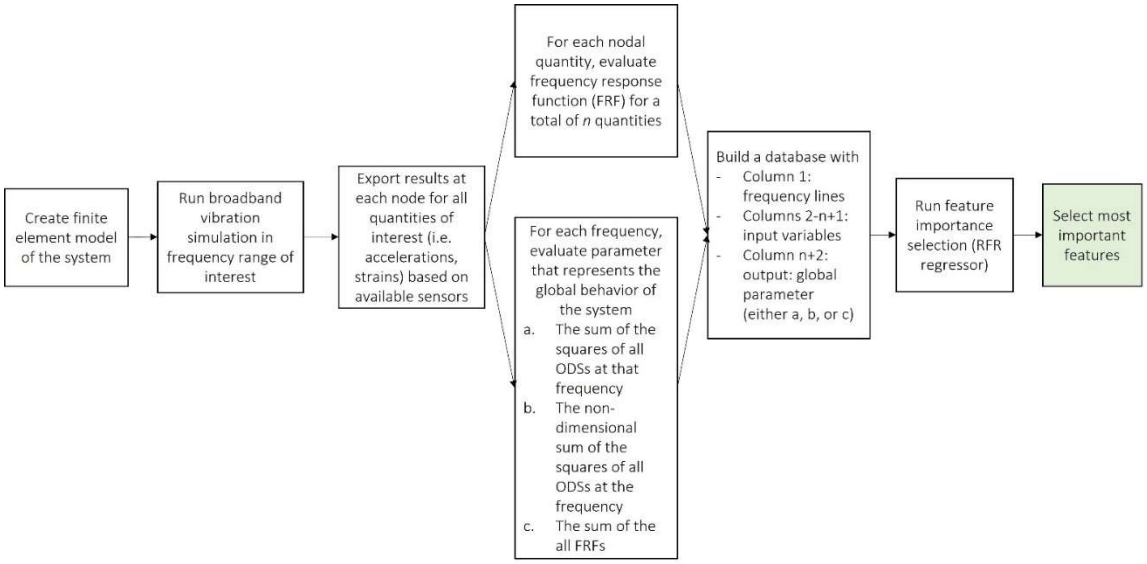


Figure 2. Flow diagram of the proposed approach.

First, a finite element (FE) model of the system is created and a broadband time-domain simulation in the frequency domain of interest is performed. The results of the simulation are exported for each node and/or potential sensor locations. These locations must correspond to all the possible/viable locations for the sensors in the modal tests. All quantities corresponding to the desired sensors should be considered, such as strains, accelerations, etc. For each of these n locations and quantities, the frequency response function (FRF) is evaluated. FRF is defined as the ratio of response (i.e. acceleration, velocity, or displacement) with respect to the excitation force which is the reference. These quantities will be the input of the machine learning model.

Then, a scalar parameter that represents the global behavior of the system should be identified for each frequency at which the input FRFs are evaluated. This parameter will be used as output of the Random Forest feature importance approach.

Three different options for output parameter are evaluated in this paper: the raw Operational Deflection Shape (ODS), the normalized ODS, and the average FRF.

1. Global parameter (a): raw ODS.

The sum of the squares of the ODS at each possible sensor location is evaluated according to:

$$\text{Output a} = ODS_\omega^T ODS_\omega$$

This expression results in a distinct scalar value for each ODS at each frequency.

Output a, however, depends on the load condition of the beam and will change depending on the magnitude of the load applied to the beam;

2. Global parameter (b): normalized ODS.

To decrease the sensitivity of the output to the load conditions, a normalized form of output a is calculated:

$$\text{Output b} = \frac{ODS_\omega^T ODS_\omega}{|ODS|^2}$$

Dividing the ODS product by the magnitude of the ODS at that frequency reduces the effect of the external load on the output used by the Random Forest feature importance approach, and therefore on the sensor placement;

3. Global parameter (c): average FRF.

The last output chosen was the average FRF at a given frequency, where n is the number of nodes and the sum of the FRF at a given frequency is taken across all nodes n.

$$\text{Output c} = \frac{\sum FRF(f)}{n}$$

After all local and global quantities are evaluated, the database can be created according to Table 1. The first column contains the frequency, columns 2 to (n+1) contain the FRF at each desired location

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

and represent the input variable, the last column (n+2) contains the global parameter and will be the output quantity for the random forest regressor. The random forest method can be run to obtain the ranking of the most important features, which can then be selected as the location and type of sensor needed for modal testing.

Table 1. Structure of Random Forest feature importance dataset.

Freq (Hz)	**Node 1**	**Node 2**	…	**Node *n***	**Output**
f	Node 1 FRF(*f*)	Node 2 FRF(*f*)		Node *n* FRF(*f*)	Output (*f*)

3.3. Dataset Creation

The dataset for the random forest regression model is extracted from a finite element model of the desired system and reformatted as specified in Table 1. Each row in the dataset corresponds to a frequency for which the FRF of each node is calculated. Each row in this case represents the operational deflection shape (ODS) for a given frequency for every node in the numerical modal model. The frequency is used primarily for tracking and is not input into the RF. The output column in this table represents the value the model should attempt to represent.

The model will output a table of all the input features (nodes) and their corresponding importance for predicting the output value. Therefore, choosing a parameter for the output is crucial to producing results that reflect an optimal sensor placement. All three global parameters will be considered as possible output parameter and the resulting sensor placement will be presented in the next section.

4. Application of Random Forest Sensor Selection Approach to a One-Dimensional Structure

The proposed method is applied to the analysis of a cantilever aluminum beam, whose properties are listed in Table 2.

Table 2. Cantilever beam properties.

Property	**Value**
Length	0.242 m
Width	0.032 m
Thickness	0.00305 m
E	70 GPa
ν	0.33
ϱ	2700 kg/m^3

The transverse behavior of the beam is modeled with 1D Euler-Bernoulli beam elements. The beam is clamped on one side, corresponding to Node 1; a transverse time-varying load is applied at the free-end of the beam, corresponding to Node n of the beam, Figure 3.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

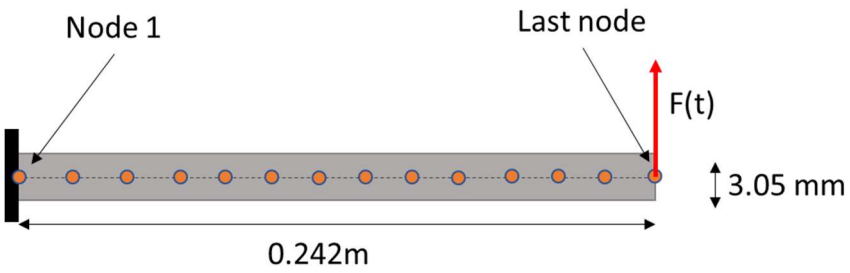


Figure 3. Schematic of beam model.

