Preflight Summary Report for: 1-s2.0-S2352012423015059-main 1111.pdf

Profile: Convert to PDF/A (Processed pages 1 to 19)

Processed by Brian Ricks, Date: 12/15/23 12:23 PM

Fixups

- Convert to PDF/A-1a (1 object)
- Make document XMP Metadata compliant with PDF/A-1 (1 object)
- Remove document structure compression (1 object)
- Compress all uncompressed objects using lossless ZIP compression (1 object)
- Recompress LZW as ZIP (1 object)
- Insert missing Type entry in StructElem objects (33 objects)
- Repair invalid bookmark hierarchies (1 object)

Results (Summary)

Error

- XMP property neither predefined nor defined in extension schema

Document information

File name: "1-s2.0-S2352012423015059-main 1111.pdf"
Path: "/Users/bricks/Library/CloudStorage/OneDrive-UniversityofNebraskaatOmaha/[redacted]Research/[redacted]Assets/[redacted]Grants/23.12 NSF SMARTI Final Report"
PDF version number: "1.4"
File size (MB): 4.5
Title: "Singular value decomposition and unsupervised machine learning for virtual strain sensing: Application to an operational railway bridge"
Author: "Emmanuel Akintunde; Saeed Eftekhar Azam; Daniel G. Linzell"
Creator: "Elsevier"
Producer: "Acrobat Distiller 8.1.0 (Windows)"
Created: "12/15/23 12:20 PM"
Modified: "12/15/23 12:22 PM"
Keywords: "Strain,Estimation,Singular,Value,Decomposition,Self-organizing,Map,K-means,Clustering,Proper,Orthogonal,Steel,Railway,Bridge"
Subject: "Structures, 58 (2023) 105417. doi:10.1016/j.istruc.2023.105417"
Trapping: "Unknown"
Number of plates: 1
Names of plates: "(Black) "

Environment

Preflight, 18.6.0 (268)
Acrobat version: 23.008
Operating system: macOS 13.4.0

Structures 58 (2023) 105417

Contents lists available at ScienceDirect

Structures

journal homepage: www.elsevier.com/locate/structures

Singular value decomposition and unsupervised machine learning for virtual strain sensing: Application to an operational railway bridge

Emmanuel Akintunde [a,*], Saeed Eftekhar Azam [b], Daniel G. Linzell [c]

[a] Olsson Inc., Omaha, NE, USA
[b] Department of Civil and Environmental Engineering, University of New Hampshire, Durham, NH, USA
[c] Department of Civil and Environmental Engineering, University of Nebraska-Lincoln, Lincoln, NE, USA

ARTICLE INFO

Keywords:
Strain
Estimation
Singular
Value
Decomposition
Self-organizing
Map
K-means
Clustering
Proper
Orthogonal
Steel
Railway
Bridge

ABSTRACT

Metal fatigue is a major concern in civil, mechanical, and offshore structures. As a result, inspections are frequently directed to critical, fatigue prone locations using visual inspections, nondestructive evaluations, or sensor data. However, fatigue assessment may be hampered if access to these fatigue-prone areas is difficult or impossible. Furthermore, when comprehensive sensor use as part of a bridge health monitoring system is desired, cost and power requirements can be prohibitive. These identified limitations have motivated studies examining virtual sensing methods that estimate strains at unmeasured locations using indirect measurements. Kalman filtering (KF) and modal expansion (ME) are two popular strain estimation processes. However, both processes are model-based, requiring calibration of a finite element (FE) model of the structure of interest, which can be time-consuming, particularly for complex structures. This constraint has spurred the need for data-driven strain estimation methods, which depends solely on data from the structure, typically provided by sensors, without any additional a priori knowledge of the structure, such as what could be provided by a FE model. This study investigated the use of a novel, data-driven, Singular Value Decomposition (SVD) based method for strain estimation on an operational railroad bridge. Left singular vector (LSV) SVD modes, also known as Proper Orthogonal Modes (POMs), were employed for strain estimation. Machine learning (ML) was implemented to reduce POM variability and subsequently increase estimation accuracy using two classification methods: k-means clustering and root mean square (RMS); and self-organizing maps (SOM) and POMs. Strains were predicted using strain time-history POMs from snapshot matrices from clustered groups of train passages to estimate unmeasured time-histories from the same group. The method was applied to operational strain measurements from approximately 300 train passages over a steel, truss, railway bridge. Results showed that use of the data driven SVD technique in conjunction with ML could predict suitable strain time signals at unmeasured locations, data that can subsequently be utilized when performing fatigue assessment.

1. Introduction

One of the challenges when assessing steel structure fatigue susceptibility is inaccessibility to fatigue prone locations. If sensors can be implemented as part of a structural health monitoring system, generally a limited number are available due to their cost and power requirements. To address this issue, several research studies have focused on using an optimized number of sensors that can predict strain response at unmeasured locations, which can include fatigue prone regions on structures, using what is termed virtual sensing. Common processes adopted for strain estimation include KF and ME. KF comes from control theory and is an optimal predictor of a system's variable of interest using a state space model of the system [1]. ME uses a linear transformation of measured response to enhance the value of structural response data for state estimation based on mode shapes of the system [2].

Papadimitriou et al. [3] proposed a methodology for estimating structure fatigue damage accumulation using spectral characteristics obtained from vibration measurements at a limited number of locations. The authors applied KF to estimate stress power spectral densities (PSD) at unmeasured locations based on limited, available response time-history measurements in conjunction with a dynamic structural model. PSD predictions were used to estimate fatigue damage

* Corresponding author.
E-mail address: eakintunde@olsson.com (E. Akintunde).

https://doi.org/10.1016/j.istruc.2023.105417
Received 29 September 2022; Received in revised form 14 October 2023; Accepted 17 October 2023
Available online 27 October 2023

accumulation and predict remaining fatigue life. The proposed methodology was evaluated using simulated measurements from a two-dimensional truss model. Results showed that fatigue life predictions using KF were sufficiently consistent with values from the reference model. The authors concluded that accuracy of the proposed methodology depends on model complexity, observation errors, and the number and location of sensors.

Smyth and Wu [4] used KF to fuse accelerations and displacements sampled at different frequencies to produce more accurate displacement predictions. Jo and Spencer [5] numerically showed that combining multi-metric measurements, which involve integration of more than one measurement metric using KF, such as acceleration and strain, improved estimation of unmeasured strains when compared to estimations based on a single acceleration or strain measurement. Palanisamy et al. [6] further verified performance of KF data fusion using strain estimation at unmeasured locations via experimental tests of a simply supported beam.

To address state estimation in non-linear dynamic systems (e.g., systems with time varying characteristics), some KF extensions have been introduced. A common application of these extensions involves concurrently estimating system states and unknown parameters, referred to as joint state and parameter estimation [7]. KF extensions include: Unscented KF (UKF) [8]; Extended KF (EKF) [9]; Augmented KF (AKF) [10]; and Particle Filters (PF) [11].

ME has been used successfully to estimate strain response for a wide range of mechanical, civil, and offshore structures. Graugaard-Jensen et al. [12] employed ME for full-field strain prediction of a laboratory tested structure and an in-service lattice tower under operational conditions. Avitabile and Pingle [13] utilized ME on a laboratory tested structure assembled from two aluminum plates to determine full-field strains using limited measurement locations. Aenlle et al. [14] adopted ME for strain estimation within a scale model of a two-story building. Skafte et al. [15] combined ME with Ritz-vectors to account for low frequency response induced by quasi-static effects of wave action on offshore structures. The method was validated via accurate strain predictions using data from a scale model offshore platform excited by shakers simulating wave spectra. Nabuco et al. [16] successfully predicted geometric strain response of two, scaled, simulated, and laboratory tested offshore platforms using ME based on parameters established from a linear model.

ME techniques can use either mode shapes from finite element (FE) analyses [13,17] or expanded experimental mode shapes to obtain strain or stress estimates [18,19]. Tarpø et al. [1] examined the effectiveness with which both ME approaches estimated full field strains and stresses for fatigue analyses of operational offshore structures. ME was applied to a scaled offshore laboratory platform model to estimate strain response using mode shapes from finite element models along with expanded mode shapes obtained from Operational Modal Analysis (OMA) of the system's response. Estimated strains were analyzed using different metrics: the coefficient of determination, time response assurance criterion (TRAC), and what was termed the normalized fatigue damage error. Results showed that ME using expanded experimental mode shapes can improve strain estimations due to its potential to reduce finite element modeling errors. However, some fitting errors were introduced because of the expansion process and the finite element model was shown to outperform expanded experimental modes in certain cases.

Certain limitations associated with utilizing KF and ME methods for strain estimation have been identified. Both methods are physics-based, requiring calibration of an FE model to determine modal parameters used for strain estimation of the structure under investigation [20]. While KF uses both vibration modes and frequencies from calibrated models for strain estimation, ME utilizes solely vibration modes. FE model calibration, especially for large and complex structures, can be tedious. In addition, these methods have been validated primarily using numerical and laboratory models. Although the literature introduces several Kalman-type filters for strain prediction of large-scale structural systems, KF application to systems subjected to moving inertial loads, such as bridges, is limited to numerical simulations [21]. Similarly, ME has largely been validated using simulated experiments [20] and measured strains from offshore structures [7].

Recently, Azam et al. [22] examined the ability with which virtual sensing methods delivered accurate strain estimation for an in-service steel railway bridge using a sparse sensor network. The authors introduced a novel application of data-driven SVD and compared strain estimation performance against estimations from physics-based AKF and ME. A portion of one of the bridge's exterior stringers was instrumented with five Bridge Diagnostic, Inc. (BDI) strain transducers and data were recorded for two train passages. Measured data from four sensors were used to estimate strain time-histories at a fifth sensor. Using vibration modes and frequencies from a calibrated FE model, AKF strain prediction performance and accuracy were compared to measured strains. Vibration modes from the FE model were used for strain estimation with ME while SVD used LSV modes, also referred to as POMs, in place of vibration modes to predict strain response. Accurate strain predictions were obtained, and results demonstrated that these approaches could be employed as virtual sensing tools in a health monitoring system, with SVD outperforming AKF and ME.

Azam et al.'s [22] findings indicated that further investigation of data driven SVD method for strain estimation is necessary as the study was limited to two train passages since SVD is a data-driven method. Train loads and speeds, as well as track locations, were also not clearly defined. Since an in-service railway bridge will be exposed to trains of different speeds, lengths, and weights traveling in different directions, inherent variability may affect the accuracy of SVD left singular vector modes (i.e., POMs) and subsequent strain estimations. Therefore, it was of interest to investigate how variability in speed, load magnitude and distribution, and train passage location affected SVD strain estimation accuracy.

As a result, in this paper strain estimation using a data driven SVD method was further investigated using significantly larger train passage data sets from the monitored railway bridge. Generated strain time-history POMs for selected train passages were also examined and two methodologies proposed to reduce their variability for better strain estimation. One of the methodologies classified the strain time-history based on the RMS utilizing k-means clustering [23], while the other used classification POM-based strain time-histories with SOM obtained from neural network clusters [24]. These methods were used to categorize strain time-histories as a function of trains having similar loads and speeds and moving across the bridge in similar transverse locations, thus reducing POM variability. Strains were estimated using POMs of strain time-history snapshot matrices from the groups/classes to predict unmeasured location strain time-histories from the same group. Prediction accuracy for each method was assessed by different quality measurements including normalized fatigue error, TRAC, and coefficient of determination (R^2).