The first seven natural frequencies of the beam are listed in Table 3.

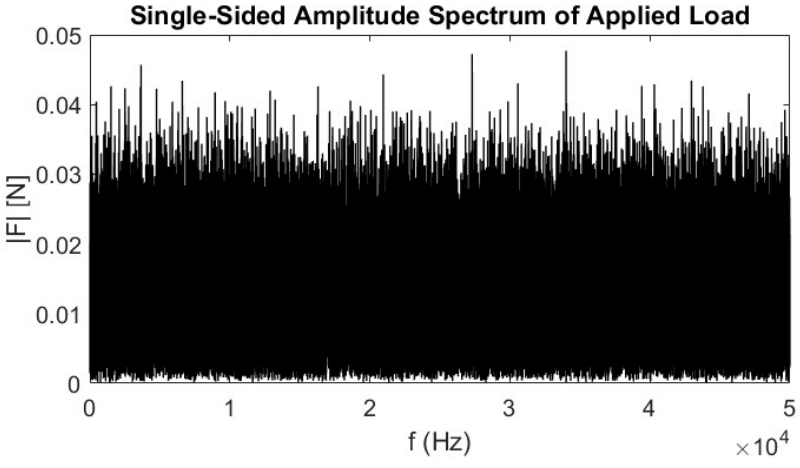
Table 3. Natural frequencies of cantilever beam.

Mode Number	FEA Natural Frequency (Hz)
1	42.83
2	268.26
3	750.45
4	1468.6
5	2423.4
6	3612.5
7	5033.1

In the first case, a fine mesh is considered (100 elements) for the RF analysis. A second case is presented in which the number of elements composing the mesh of the beam is reduced to 20 elements. The comparison of these two cases will determine whether the method is sensitive to the mesh of the model. In the third case, the time history of the applied load is changed to determine the sensitivity of the approach to the loading condition.

4.1. Densely Meshed Cantilever Beam (Case 1)

For this first case, a finite element model of the cantilever beam was created using 100 linear beam elements, corresponding to an element size and distance between nodes of 2.4mm. The beam was subjected to a transverse broadband gaussian white noise excitation from 0 Hz to 50 kHz, as shown in Figure 4, applied at the free-end of the beam.



Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

Figure 4. Applied load for case 1.

The database used by the RF feature importance approach was created from the transverse acceleration at each node. Transverse acceleration was chosen as the input parameter due to the wide availability of linear accelerometers for modal testing.

In this first example, all three definitions of the global parameter are explored. For each output option, an RF model is trained with the data available. To understand the capability of the RF to represent the output, the R2 and MSE are shown in Table 4. The MSE and R2 values in Table 4 appear excellent, giving confidence that the RF model is a good representation of the system and that the feature importance algorithm is reliable.

Table 4. R2 and MSE of RFR for case 1.

	(a) ODS	(b) Normalized ODS	(c) Avg. FRF
R2	0.900	0.940	0.940
MSE	7.0E-15	4.4E-9	2.8E-11

The ten most important features from each global parameter selection are listed in Table 5. These features represent the first ten candidate sensor locations identified by the RF algorithm. The nodal numbering starts with "node 1" at the root of beam and ends with node 101 at the tip of the beam.

The contribution of each feature/sensor to the variance of the data is depicted in Figure 5. The first ten features of the ODS RF account for 31% of the variance (global parameter a) while the first ten features of the normalized ODS account for 34% of the variance (global parameter b) The first 10 features of the average FRF account for 29% of the variance (global parameter c). The position of the first ten sensors identified by the RF approach for the three outputs are depicted in Figure 6; some of the sensors are overlapping or very close to each other (e.g. 2.4mm between sensor location 43 and 44 for output a), which is not physically possible in a real testing environment.

Table 5. Most important sensor locations (nodes) for case 1 based on different selections of the global parameter.

Feature Importance Rank	(a) ODS	(b) Normalized ODS	(c) Avg. FRF
1	86	66	61
2	69	100	47
3	44	68	94
4	12	92	42
5	43	73	41
6	70	34	48
7	53	82	5
8	101	20	82
9	96	88	101
10	87	83	57

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

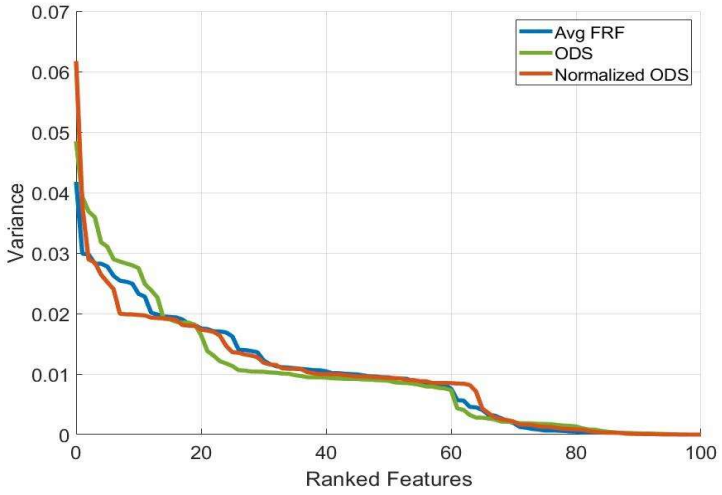


Figure 5. Variance as a function of number of features for case 1.

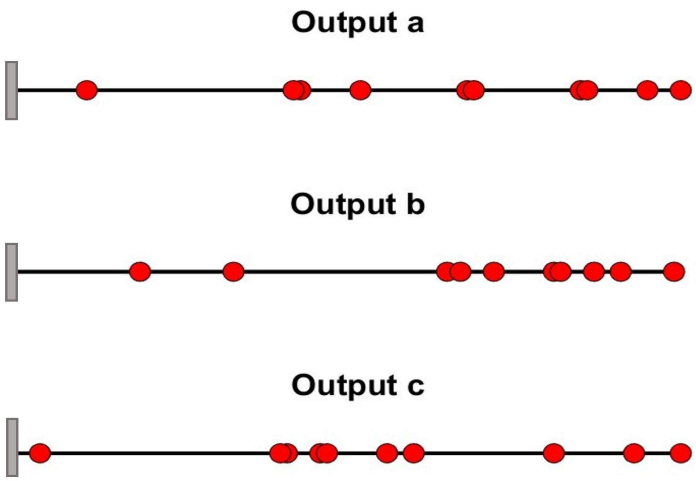


Figure 6. Sensor locations along the beam for case 1.

For all three choices of output, approximately 20 features are needed before at least 50% of the variance is accounted for; however, output (a) exhibits slightly higher individual variance in the first three features.

To estimate the mode shapes that the selected sensors will predict, the value of the actual (FEA) mode shape was taken at each candidate sensor location. To obtain the mode shapes in the figures, the numerical mode shapes were evaluated at the selected sensors' locations to verify that minimum aliasing is present with the proposed choice. The mode shapes are depicted in Figure 7 (Modes 1 – 4) and Figure 8 (Modes 4-6). Each column of the charts represents a mode, and the rows display the different global parameter choices (a: raw ODS, b: normalized ODS, c: average FRF). Sensor locations do not vary between modes but are plotted on top of the different mode shapes to visually evaluate the ability of the sensors to measure a given mode shape.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

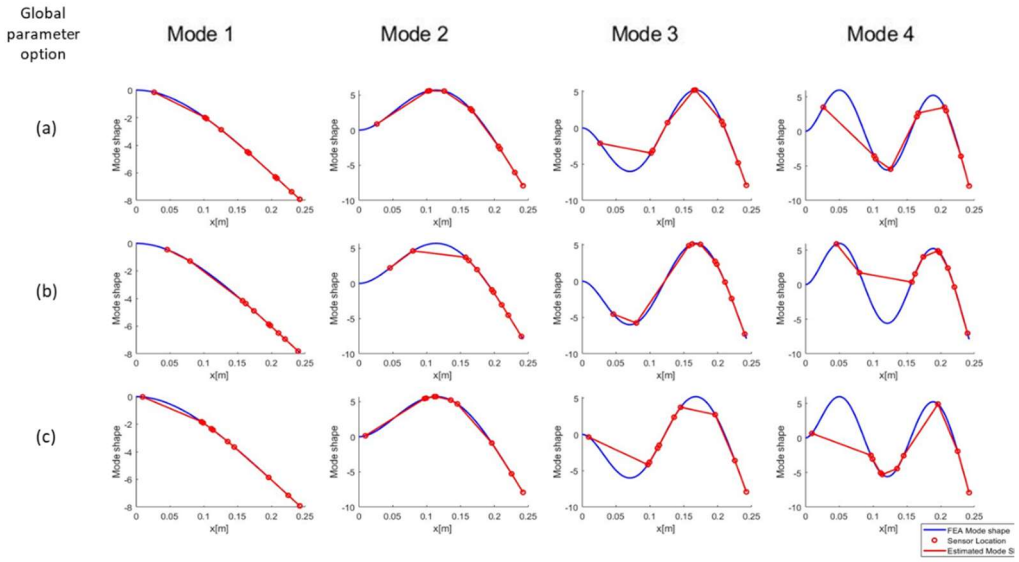


Figure 7. Modes 1-4 as predicted with first 10 features, case 1 (a: raw ODS, b: normalized ODS, c: average FRF).

All three sensor sets obtained with the different choices of output parameters can predict the first four modes with relative accuracy. Starting at mode 4, mode peak clipping can be noted with all global parameter options.

As mode number increases (Figure 8) the predictions made using the machine learning modeling fail to capture the behavior of the first third of the beam as all methods weight the free end of the beam more heavily. The method is able to capture the number of nodes for each mode, and do not exhibit large aliasing errors in the mode shape representation. At mode 7 and above, aliasing of the modes along the length of the beam starts becomes apparent, which is expected with only ten sensor placements on the structure.

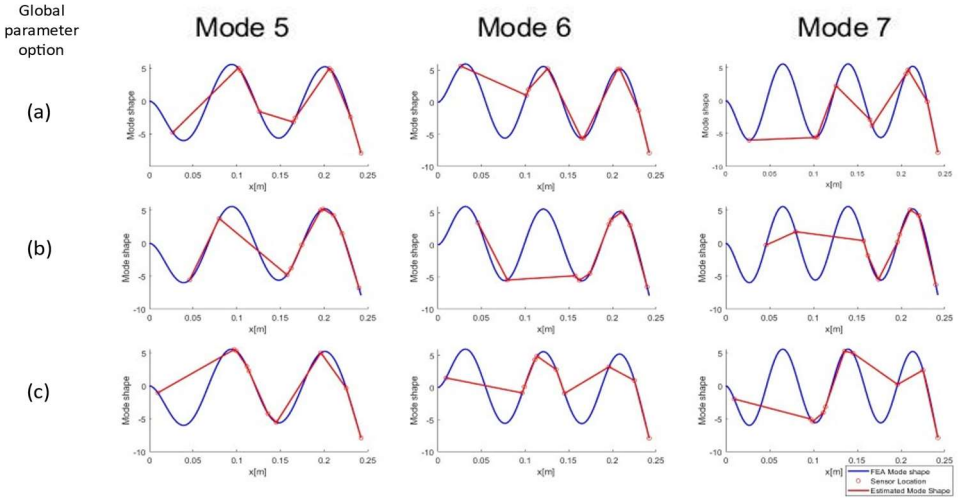


Figure 8. Modes 6-7 as predicted with first 10 features, case 1.

4.2. Effect of Mesh Density on Sensor Selection Using Random Forest Regressor (Case 2)

The previous subsection defined optimal sensor locations for modal analysis with a mesh of 100 elements. This mesh results in an element size of 2.42mm. Since the method allows for the placement of a sensor at any given node, it can select adjacent nodes for sensor placement (Figure 6 and Table 5). This distance between nodes (element size) is impractical for physical sensors. This subsection discusses the sensitivity of the approach to the mesh size.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

For the second case, the mesh is reduced to 20 elements, resulting in minimum sensor distances of 12.1mm, which is more reasonable. The rest of the parameters for the finite element and random forest analyses are the same as in case 1, including the applied excitation and the physical properties of the beam.

Table 6 lists the first eight positions selected as the best sensor locations by the RF feature importance applied to a coarser mesh. A nodal location of 1 corresponds to the root of the cantilever beam and 21 corresponds to the tip of the beam.