2. SVD for strain estimation

2.1. Singular value decomposition

SVD is a powerful numerical analysis tool, particularly for linear domains to complete matrix decomposition. Beltrami and Jordan were reported to have developed SVD for square matrices in the 1870 s [25], Autonne [26] for complex matrices, and Eckart and Young [27] for rectangular matrices. Sylvester [28], Schmidt [29], and Weyl [30] are other mathematicians who contributed to SVD and development for matrix decomposition.

SVD factors a matrix (A), into products of a unitary matrix (Ψ), a diagonal matrix (Ω), and another unitary matrix, ($\mathbf{V}^T$): $\Psi\Omega\mathbf{V}^T$[31]. Several advantages make SVD a very useful tool for matrix

Fig. 1. Monitored railway bridge.

approximation. Decomposition accomplished using unitary matrices makes SVD a great tool for n-space geometries. Since SVD does not add to a problem's intrinsic sensitivity to perturbations, it is numerically stable. This implies that minor changes in matrix A correlate to small changes in Ω, and visa-versa. Furthermore, Ω's diagonality allows for easy observation when A becomes a rank-degenerate matrix. When this is observed, decomposition yields optimal low rank approximations for A. Golub and Reinsch [32] helped develop efficient and stable algorithms to complete SVD.

Strain time-histories from the studied, steel, railway bridge were saved in a snapshot matrix $A \in \mathbb{R}^{n_m \times n_s}$ from n_m measurements and n_s samples having the following form:

$$A = [\alpha_1 \quad \cdots \quad \alpha_{n_s}], \tag{1}$$

where: $\alpha_k \in \mathbb{R}^{n_m}$ is data measured from n_m sensors at discrete time intervals for $k = 1, 2, \cdots, n_s$. SVD of the snapshot matrix A yields the following:

$$\text{svd}(A) = \Psi\Omega V^T, \tag{2}$$

and: $\Psi \in \mathbb{R}^{n_m \times n_m}$ is a unitary matrix, whose columns are left singular vectors of A (i.e., POMs); $\Omega \in \mathbb{R}^{n_m \times n_s}$ is a diagonal semi-matrix whose components Ω_{ii} are singular values of Ψ; and $V \in \mathbb{R}^{n_s \times n_s}$ is a unitary matrix whose columns are right singular vectors of A.

2.2. Poms for strain estimation

The equation of motion for a damped, n-degree of freedom (DOF) system can be expressed as follows:

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{p}(t), \tag{3}$$

where: $\mathbf{M} \in \mathbb{R}^{n \times n}$, $\mathbf{C} \in \mathbb{R}^{n \times n}$, and $\mathbf{K} \in \mathbb{R}^{n \times n}$ are mass, damping, and stiffness matrices respectively;$\ddot{\mathbf{u}} \in \mathbb{R}^n$, $\dot{\mathbf{u}} \in \mathbb{R}^n$, and $\mathbf{u} \in \mathbb{R}^n$ are acceleration, velocity, and displacement column vectors of the system at time (t) respectively, and $\mathbf{p}(t) \in \mathbb{R}^n$ is the applied load on the system at any time (t).

Response of the system can be written via ME as shown below:

$$\mathbf{u}(t) = \sum_{i=1}^{n} \phi_i \mathbf{q}_i(t) = \Phi\mathbf{q}(t), \tag{4}$$

where: $\Phi \in \mathbb{R}^{n \times n}$ is the modal matrix containing all vibration modes ϕ_i as column vectors; and $\mathbf{q}(t) \in \mathbb{R}^n$ are the modal coordinates [33].

Since SVD is to be utilized instead of ME, Eqn. (4) can be rewritten as a linear combination of left singular vectors:

$$\mathbf{u}(t) = \sum_{i=1}^{n} \psi_i \mathbf{q}_i(t) = \Psi\mathbf{q}(t), \tag{5}$$

where $\Psi \in \mathbb{R}^{n \times n}$ contains the POMs.

Assuming that measured locations are where physical sensors are located, and unmeasured locations are where system response will be predicted, $\mathbf{u}(t)$ and Ψ from Eqn. (5) can then be presented as follows:

$$\mathbf{u}(t) = \begin{bmatrix} \mathbf{u_m}(t) \\ \mathbf{u_e}(t) \end{bmatrix}; \quad \Psi = \begin{bmatrix} \Psi_m \\ \Psi_e \end{bmatrix}, \tag{6}$$

where: $\mathbf{u_m}(t) \in \mathbb{R}^{n_m}$ is the measured response; $\mathbf{u_e}(t) \in \mathbb{R}^{n_e}$ is the estimated response at unmeasured location; and $\Psi_m \in \mathbb{R}^{n_m \times n_m}$ and $\Psi_e \in \mathbb{R}^{n_e \times n_e}$ are the POMs for measured and estimated responses, respectively.

Structural response of a linear system can be written as a linear combination of its left singular vectors. In case first few left singular vectors would be kept in the analysis, the reconstructed response would

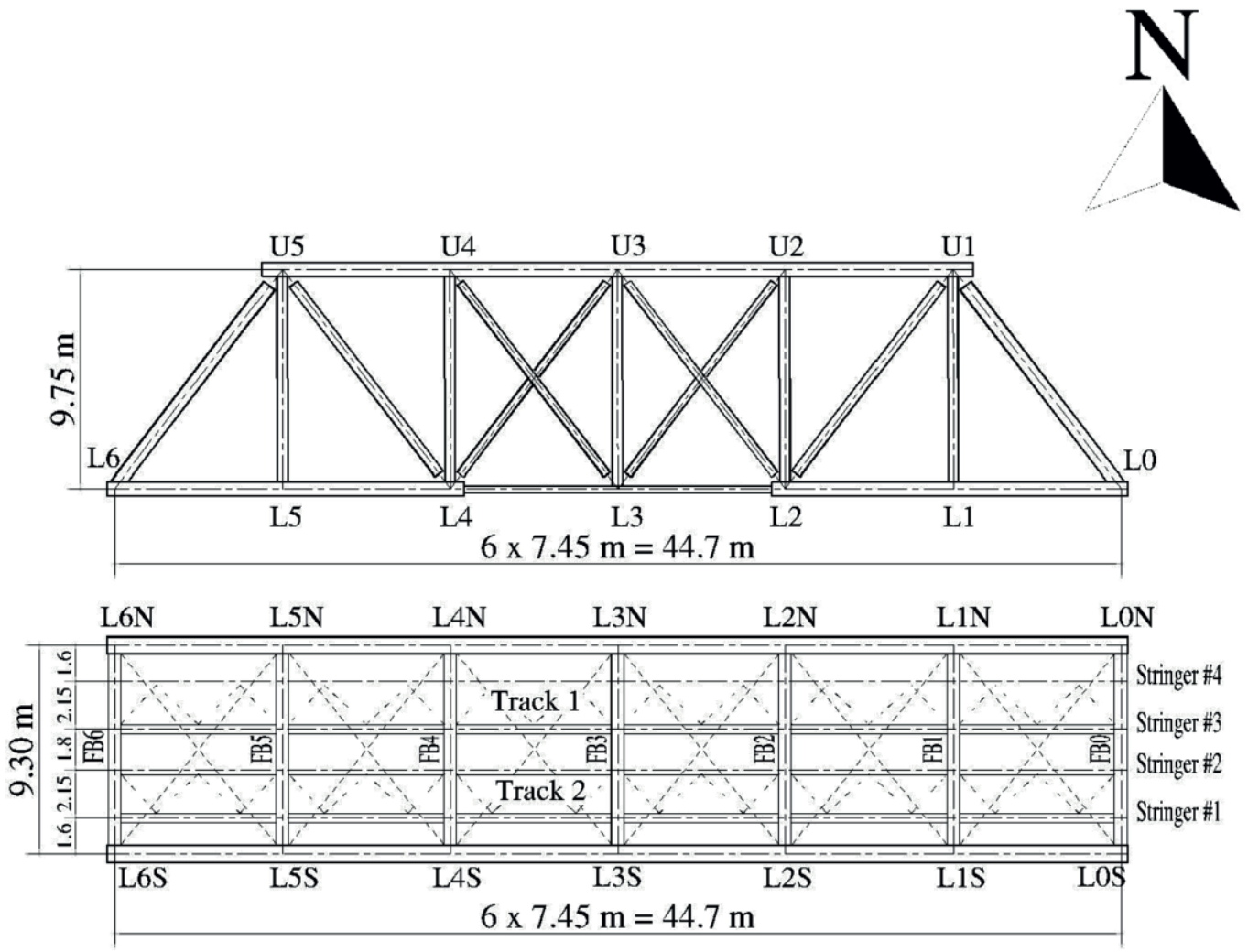

Fig. 2. Truss span elevation and plan views.

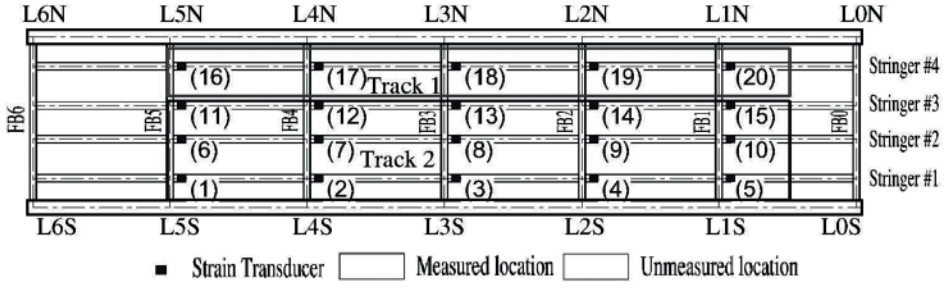

Fig. 3. Truss floor system instrumentation plan.

be an approximation. Using Eqn. (6), Eqn. (5) can further be expressed as follows:

$$\mathbf{u}(t) = \begin{bmatrix} \mathbf{u_m}(t) \\ \mathbf{u_e}(t) \end{bmatrix} = \begin{bmatrix} \Psi_\mathbf{m} \\ \Psi_\mathbf{e} \end{bmatrix} \mathbf{q}(t). \tag{7}$$

Hence, Eqn. (7) can further be expressed as shown below [22]:

$$\widehat{u}_e(t) = \Psi_e \widehat{q}(t) = \Psi_e \left(\Psi_m^T \Psi_m\right)^{-1} \Psi_m^T u_m(t) \tag{8}$$

For further details on using SVD for virtual sensing see [22]. (See Fig. 1)

3. SVD strain estimation: Application to a railway steel bridge

3.1. Monitored railway bridge

The studied bridge is a functioning open deck, double-track, multi-span truss and through girder riveted steel railway bridge in central Nebraska. One major concern for many riveted steel railway bridges is fatigue of stringer to floor-beam connections, which can result in connection and, potentially, structural failure. The truss span under investigation is simply supported and has a total length of 44.7 m. The span consists of seven floor beams supporting four equally spaced stringers at 2.15 m center to center. Fig. 2 shows the span plan and

Fig. 4. Partial cross section of stringers and floor beam system with installed strain transducers.