The first eight features of the ODS RF (global parameter a) account for 55% of the variance, which represents a considerable improvement with respect to case 1. Similar changes pertain to the other two choices of global parameter: the first ten features of the normalized ODS RF (global parameter b) account for 56% of the variance, while the first eight features of the Avg. FRF RF (global parameter c) account for 52% of the variance. This improvement is expected as the eight most important features in case 2 account for 40% of the total nodes, whereas in case 1 the ten most important features account for only 10% of the total nodes.

Table 6. Most important sensor locations for case 2, ranked by feature importance.

Feature Importance Rank	(a) ODS	(b) Normalized ODS	(c) Avg. FRF
1	12	20	14
2	11	21	13
3	20	14	19
4	8	19	7
5	19	12	12
6	21	7	17
7	14	6	18
8	6	10	3

The position of the first eight sensors identified by the RF approach for the three outputs are depicted in Figure 9; it is clear that the overlapping problems identified in case 1 have been eliminated.

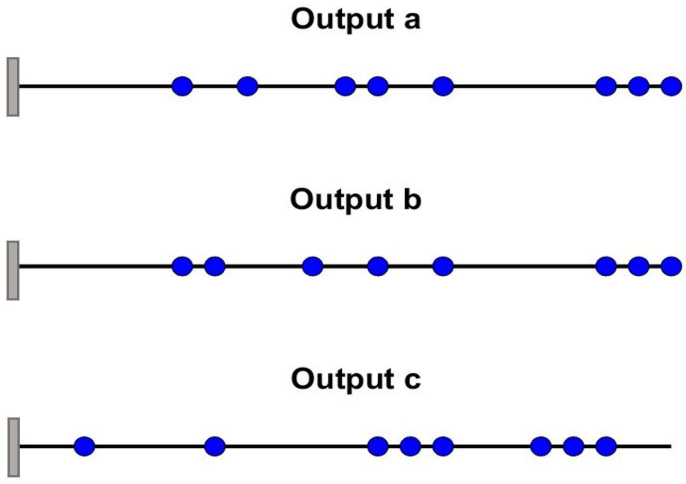


Figure 9. Sensor positioning for Case 2.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

A visualization of the sensors ability to identify the modes of the beam are depicted in Figure 10 and Figure 11. The plots were created with the same methodology as Figure 7 and Figure 8. Examining the plots, the ODS sensor selector (global parameter a) appears to perform worse, with almost all the sensors placed at inflection points for the 7th natural frequency.

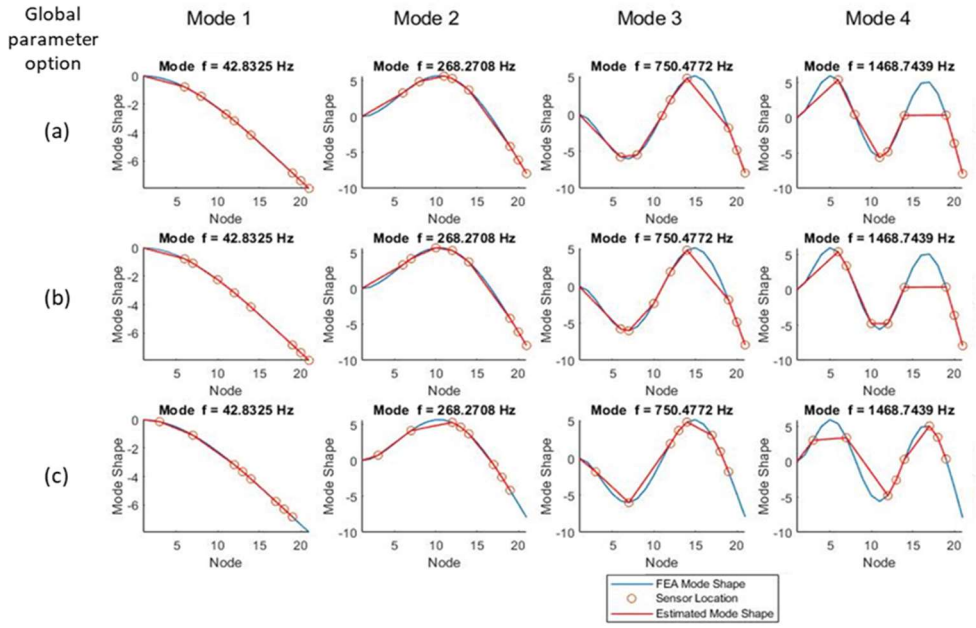


Figure 10. Modes 1-4 as predicted with first 10 features, case 2.

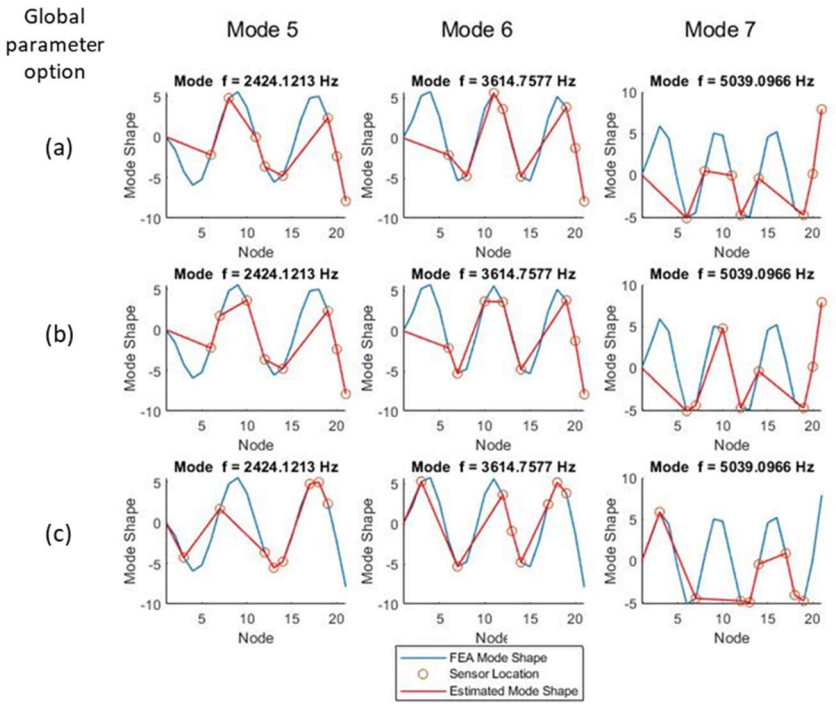


Figure 11. Modes 5-7 as predicted with first 10 features, case 2.