Fig. 5. Installed strain transducer on stringer bottom flange.

elevation. Upper and lower chord panel points are labelled U1-U5 and L0-L6, respectively, in the figure. Lower chord panel points on the north and south trusses are represented as L*X*N and L*X*S respectively and FB*X* denotes a floor beam between the same lower chord panel points. More details about the bridge can be found elsewhere [34].

A sensitivity analysis was performed before instrument placement

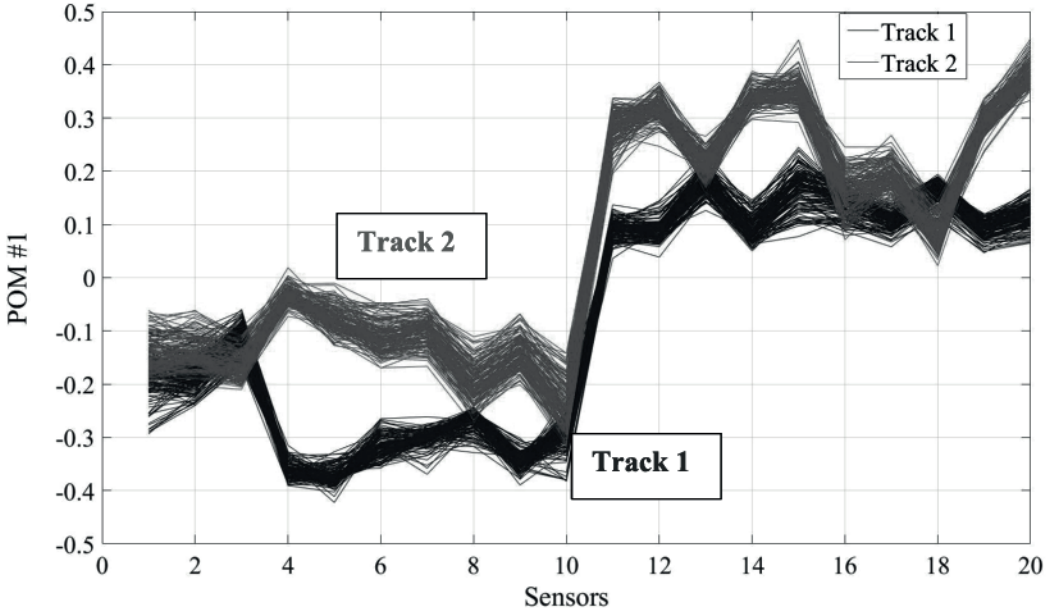

Fig. 6. RMS categorization of 300 train passages POMs based on transverse location.

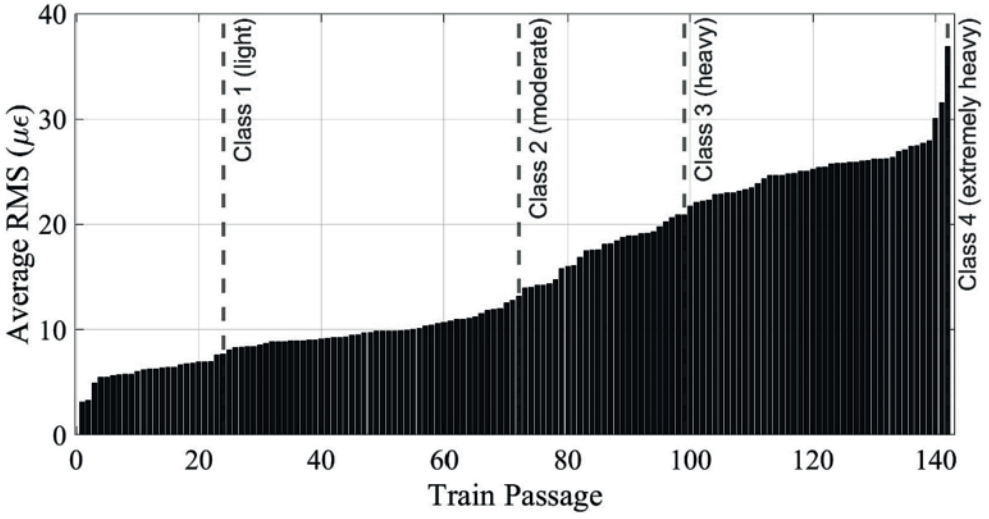

Fig. 7. Average RMS sub-categorization using k-means clustering (Track 1).

and 20 strain transducers were identified as the minimum number needed to detect stringer to floor beam connection defects [34]. The monitoring system is comprised of a BDI data logger, a wireless base station, and 4-channel wireless nodes, with each connected to 4 strain transducers via cables. The strain transducer (popular known as ST350) is resistive-based, and its circuit includes a full Wheatstone bridge with 4 active 350 O strain gages specifically designed to provide high electrical output for a given strain magnitude. The wireless node serves as a data acquisition device that facilitate the transmission of data from the strain transducers for collection and viewing (on a computer) through the wireless network created by the base station. Six 24-volt batteries, recharged by two solar panels were employed to power the system. The system was configured to activate and record strains at a sampling rate of 50 Hz whenever a train crossed the bridge.

The examined truss span was instrumented with 20 BDI strain transducers located on stringer bottom flanges adjacent to floor beam locations. Fig. 3 details strain transducer location on the truss floor system. The blue and red boxes differentiate between strain transducers at measured (Sensors 1 – 15) and what were deemed unmeasured (Sensors 16–20) locations, respectively, as discussed in Section 3.2.1. Fig. 4 shows a cross section of the stringer and floor beam system with installed strain transducers while Fig. 5 focuses on one of the strain transducers at the stringer bottom flange.

3.2. Output classification

Data-driven strain prediction method are solely dependent on measured data and, as a result, calibrated FE models are not necessary. Since moving train location, load, and speed variability can lead to variations in POMs, which can significantly impair data-driven strain prediction accuracy [35], it is therefore essential to first classify bridge inputs (e.g., moving loads) prior to strain estimation to minimize prediction variations. It may be difficult to directly determine moving train load magnitudes and locations and, as a result, categorization of train passages based on measured response could be beneficial. This study investigated two methodologies for categorization a single week dataset of 300 train passages into groups having similar loads and traveling across the bridge in the same direction. The first methodology was k-

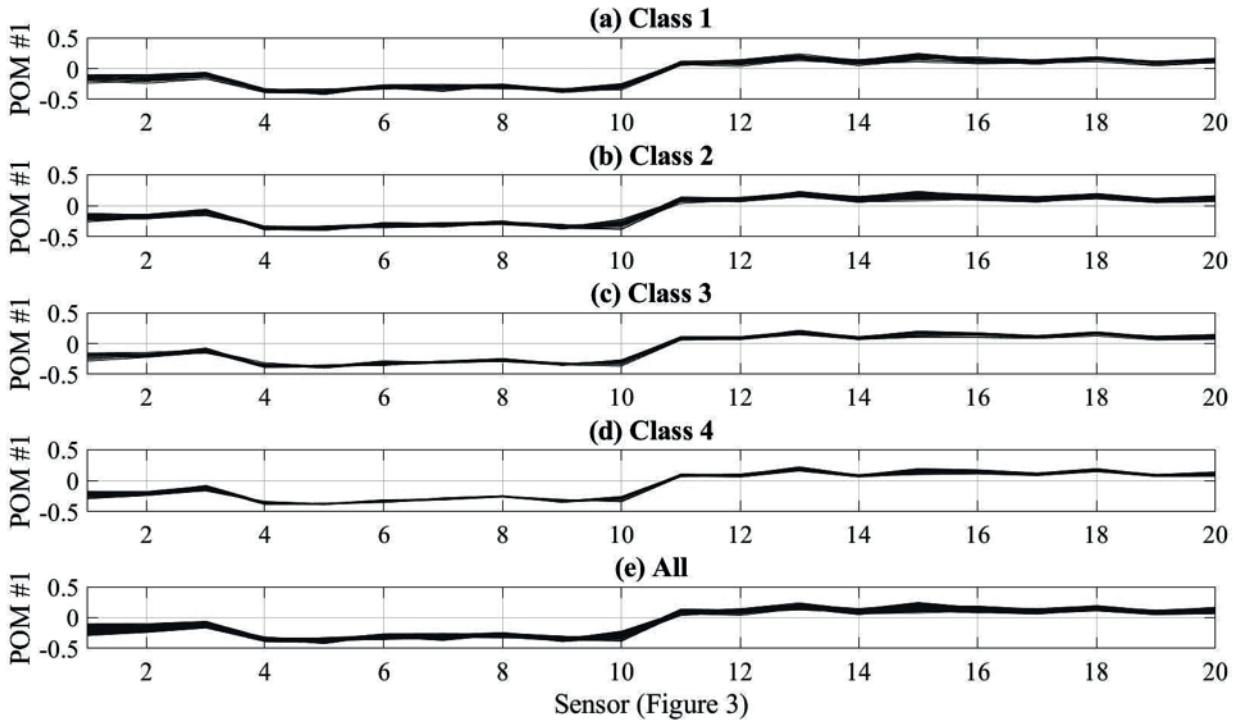

Fig. 8. Track 1 POMs for sub-categorized train passages classes using k-means clustering.

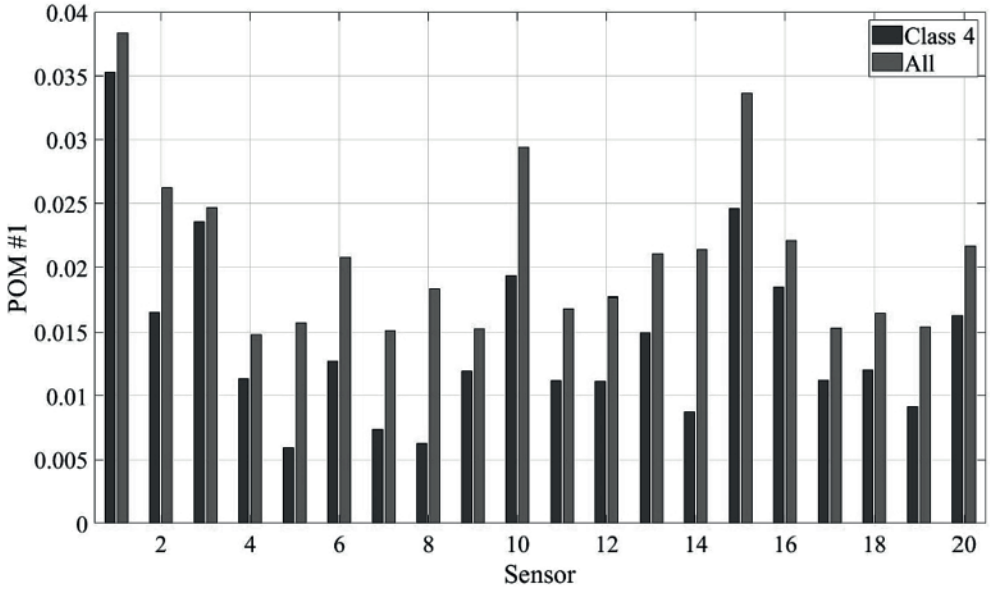

Fig. 9. Standard deviations, POM 1, all trains and Class 4 for Track 1 (Fig. 7).

means clustering [23] and peak picking using strain time-history RMSs for each passage. The second methodology performed clustering using SOMs of neural network clustering [24] from POMs of train passage snapshot matrices.

3.2.1. Classification using k-means clustering and RMS

Azam et al. [35] showed there is a strong correlation between statically equivalent uniform loads (i.e., input) and RMSs of resulting strain time-histories for each passage (i.e., output). The higher the loads, the higher the strain and RMS for each strain signal. This relationship was explored to help determine train location for each passage. RMSs from sensors on the south side of each track (i.e., Sensors 1 –5 on the south side of Track 2, Sensors 11 – 15 on the south side of Track 1 in Fig. 3) were compared for each train passage. It was shown that the train is on the track whose sensors had the highest RMS value. Fig. 6 shows snapshot matrix POMs from 300 train passages and demonstrates clear distinction between POMs generated from trains on each track. Track 1 had a total of 142 train passages while Track 2 had a total of 158 passages. It should be noted that POMs have not unit, and the first mode or column of POM's unitary matrix was utilized, hence the reason for the nomenclature POM #1 on Fig. 6 and other corresponding figures. Y-axis range −0.5 to 0.5 was employed since POMs values were within the range and for easy comparison. Classification based on RMS was utilized to categorize train passages based on track location and consequently

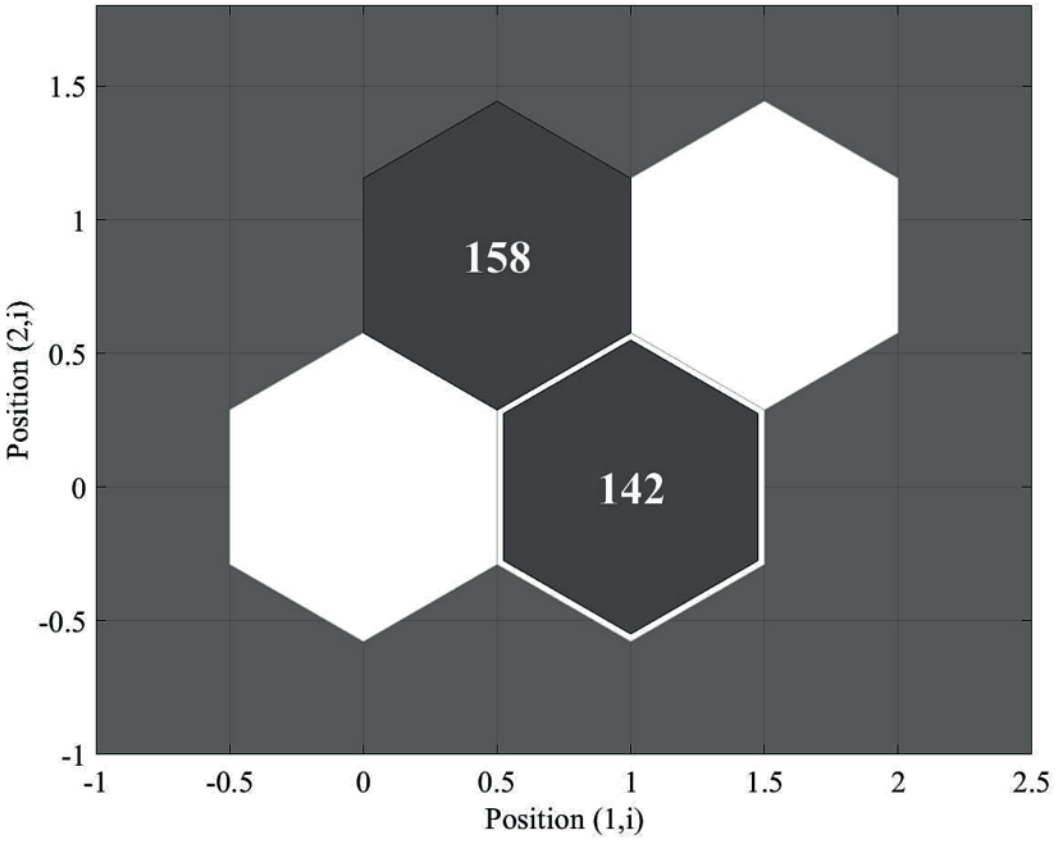

Fig. 10. SOM sample hits from neural network clustering of 300 train passage POMs.

reduced POM dispersion due to train location as shown in Fig. 6.

To further reduce POM dispersion, the RMS of each track snapshot matrix was subcategorized using k-means clustering. K-means clustering is a popular unsupervised machine learning approach that divides a set of n observations into k clusters by grouping observations with similar underlying patterns [23]. Grouping is achieved by assigning every observation to its nearest cluster such that the sum of squared distances between the observations and the cluster's centroid (i.e., arithmetic mean of all data points that belong to the cluster) is minimized. Each observation can only be assigned to one cluster, and the less variation there is within each cluster the more similar the data points.

Mean RMSs for all sensors for each train passage were determined, sorted, and sub-grouped using k-means clustering. Train passages for each track were arbitrarily sub-categorized into four classes to represent the different possible categories of train loads (i.e., "light," "moderate," "heavy," and "extremely heavy") using k-means clustering. Class 1 represented "light" train loads while Class 4 denoted "extremely heavy" loads. Fig. 7 shows resulting sub-categorizations of average snapshot matrix RMSs for train passages on Track 1.

Additional sub-categorization of train passages on each track into the four classes further reduced POM dispersion. Fig. 8 shows POMs in each class after k-means clustering for Track 1. This k-means clustering resulted in additional reduction in POM variability, especially for Classes 3 and 4, when compared to all POMs. Fig. 9 shows reduction in POM variation for Class 4 train passages when compared to that for all train passages on Track 1. Strain estimation was subsequently achieved by employing measured strain snapshot matrix POMs from select passages in each class to predict time-histories in the same class at designated unmeasured locations using Eqn. (8).

3.2.2. *Classification using self organizing maps and POMs*

Several neural network architectures have been designed for clustering analysis. SOM is a promising technique [24] that uses a set of connected neurons and computes an input pattern's Euclidean distance to each neuron. The neuron with the weight vector closest to the input pattern is considered the winning neuron, and its weight vector is integrated into the input pattern for that neuron and its neighboring neurons [36]. As a result, input data with similar patterns cluster together in the same neuron.

SOM performs clustering while preserving topology [37]. It is also useful for vector quantification, feature extraction, and data visualization [38]. SOM is particularly useful for visualizing high-dimension data as it converts complex nonlinear statistical relationships into simple geometric relationships in a low-dimension display. Complex information processing systems can further be decomposed into a series of simpler subsystems using the SOM [39].

For the current study, SOM was implemented using the Neural Network Clustering App in MATLAB [40]. The clustering app assists with creation, visualization, and training of SOM networks. The SOM has one layer with neurons organized in a grid and learns to cluster data based on similarity. Data was imported, and a neural network was defined and trained. To define the network, the two-dimensional map of size n, which corresponds to an output of clusters n^2 (i.e., 'n' rows by 'n' columns), is specified. This implies that when a map size of n = 1, 2, 3, …, k is defined, it corresponds to an output of 1, 4, 9, …, k^2 number of clusters, respectively. A minimum map size of 2 is required to produce more than one cluster.

POMs from snapshot matrix of 300 strain time-histories were imported with a network and map size of 2 being defined. As a result, observations (POMs) were categorized into output having 4 clusters. The number of POMs associated with each neuron (cluster) as well as neuron locations in the topography are captured via SOM sample hits. Fig. 10 shows resulting SOM sample hits from the clustering analysis. The proper orthogonal modes were largely categorized into two hits (i.e.,

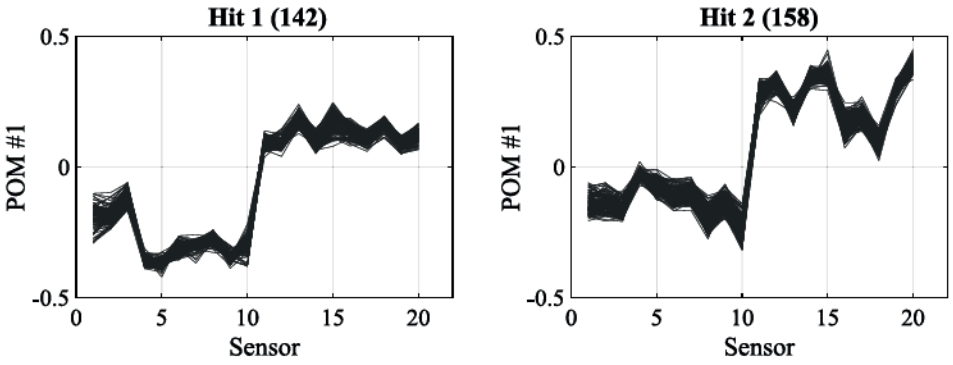

Fig. 11. Clustered POMs from 300 train passages.

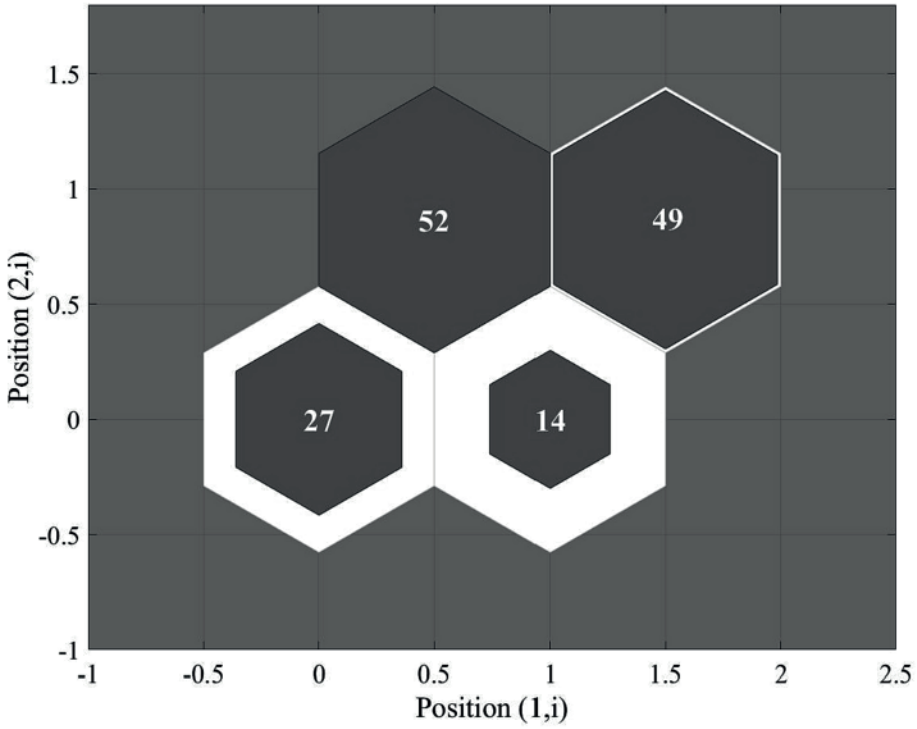

Fig. 12. SOM sample hits from Neural Net clustering of 142 train passage POMs (Track 1).