It is noticeable comparing the figures of the modes of case 1 (Figure 6) and 2 (Figure 9) that the selected sensor locations are similar for both cases, suggesting the robustness of the method with mesh size.

The ability of the method to properly select sensor locations for modal testing is further verified by extracting natural frequencies from the selected FRF using a modal analysis procedure. The

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

natural frequencies extracted by these signals using the gaussian white noise excitation are presented in Table 7. All choices for global output yield results that closely match the natural frequencies derived by the modal analysis of the finite element model; however, all sensor configurations are poor predictors of the first natural frequency, with the normalized ODS parameter performing better with an error of 27% (Table 7). In the mid-range frequencies, all three methods perform quite well, with errors below 6% from the numerical frequency.

Table 7. Extracted natural frequencies, case 2.

	FEA	(a) ODS		(b) norm ODS		(c) Avg. FRF	
Mode #	f (Hz)	f (Hz)	% Err	f (Hz)	% Err	f (Hz)	% Err
1	42.8	–	100%	54.3	27%	–	100%
2	268.3	277.5	3%	281.6	5%	278.0	4%
3	750.5	755.7	1%	757.5	1%	751.7	0%
4	1468.7	1471.0	0%	1462.8	0%	1462.6	0%
5	2424.1	2397.1	1%	2395.1	1%	2398.2	1%
6	3614.8	3526.0	2%	3531.6	2%	3499.3	3%
7	5039.1	4834.3	4%	4747.1	6%	–	–

4.3. Effect of Excitation Signals on Sensor Selection Using Random Forest Regressor (Case 3)

To study the effect that the choice of excitation signal has on sensor selection using the proposed methodology, two different excitation signals were chosen for comparison. Although traditional methods should select sensors for modal testing independently of the excitation signal, the random forest method is sensitive to the chosen excitation frequency due to how the input database is constructed. The first excitation consists of the gaussian white noise input signal used in the previous sections (cases 1 and 2), while the second excitation signal is a linear chirp as described below (case 3). The comparison will be based on a mesh size of 20 elements identical to case 2.

The beam was excited with a linear chirp signal, whose single-sided amplitude is depicted in Figure 12 as a function of frequency.

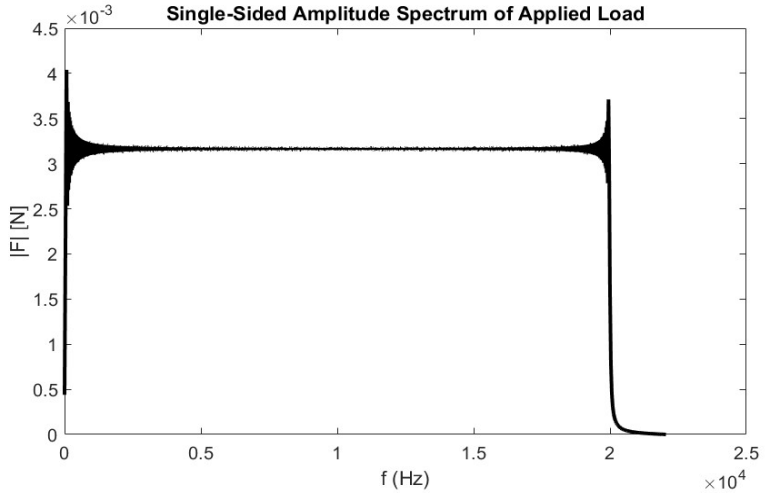


Figure 12. Single-sided amplitude of linear chirp from 20 Hz to 20 kHz, case 3.

The same three global parameters were considered for this case. The sensors selected by the approach for this different excitation signal are listed in Table 8.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

Table 8. Most important sensor locations for Case 3, ranked by feature importance.

Feature Importance Rank	(a) ODS	(b) Normalized ODS	(c) Avg. FRF
1	10	19	20
2	19	13	9
3	4	8	19
4	14	16	21
5	8	12	14
6	3	10	13
7	16	11	5
8	18	17	8

The position of the first eight sensors identified by the RF approach for the three outputs are depicted in Figure 13.

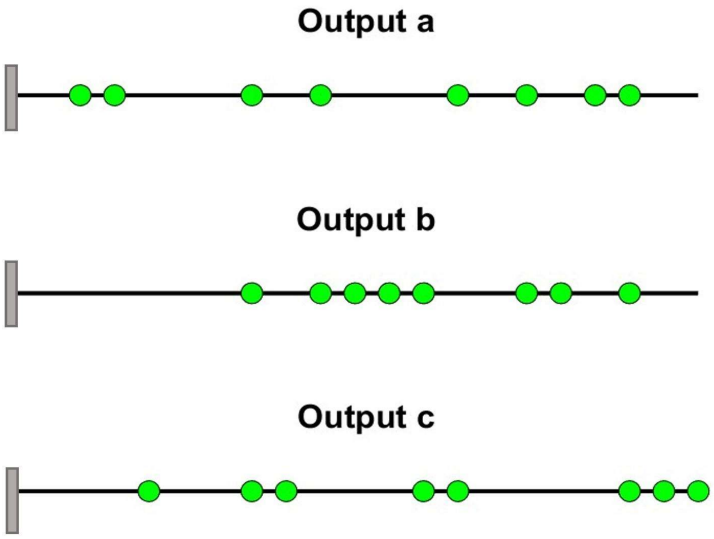


Figure 13. Sensor locations for case 3.

To better visualize the locations of the sensors and the potential for the sensors to capture the desired mode shapes, each selected sensor location is also plotted on the finite element derived mode shape for the first seven modes (Figure 14 and Figure 15).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