158 and 142 train passages in each), leaving the other hits empty. This indicates that the input data (train passages) can largely be grouped initially into two groups. Results from this categorization follow similar trends to RMS categorization for Tracks 1 and 2 by showing that each moving train is in one of the two locations (i.e., either Track 1 or 2).

Fig. 11 displays POMs from each populated sample hit and demonstrates that this clustering analysis is equivalent to classification of train passages based on transverse location on the bridge. Hits 1 and 2 correspond to train passages on Tracks 1 and 2, respectively, when compared to categorization using strain signal RMS. This initial clustering analysis reduced POM dispersion due to train passages on both tracks.

To further reduce POM variability caused by differing train loads and speeds, Hit 1 from Fig. 11 was subsequently subcategorized into four classes using a map size of 2. In similar fashion to RMS classification, four classes were chosen to represent different possible categories of train loads (i.e., "light," "moderate," "heavy," "extremely heavy") for train passages on each track. It was difficult to independently ascertain which of the classes corresponded to categories of train load shown in Fig. 7 from this clustering analysis. Resulting clustered hits and corresponding POMs are shown in Fig. 12 and Fig. 13, respectively. Strain estimation at unmeasured locations in the each class from Fig. 13 was then performed using measured strain snapshot matrix POMs from select passages in the same class.

A flow chart summarizing both classification approaches is presented in Fig. 14.

4. Results and discussion

4.1. Quality measurements

To assess strain estimate quality from the two clustering approaches and SVD against the actual response obtained during field monitoring at the same locations, three measures were introduced: normalized fatigue error; TRAC; and coefficient of determination (R^2). The methods are summarized, and their selection justified in the following sections.

4.1.1. Normalized fatigue error

One important potential application for strain estimation is fatigue assessment, as fatigue life largely depends on corresponding stress

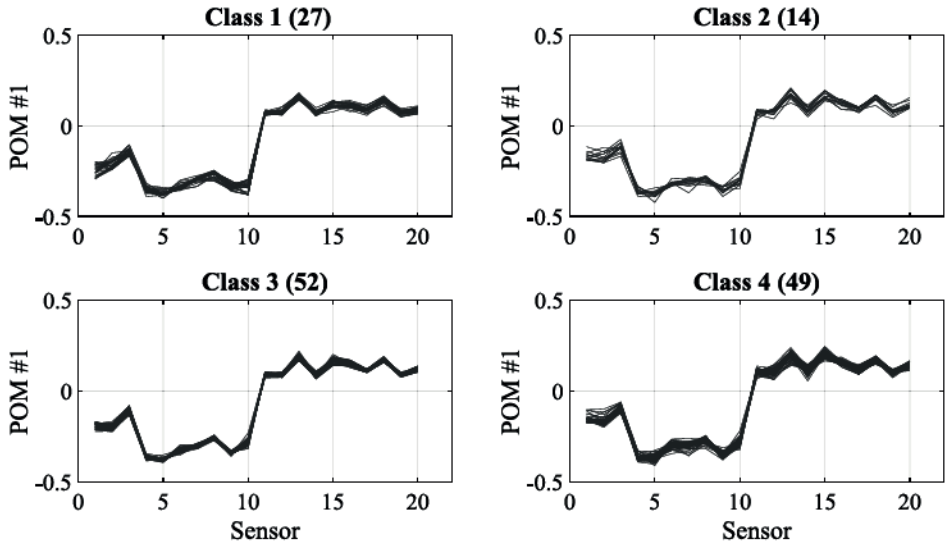

Fig. 13. Clustered POMs from 142 train passages of Track 1.

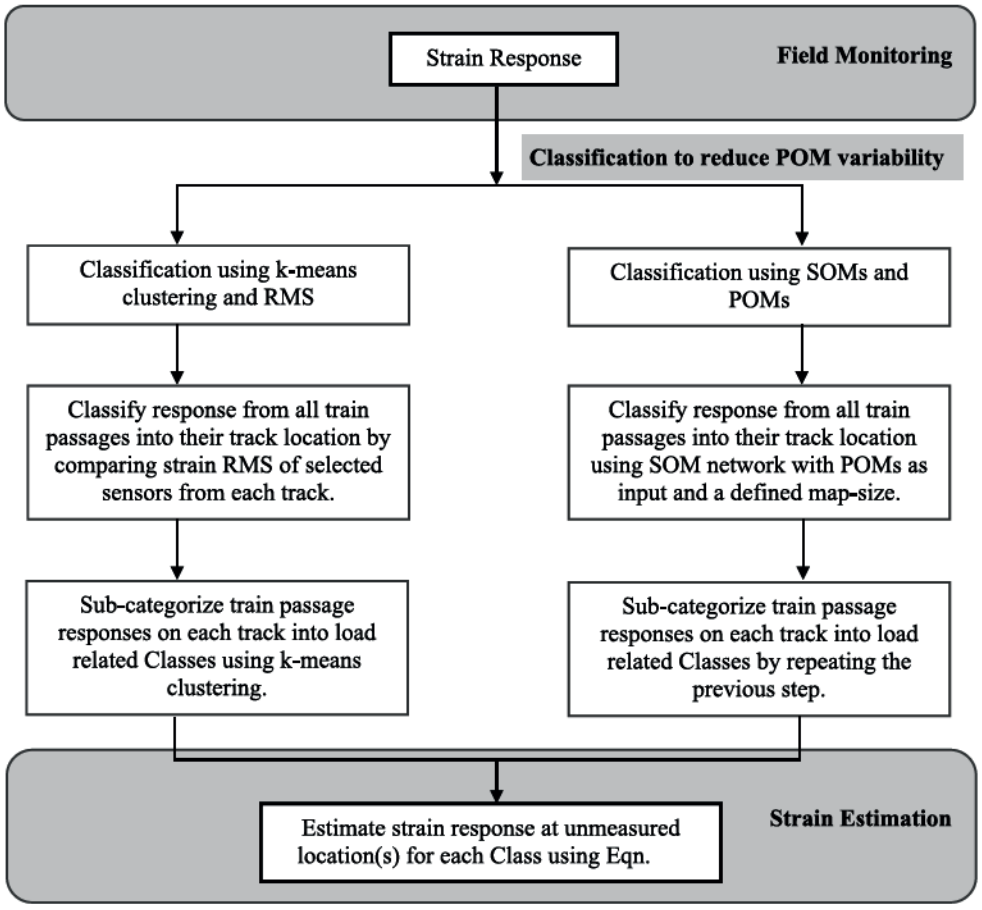

Fig. 14. Strain estimation process using k-mean clustering or SOMs and POMs.

ranges and frequencies [41,42]. It is thus vital to compare estimated and measured strains utilized in cycle counting that helps determine fatigue life. This can be achieved using the normalized fatigue error, which measures fatigue based on strain response from an estimated and measured signal [1].

Use of the Palgreen-Miners Rule [43] and exclusion of the effects of bilinear S-N curves allows fatigue damage to be expressed as follows [1]:

$$D_i = \sum_{j=1}^{n_{cycles}} \frac{1}{N_j} = \sum_{j=1}^{n_{cycles}} \frac{\Delta\sigma_j^m}{C} = \frac{E^m}{C} \sum_{j=1}^{n_{cycles}} \Delta\varepsilon_j^m, \quad (9)$$

where: D_i = cumulated fatigue damage at the i^{th} fatigue location; n_{cycles} = number of counted cycles; N_j = number of cycles for fatigue failure at the given stress range, $\Delta\sigma_j$; C = fatigue capacity (i.e., S-N curve log(N)

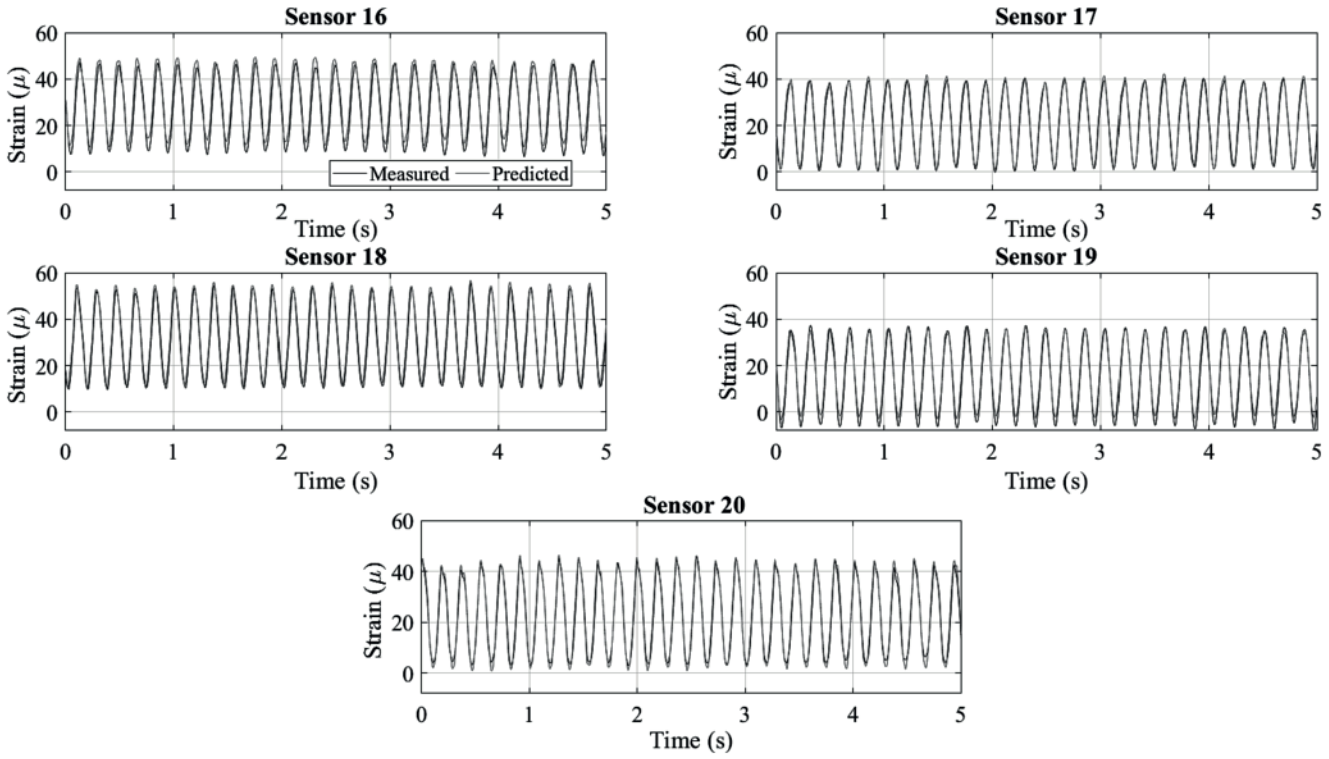

Fig. 15. Classification using RMS: predicted and measured strain time-histories for first train passage, Class 4 ("extremely heavy" train passage, Fig. 3).

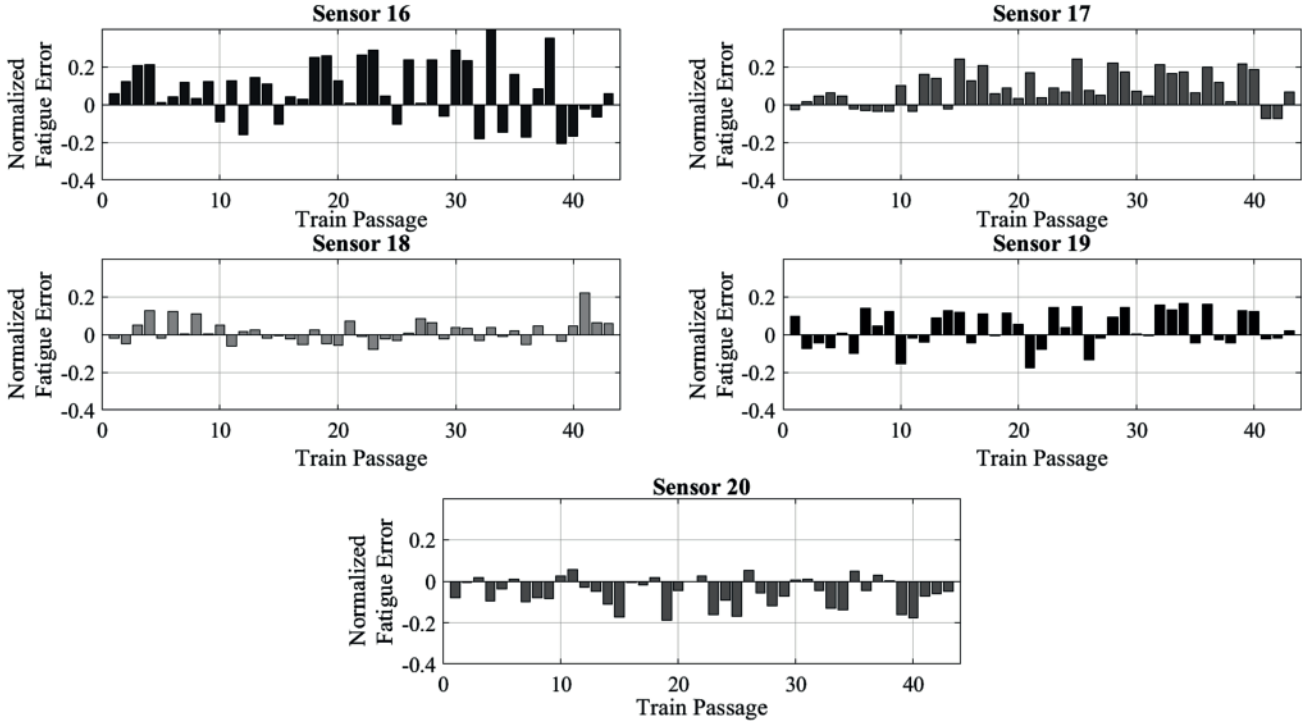

Fig. 16. Classification using RMS: normalized fatigue error, Class 4 ("extremely heavy" train passage, Fig. 3).

axis intercept); m = "slope" of the S-N curve; E = modulus of elasticity; and $\Delta\varepsilon_j$ = strain range from cycle counting.

Tarpo et al. [1] formulated normalized error between estimated and measured fatigue damage as:

$$\eta_i = \frac{\widehat{D}_i - D_i}{D_i} = \frac{\sum_{j=1}^{n_{cycles}} \Delta\widehat{\varepsilon}_j^m}{\sum_{j=1}^{n_{cycles}} \Delta\varepsilon_j^m} - 1, \tag{10}$$

where $\eta_i = 0$ indicates a perfect estimation of strain.

A negative η_i indicates underestimation of fatigue damage and a

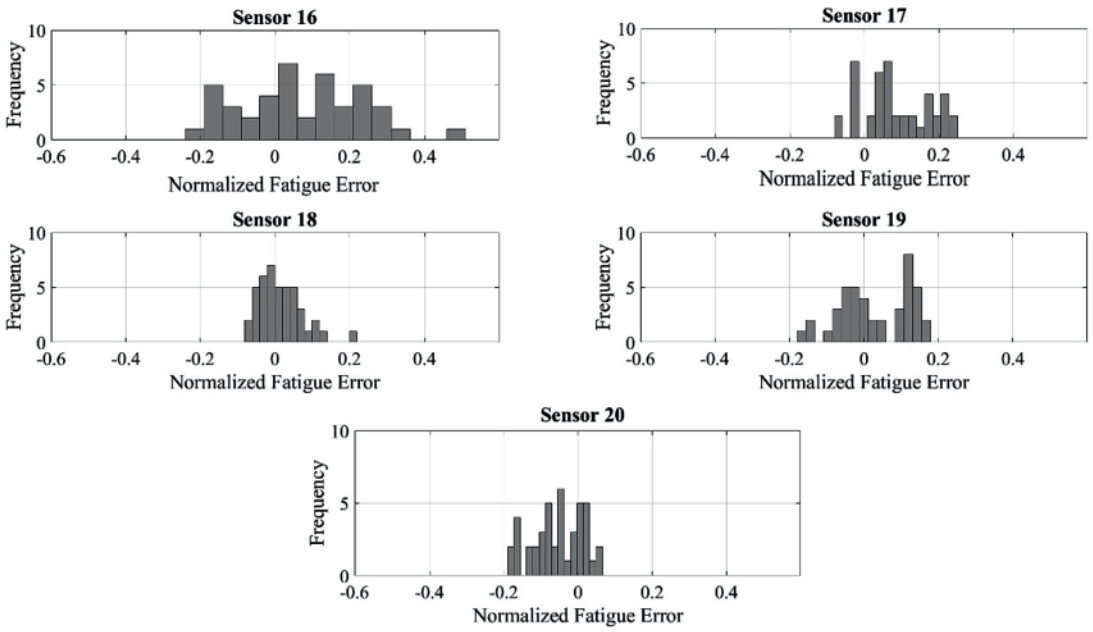

Fig. 17. Classification using RMS: normalized fatigue error histogram, Class 4 ("extremely heavy" train passage, Fig. 3).

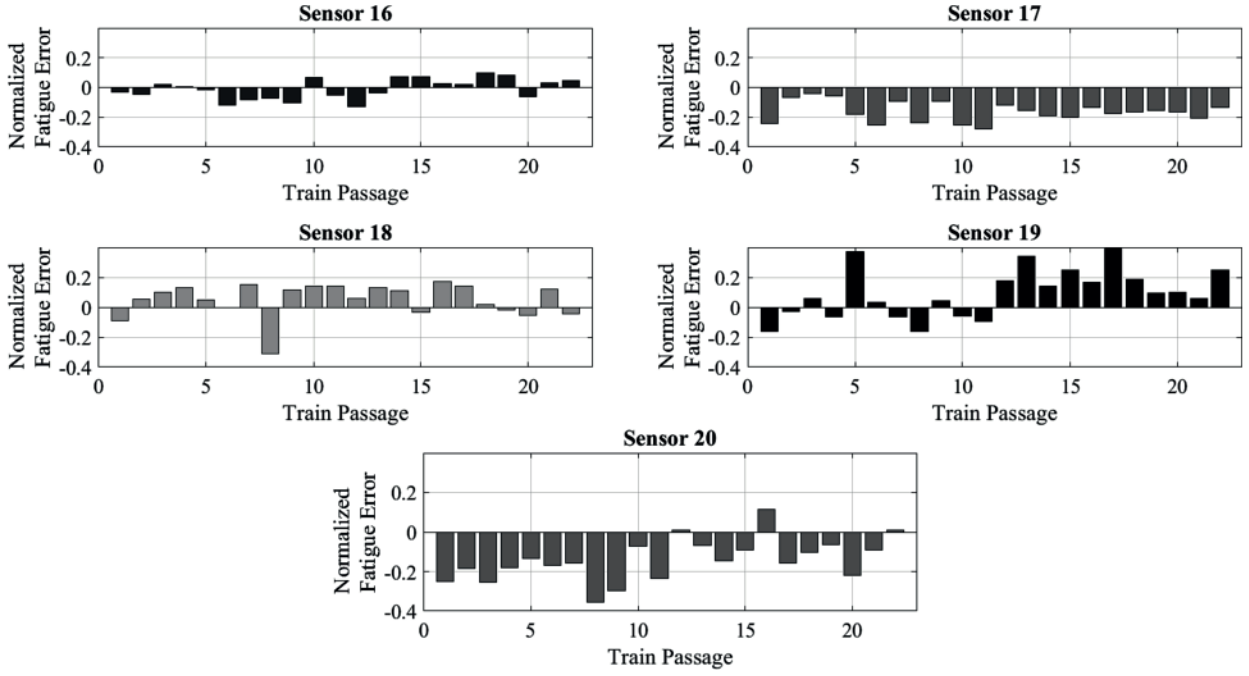

Fig. 18. Classification using RMS—normalized fatigue error for Class 1 ("light" train passage, Fig. 3).

positive η_i indicates damage overestimation.

4.1.2. Time response assurance criterion (TRAC)

TRAC is a measure of overall correlation between estimated and measured strain time-history signals [13] and is expressed as follows:

$$\text{TRAC} = \frac{(\varepsilon^{T}.\hat{\varepsilon})^2}{(\varepsilon^{T}.\varepsilon)(\hat{\varepsilon}^{T}\hat{\varepsilon})}, \tag{11}$$

where ε is the measured strain time-history; and $\hat{\varepsilon}$ is the predicted measured strain time-history.

TRAC is a popular quality assurance quantitative measurement and was utilized to compare and examine whether measured and estimated strain time-history signals had similar general shape and trends [15]. TRAC values are positive, real quantities between 0 and 1 and are independent of amplitude differences [1]. Values close to unity indicate strong agreement between strain time-histories, while values close to 0 indicate minimal or no similarity [44]. However, TRAC does not consider amplitude differences, but was however considered for reader (s) to see since this is a commonly used assurance quantitative measurement metric [1].