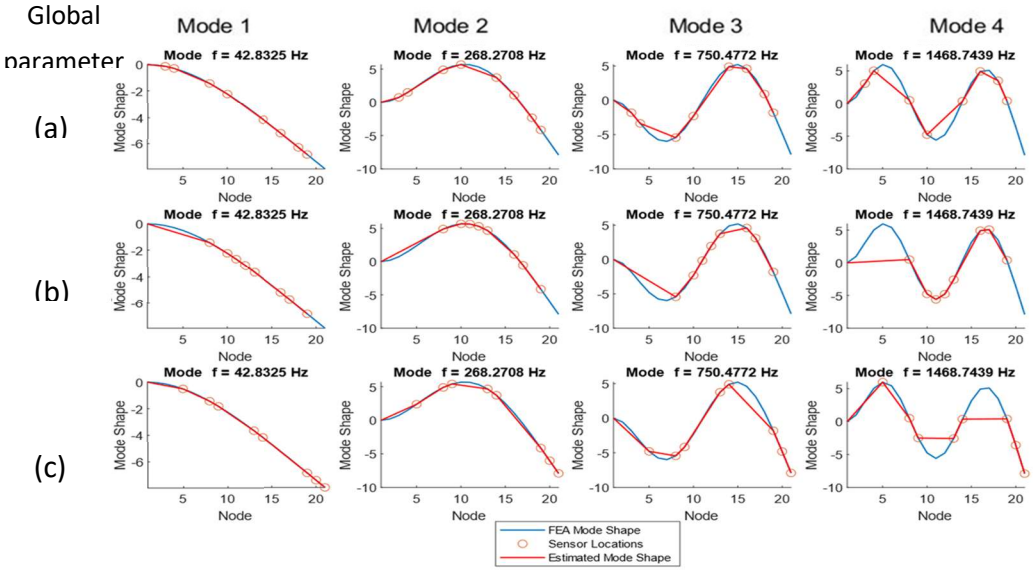


Figure 14. Sensor locations from chirp excitation for mode 1 through 4, case 3.

All global parameter options appear to track the first two mode shapes adequately; however, the normalized ODS (global parameter b) and FRF based (global parameter c) methodologies miss more peaks than the ODS (global parameter a) method especially at higher frequencies. Both the normalized ODS and FRF based methods exhibit peak clipping starting at mode 4 and place sensors at inflection points and so will not be able to capture those frequencies.

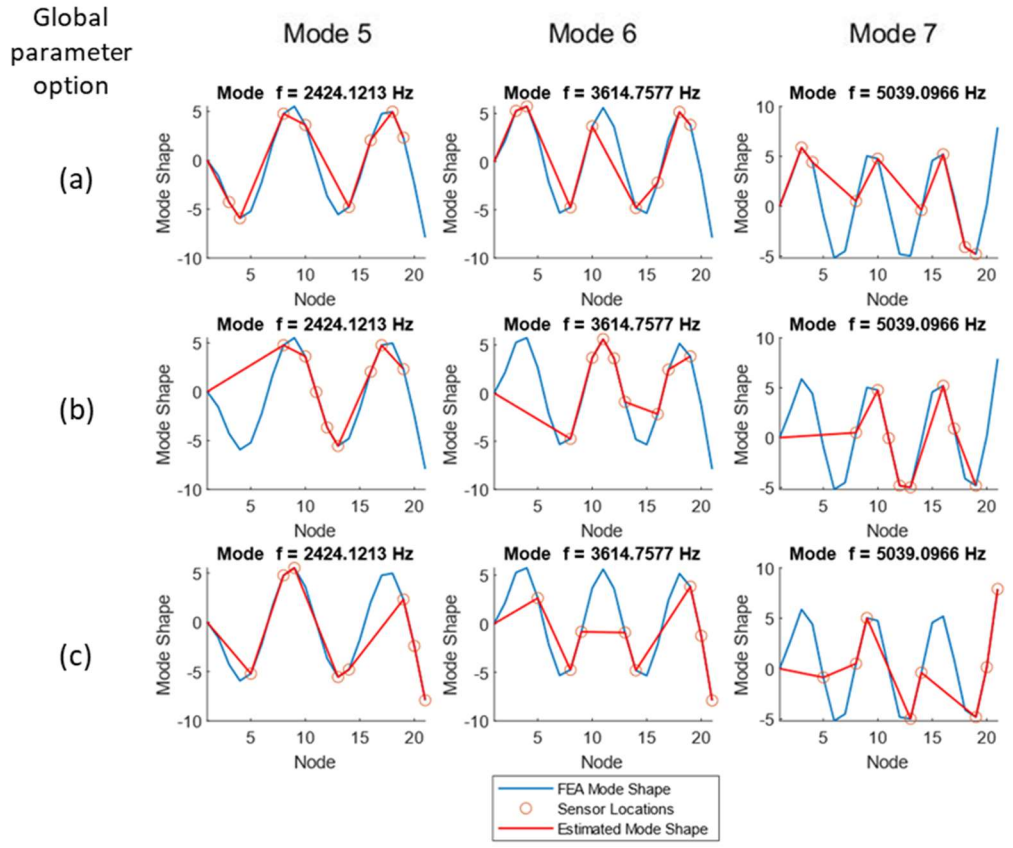


Figure 15. Sensor locations from chirp excitation for mode 5 through 7, case 3.

As a validation of the approach, the first seven natural frequencies are extracted using modal analysis of the first eight selected locations. The calculated natural frequencies are listed in Table 9.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

All three approaches are able to identify the first three natural frequencies, as the chirp excitation is better able to excite this frequency and mode. The error on the first natural frequency is however still large, ranging from 16% to 21% with respect to the first natural frequency calculated from the eigenvalues of the numerical system. The errors on the 2nd-7th natural frequencies are in line with case 2, suggesting that the method is reliable independently on the choice of excitation signal.

Table 9. Extracted natural frequencies, case 3.

FEA	(a) ODS		(b) norm ODS		(c) Avg FRF	
f (Hz)	f (Hz)	% Err	f (Hz)	% Err	f (Hz)	% Err
42.8	49.5	16%	52.0	21%	51.1	19%
268.3	279.7	4%	291.8	9%	280.1	4%
750.5	767.4	2%	765.8	2%	767.3	2%
1468.7	1463.0	0%	1463.3	0%	1463.5	0%
2424.1	2401.0	1%	2400.7	1%	2399.8	1%
3614.8	3539.1	2%	3537.1	2%	3525.5	2%
5039.1	4832.8	4%	–	–	–	–

5. Comparison of Proposed Methodology with Traditional Approaches and Discussion

This section compares the proposed methodology with results from a traditional sensor placement methodology for modal analysis, specifically the Effective Independence Method (EIM) [22,35]. In the case of the EIM, the sensors are chosen based on the numerical modes of the beam, and therefore do not depend on the applied excitation. EIM is applied to a mesh with 100 elements, similarly to case 1.

The first ten nodes identified by the EIM as best sensor locations for the analysis of the beam are listed in Table 10, and the location is plot in Figure 16.

Table 10. Optimal sensor location identified by traditional approach (EIM).

Effective Independence Method
9
16
25
32
39
48
63
70
78
94

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

Figure 16. Optimal sensor locations defined by Effective Independence Method.

A comparison between the natural frequencies identified with the optimal sensors' locations selected by the EIM and the random forest feature selection approach is provided in Table 11Table 12.

Table 11. Comparison between natural frequencies identified with the optimal sensors' locations selected by the EIM and the random forest feature selection approach for case 2.

FEA	EIM		(a) ODS		(b) norm ODS		(c) Avg FRF	
f (Hz)	f (Hz)	% Err	f (Hz)	% Err	f (Hz)	% Err	f (Hz)	% Err
42.8	–	100%	–	100%	54.3	27%	–	100%
268.3	278.9	4%	277.5	3%	281.6	5%	278.0	4%
750.5	755.6	1%	755.7	1%	757.5	1%	751.7	0%
1468.7	1482.5	1%	1471.0	0%	1462.8	0%	1462.6	0%
2424.1	2396.0	1%	2397.1	1%	2395.1	1%	2398.2	1%
3614.8	3523.6	3%	3526.0	2%	3531.6	2%	3499.3	3%
5039.1	4837.4	4%	4834.3	4%	4747.1	6%	–	–

Table 12. Comparison between natural frequencies identified with the optimal sensors' locations selected by the EIM and the random forest feature selection approach for case 3.

FEA	EIM		(a) ODS		(b) norm ODS		(c) Avg FRF	
f (Hz)	f (Hz)	% Err	f (Hz)	% Err	f (Hz)	% Err	f (Hz)	% Err
42.8	49.4	15%	49.5	16%	52.0	21%	51.1	19%
268.3	279.7	4%	279.7	4%	291.8	9%	280.1	4%
750.5	767.3	2%	767.4	2%	765.8	2%	767.3	2%
1468.7	1463.0	0%	1463.0	0%	1463.3	0%	1463.5	0%
2424.1	2401.0	1%	2401.0	1%	2400.7	1%	2399.8	1%
3614.8	3539.1	2%	3539.1	2%	3537.1	2%	3525.5	2%
5039.1	4833.1	4%	4832.8	4%	–	–	–	–
6696.0	6317.7	6%	6317.1	6%	6306.8	6%	6459.3	4%

Both the proposed method and the traditional EIM method appear to perform similarly in the beam problem, yielding similar natural frequencies to each other in both loading conditions (Table 11Table 12). The natural frequencies resulting from the choice of output a (ODS) are generally characterized by a lower percent error than the use of output b (normalized ODS) and c (average FRF). The proposed method is therefore considered a feasible approach with traditional methodologies.

The choice of global output a seems to be more reliable than global outputs b and c in its ability to identify a set of sensors that maximizes the larger number of natural frequencies that can be

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

extracted through modal analysis. The approach is also robust with respect to the applied excitation; although the optimal sensor location slightly changes when different excitations are used to generate the database for input to the RF method, the identified natural frequencies do not show large differences. Additionally, an excitation needs to be applied to perform modal testing, and we could therefore argue that the use of the expected excitation during testing to create the database will result in optimal sensor positioning for a given excitation.

The application of the proposed method also has the advantage of not requiring an iterative approach, and can be quickly applied to preexisting finite elements results. It does however require the solution of transient vibration simulations, which can however be time consuming.

6. Conclusions

Machine learning represents an appealing solution to the issue of sensor selection for modal testing. Current algorithms used for sensor selection are iterative when implemented and, for large geometries and complex models, the computational time can be substantial. The method presented here exploits the ability of the random forest approach to select the most important feature of a database for modal analysis sensor selection based on finite element models of the system. A database is created in the frequency domain; the responses of interest at every nodal location of the model are considered as possible sensor locations. The output of the random forest feature selection approach is defined based on a global output that characterizes the system. Three options are evaluated in the paper: the ODS of the system, a normalized ODS, and the average FRF at the specific frequency. The approach is applied to a one-dimensional model of a vibrating cantilever beam. The optimal sensors' locations identified by the proposed approach is compared to the sensors selected by the Effective Independence Method and appears to yield similar results to the proposed approach. Within the proposed approach, the choice of the operational deformed shape (ODS) as global parameter appears to be more robust than the other proposed options. The approach is evaluated for sensitivity with respect to mesh size and type of excitation signal, and is robust to any changes in these parameters.

Author Contributions: "Conceptualization, M.C and T.K.; methodology, M.C and T.K.; software, M.C and T.K; validation, M.C and T.K.; formal analysis, T.K.; investigation, M.C and T.K; resources, M.C. and F.D..; data curation, T.K.; writing—original draft preparation, M.C and T.K.; writing—review and editing, M.C, T.K. and F.D.; visualization, M.C and T.K; supervision, M.C and F.D.; project administration, M.C.; funding acquisition, M.C. and F.D.. All authors have read and agreed to the published version of the manuscript."

References

1. D. J. Ewins, *Modal Testing: Theory and Practice*. Research Studies Press Ltd., 1984.
2. C. M. Harris, *Shock and Vibration Handbook*, 2nd ed. New York: McGraw-Hill Book Company, 1976.
3. C. Stephan, "Sensor placement for modal identification," *Mech Syst Signal Process*, vol. 27, no. 1, pp. 461–470, Feb. 2012, doi: 10.1016/j.ymssp.2011.07.022.
4. C. Papadimitriou, "Optimal sensor placement methodology for parametric identification of structural systems," *J Sound Vib*, vol. 278, no. 4–5, pp. 923–947, Dec. 2004, doi: 10.1016/j.jsv.2003.10.063.
5. C. Papadimitriou and G. Lombaert, "The effect of prediction error correlation on optimal sensor placement in structural dynamics," *Mech Syst Signal Process*, vol. 28, pp. 105–127, Apr. 2012, doi: 10.1016/j.ymssp.2011.05.019.
6. C. Papadimitriou, C. P. Fritzen, P. Kraemer, and E. Ntotsios, "Fatigue predictions in entire body of metallic structures from a limited number of vibration sensors using Kalman filtering," *Struct Control Health Monit*, vol. 18, no. 5, pp. 554–573, Aug. 2011, doi: 10.1002/stc.395.
7. C. R. Farrar and K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*. 2012. doi: 10.1002/9781118443118.
8. P. Vurtur Badarinath, M. Chierichetti, and F. Davoudi Kakhki, "A Machine Learning Approach as a Surrogate for a Finite Element Analysis: Status of Research and Application to One Dimensional Systems," *Sensors*, vol. 21, no. 5, p. 1654, Feb. 2021, doi: 10.3390/s21051654.
9. K. Demirlioglu, S. Gonen, and E. Erduran, "On the Selection of Mode Shapes Used in Optimal Sensor Placement," *Sensors and Instrumentation, Aircraft/Aerospace and Dynamic Environments Testing, Volume 7. Conference Proceedings of the Society for Experimental Mechanics Series*, pp. 85–92, 2023, doi: 10.1007/978-3-031-05415-0_8.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