4.1.3. Coefficient of determination (R^2)

An additional quality measurement, the coefficient of determination (R^2), is employed to compensate for TRAC's limitation associated with accounting for amplitude differences. The coefficient of determination is

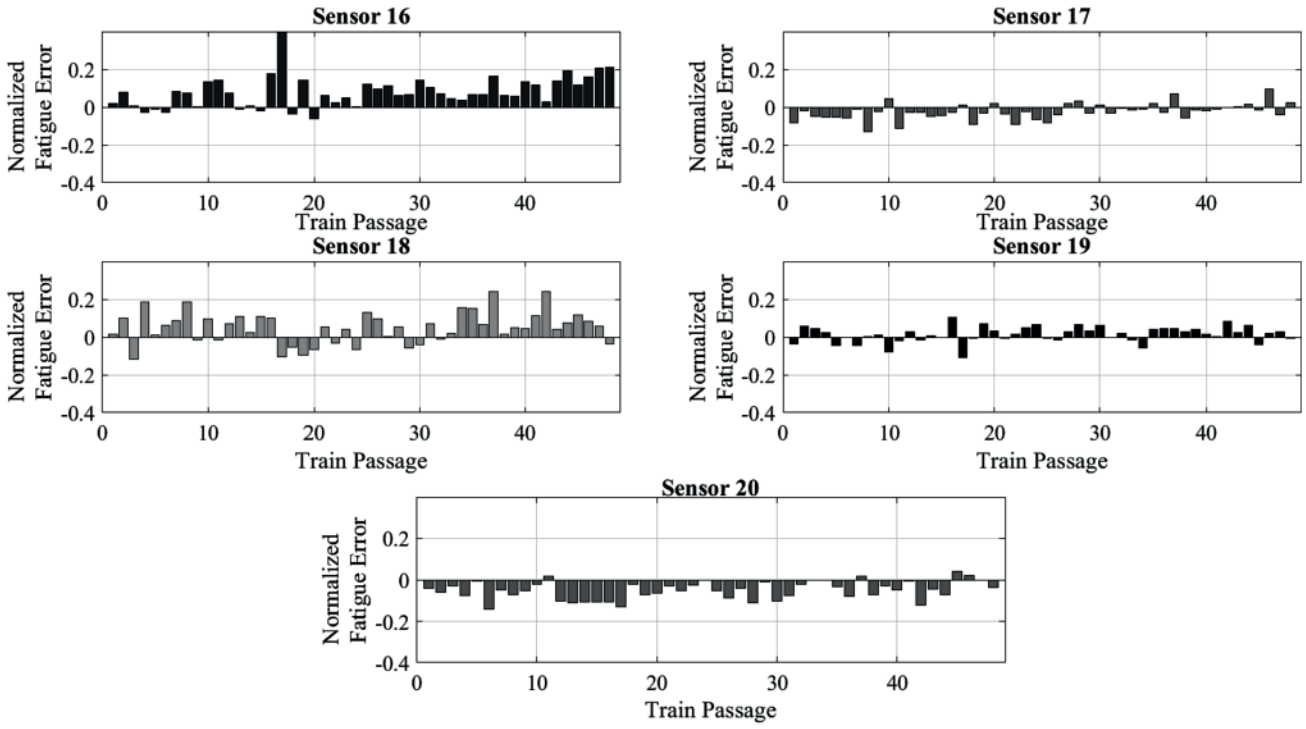

Fig. 19. Classification using RMS—normalized fatigue error for Class 2 ("moderate" train passage, Fig. 3).

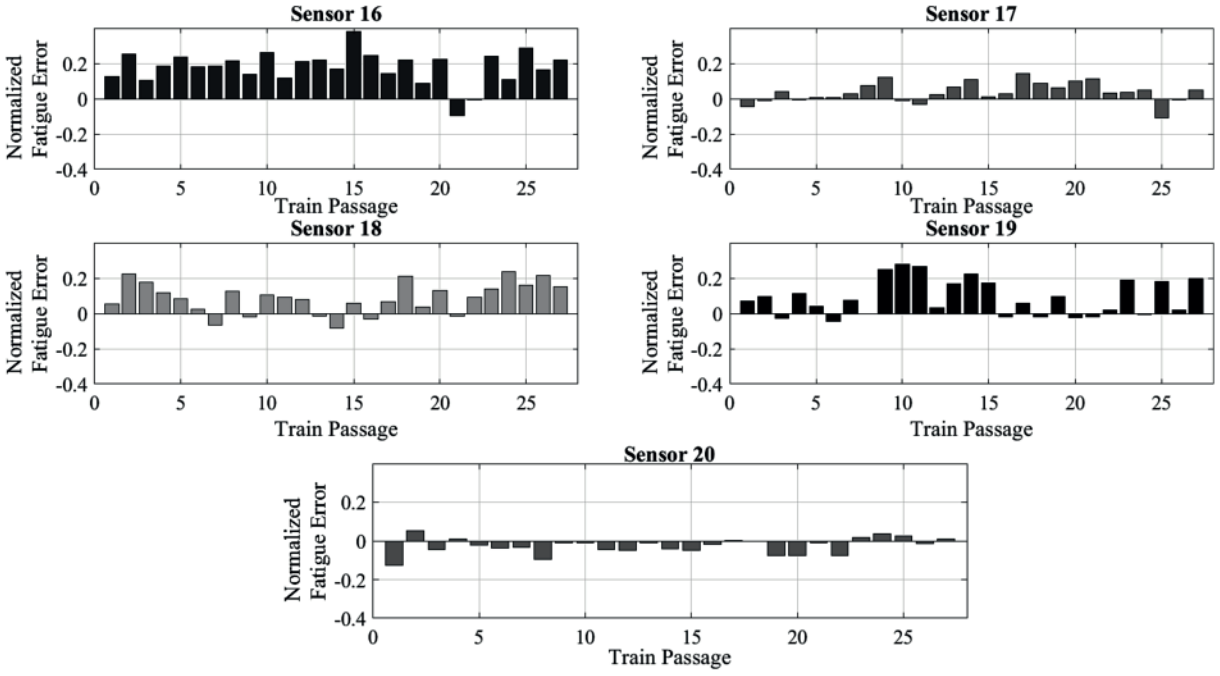

Fig. 20. Classification using RMS—normalized fatigue error for Class 3 ("heavy" train passage, Fig. 3).

mean squared error normalized with respect to the variance of the measured strain signal and is frequently used for model validation [45]. As a result, it considers both amplitude discrepancies and overall amplitude [46] and is defined as follows:

$$R_i^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \widehat{\varepsilon}_i)^2}{Var(\varepsilon_i)}, \tag{12}$$

where ε_i is the measured strain time-history; and $\widehat{\varepsilon}_i$ is the predicted measure strain time-history. When the coefficient of determination is 1, amplitudes are perfectly correlated.

4.2. SVD estimated strain

After clustering strain time-histories into classes to minimize POM dispersion due to variation in train speeds, loads, and locations, strains were estimated for each group using data driven SVD. After multiple trials, the minimum number of strain signals employed as the training set for a selected clustered class was identified as 25 % of the total

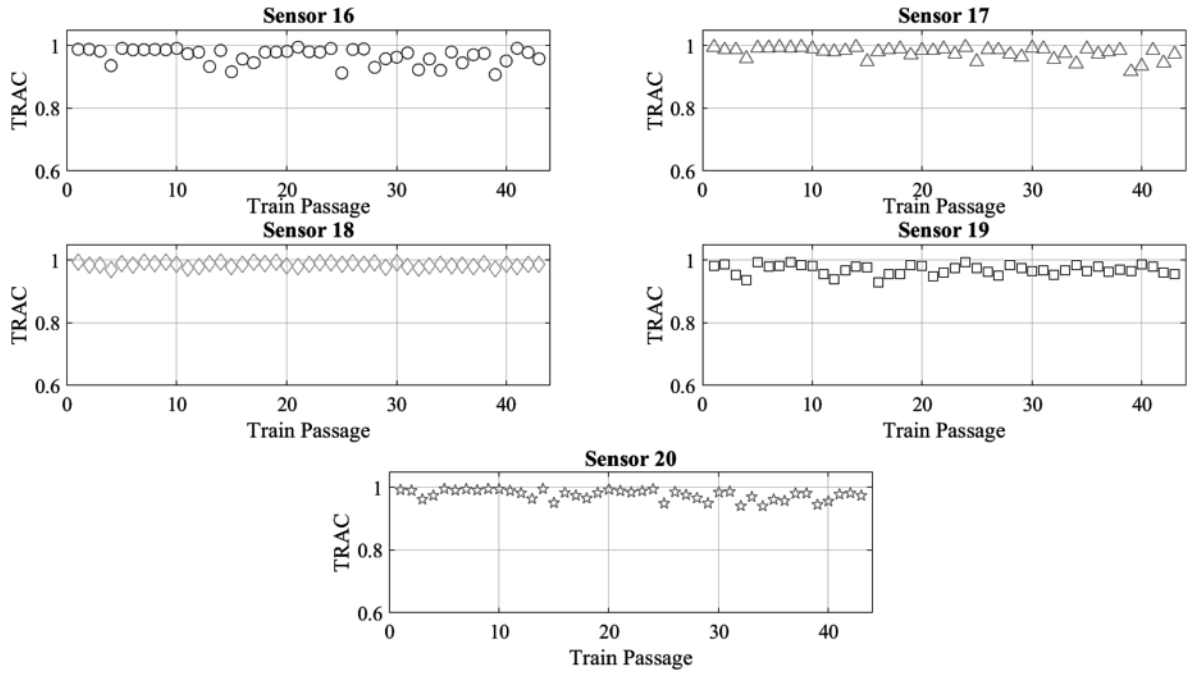

Fig. 21. Classification using RMS: TRAC values, Class 4 ("extremely heavy" train passage, Fig. 3).

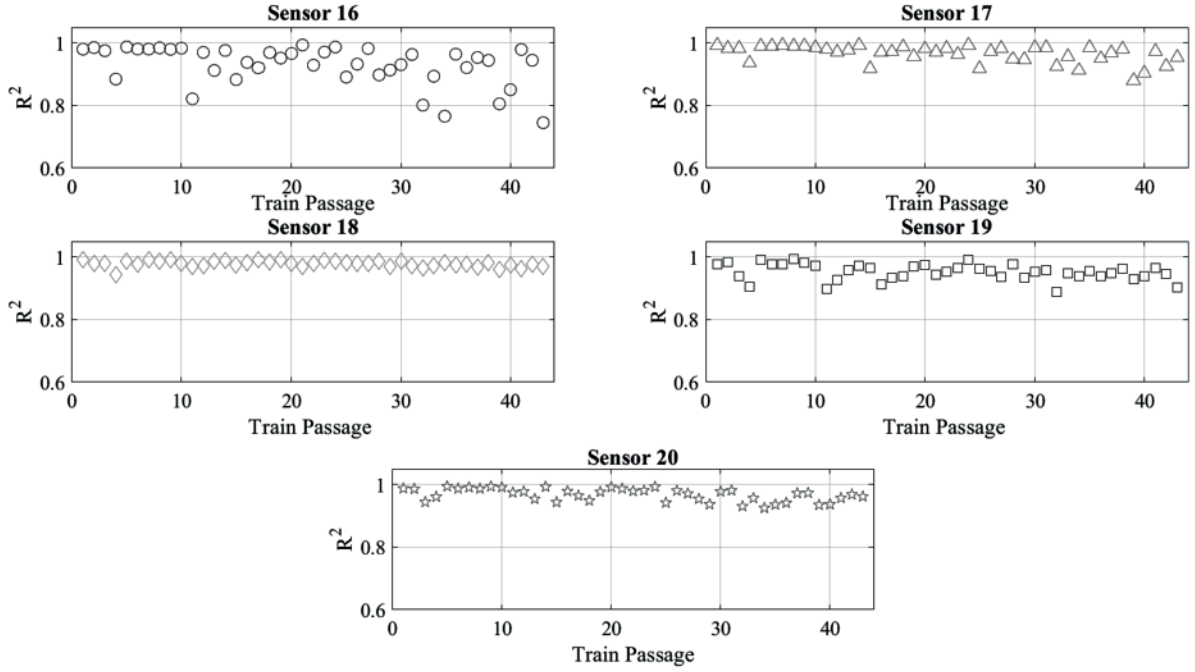

Fig. 22. Classification using RMS: R^2, Class 4 ("extremely heavy" train passage, Fig. 3).

number of signals. Training set strain time-histories include response at measured and unmeasured locations, and they were stored in a snapshot matrix. It was arbitrarily decided to sequentially use Sensors 1–15 as measured locations with remaining sensors being estimated and subsequently compared to actual measurements at those locations (Fig. 3).