10. B. Y. Zhang and Y. Q. Ni, "A data-driven sensor placement strategy for reconstruction of mode shapes by using recurrent Gaussian process regression," *Eng Struct*, vol. 284, p. 115998, Jun. 2023, doi: 10.1016/J.ENGSTRUCT.2023.115998.
11. A. Castillo and A. R. Messina, "Data-driven sensor placement for state reconstruction via POD analysis," *IET Generation, Transmission & Distribution*, vol. 14, no. 4, pp. 656–664, Feb. 2020, doi: 10.1049/IET-GTD.2019.0199.
12. G. Suryanarayana, J. Arroyo, L. Helsen, and J. Lago, "A data driven method for optimal sensor placement in multi-zone buildings," *Energy Build*, vol. 243, p. 110956, Jul. 2021, doi: 10.1016/J.ENBUILD.2021.110956.
13. P. Sashittal and D. J. Bodony, "Data-driven sensor placement for fluid flows," *Theor Comput Fluid Dyn*, vol. 35, no. 5, pp. 709–729, 2021, doi: 10.1007/s00162-021-00584-w.
14. K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, "Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns," *IEEE Control Syst*, vol. 38, no. 3, pp. 63–86, Jun. 2018, doi: 10.1109/MCS.2018.2810460.
15. M. Farid and D. Solav, "Data-driven sensor placement optimization for accurate and early prediction of stochastic complex systems," *J Sound Vib*, vol. 543, p. 117317, Jan. 2023, doi: 10.1016/J.JSV.2022.117317.
16. E.-T. Lee and H.-C. Eun, "Optimal Sensor Placement in Reduced-Order Models Using Modal Constraint Conditions," *Sensors*, vol. 22, no. 2, 2022, doi: 10.3390/s22020589.
17. T. Kelmar and M. Chierichetti, "Machine Learning Based Sensor Selection for Modal Testing," in *AIAA SCITECH 2024 Forum*, Orlando, FL, Jan. 2024. doi: 10.2514/6.2024-0196.
18. M. Papadopoulos and E. Garcia, "Sensor placement methodologies for dynamic testing," *AIAA Journal*, vol. 36, no. 2, pp. 256–263, 1998, doi: 10.2514/2.7509.
19. J. E. Coote, N. A. J. Lieven, and G. W. Skingle, "Sensor placement optimization for modal testing of a helicopter fuselage," in *Proceedings of the 24th International Modal Analysis Conference (IMAC-XXIII), Orlando, FL, USA*, 2005, pp. 7–10.
20. Y. Jiang, D. Li, and G. Song, "On the physical significance of the Effective Independence method for sensor placement," *J Phys Conf Ser*, vol. 842, no. 1, p. 12030, May 2017, doi: 10.1088/1742-6596/842/1/012030.
21. F. E. Udwadia, "Methodology for Optimum Sensor Locations for Parameter Identification in Dynamic Systems," *J Eng Mech*, vol. 120, no. 2, pp. 368–390, 1994, doi: 10.1061/(ASCE)0733-9399(1994)120:2(368).
22. D. C. Kammer, "Sensor placement for on-orbit modal identification and correlation of large space structures," *Journal of Guidance, Control, and Dynamics*, vol. 14, no. 2, pp. 251–259, 1991, doi: 10.2514/3.20635.
23. J. Lollock and T. Cole, "The Effect of Mass Weighting on the Effective Independence of Mode Shapes," in *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, 2005, pp. 449–456. doi: 10.2514/6.2005-1836.
24. R. N. Coppolino, "Systematic Modal Test Planning," in *The Integrated Test Analysis Process for Structural Dynamic Systems*, Cham: Springer International Publishing, 2020, pp. 37–64. doi: 10.1007/978-3-031-79729-3_3.
25. R. Coppolino, "Aerospace Perspective for Modeling and Validation," in *Handbook of Experimental Structural Dynamics*, P. Allemang Randall and Avitabile, Ed., New York, NY: Springer New York, 2020, pp. 1–36. doi: 10.1007/978-1-4939-6503-8_30-1.
26. J. Brownlee, "How to Perform Feature Selection With Numerical Input Data." Accessed: May 27, 2021. [Online]. Available: https://machinelearningmastery.com/feature-selection-with-numerical-input-data/
27. Scikit-learn Developers, "sklearn.feature_selection.SelectKBest — scikit-learn 0.24.2 documentation." Accessed: May 27, 2021. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html#sklearn.feature_selection.SelectKBest
28. F. Pedregosa *et al.*, "Skikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
29. P. Płoński, "Random Forest Feature Importance Computed in 3 Ways with Python | MLJAR Automated Machine Learning." Accessed: May 27, 2021. [Online]. Available: https://mljar.com/blog/feature-importance-in-random-forest/
30. M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, and E. Tuzhilina, "Principal component analysis," *Nature Reviews Methods Primers*, vol. 2, no. 1, p. 100, 2022, doi: 10.1038/s43586-022-00184-w.
31. F. Song, Z. Guo, and D. Mei, "Feature selection using principal component analysis," in *Proceedings - 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, ICSEM 2010*, 2010, pp. 27–30. doi: 10.1109/ICSEM.2010.14.
32. S. Choppala, T. W. Kelmar, M. Chierichetti, F. Davoudi, and D. Huang, "Optimal sensor location and stress prediction on a plate using machine learning," in *AIAA SCITECH 2023 Forum*, National Harbor: AIAA, 2023, pp. 1–12. doi: 10.2514/6.2023-0370.
33. P. Płoński, "Random Forest Feature Importance Computed in 3 Ways with Python | MLJAR Automated Machine Learning." Accessed: May 28, 2021. [Online]. Available: https://mljar.com/blog/feature-importance-in-random-forest/

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0604.v1

34. M. Chierichetti and F. Davoudi, "Optimal sensor location along a beam using machine learning," in *AIAA SCITECH 2022 Forum*, Jan. 2022. doi: 10.2514/6.2022-1465.
35. D. C. Kammer, "Optimal sensor placement for modal identification using system-realization methods," *Journal of Guidance, Control, and Dynamics*, vol. 19, no. 3, pp. 729–731, 1996, doi: 10.2514/3.21688.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.