Training set POMs were calculated from SVD of the snapshot matrix and were subsequently used in Eqn. (8) to determine predicted signals for subsequent train passages at unmeasured locations. It was observed that strain estimation accuracy using SVD largely depended on the quality of the entire monitored data set. For instance, in cases where sensors were not performing adequately, measured response could impair the accuracy of estimated strain. To avoid ill-conditioning, adequate left singular vectors were needed to calculate generalized coordinates. Trial-and-error was used and 10 left singular vectors were selected. Representative results from high load categories (i.e., Classes 3 and 4) are presented in the following sections since maximum stress ranges are produced for higher loads.

4.2.1. Estimated strains using RMS, classification by k-means clustering

Figs. 15 to 22 show strain estimation results from k-means clustering classification using RMS. Fig. 15 shows a portion of the estimated and measured strain signals at Sensors 16–20 (unmeasured locations) for the

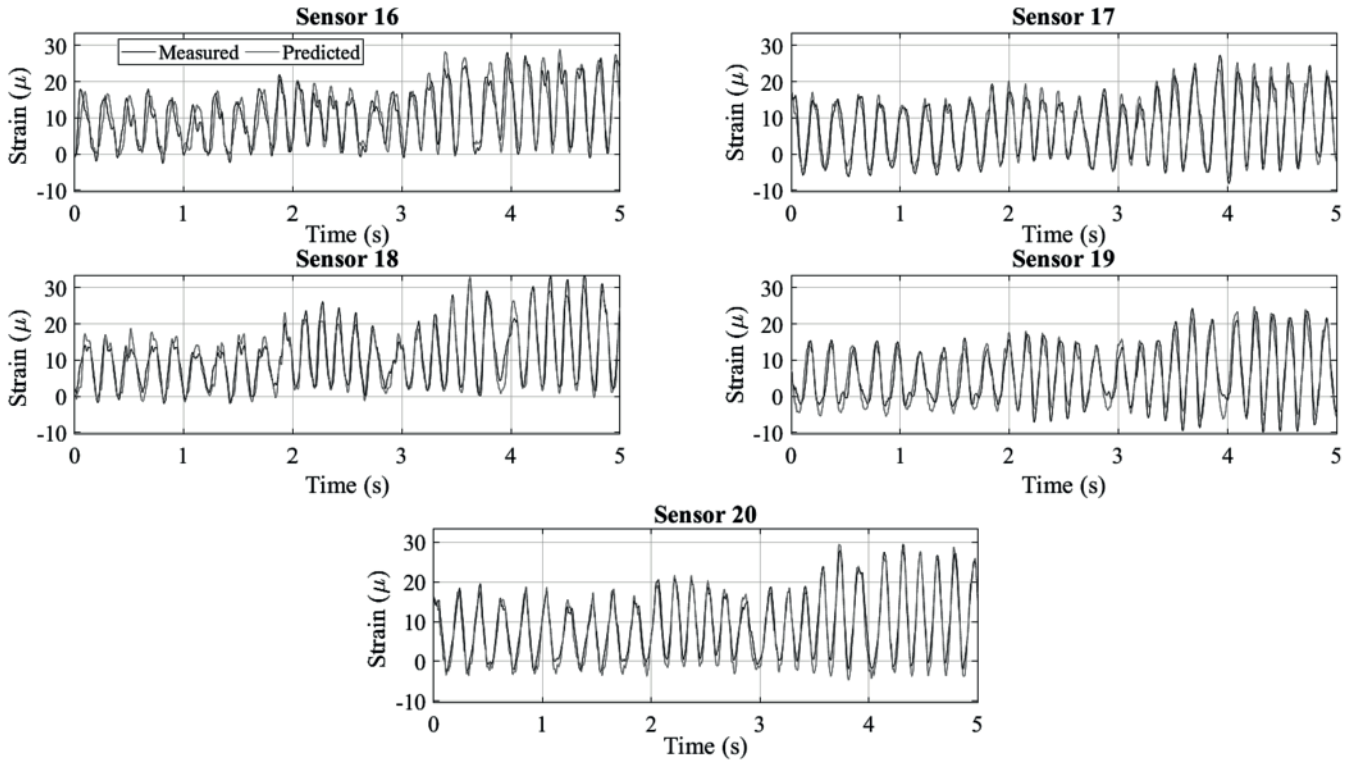

Fig. 23. Classification using POMS—predicted and measured strain time-histories for first train passage in Class 3 ("heavy" train passage, Fig. 3).

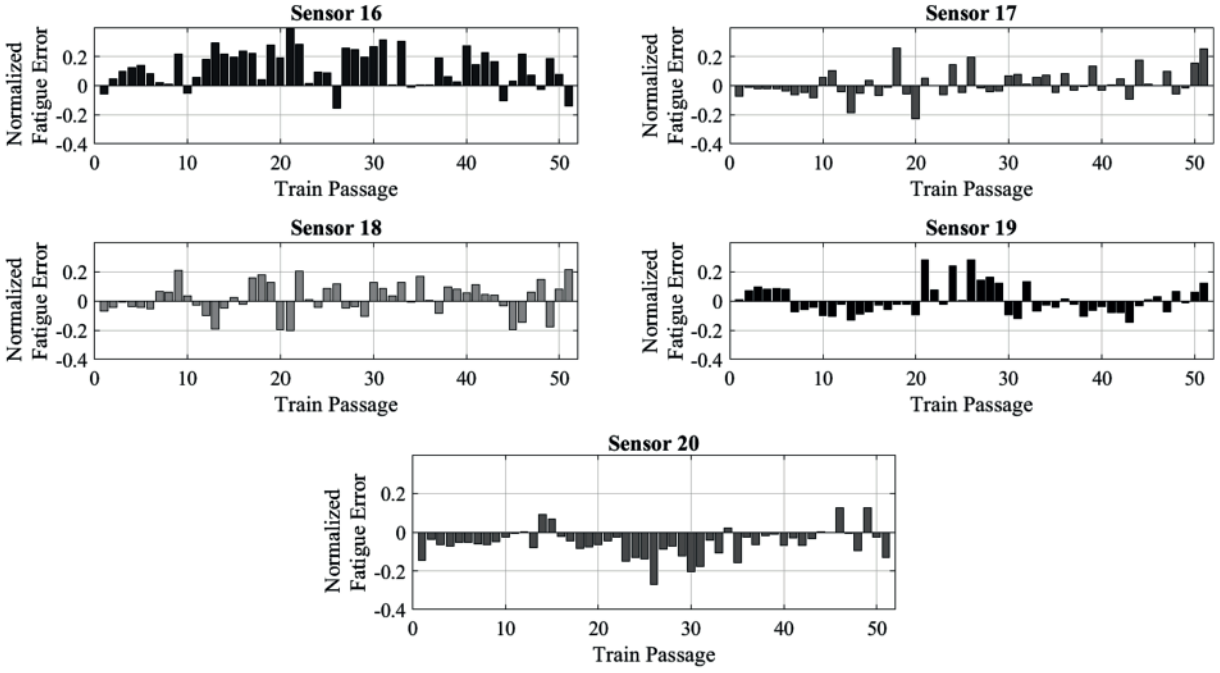

Fig. 24. Classification using POMS—normalized fatigue error for Class 3 ("heavy" train passage, Fig. 3).

first Class 4 ("extremely heavy") train passage. Estimated and measured signals match quite well at each location. However, some minimal amplitude differences are observed at Sensors 16 and 19. Quality measurements are subsequently shown in Fig. 16 and Fig. 22. Fig. 16 details normalized fatigue error for all Class 4 train passages at each unmeasured location. Normalized fatigue error at these locations ranges between 0.4 to −0.2, i.e., 40 % to −20 %.in terms of percentage. As presented earlier, results closer to 0.0 represent good estimates. For Sensors 18–20, errors are substantially reduced, ranging between 0.2 and −0.2 (i.e., 20 % to −20 %). The range and frequency of the normalized fatigue error are shown in the histogram in Fig. 17. Normalized fatigue errors for Classes 1–3 are shown in Figs. 18 to 20. Normalized fatigue error again ranges between 0.4 and −0.4 (40 % and −40 %) for all train passages in Classes 1–3, though high but considered to be an appreciable improvement over reported error ranges (1.5 to −0.5, i.e., 150 % to −50 %) for an offshore platform using ME [1]. The

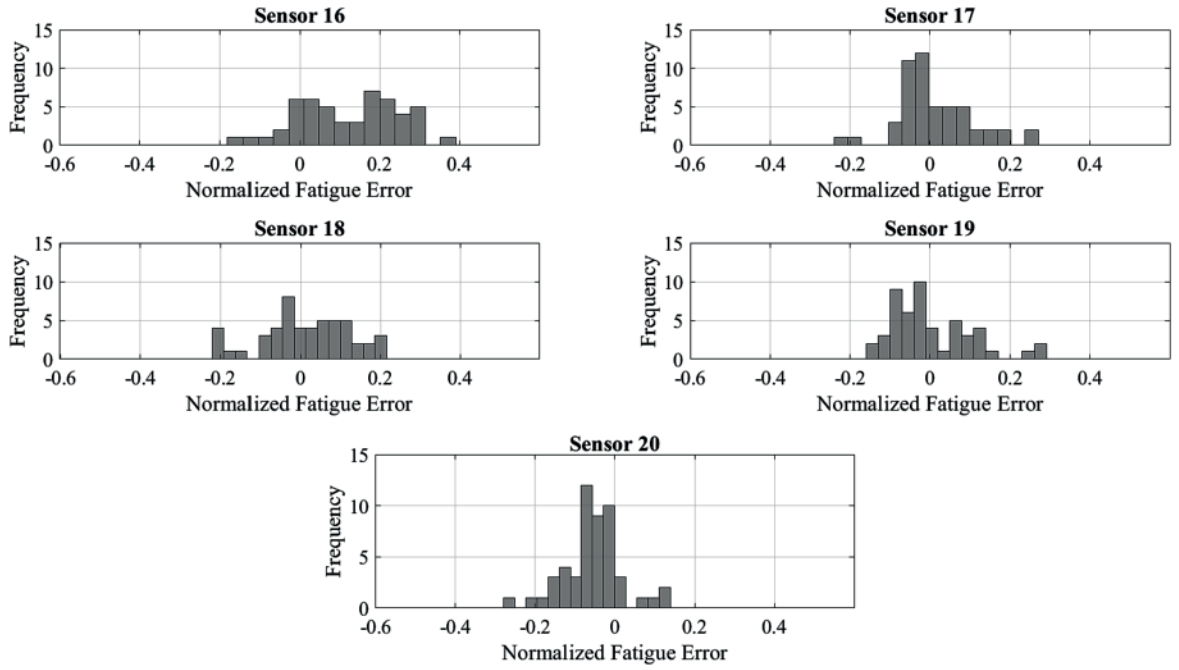

Fig. 25. Classification using POMS—histogram of normalized fatigue error for Class 3 ("heavy" train passage, Fig. 3).

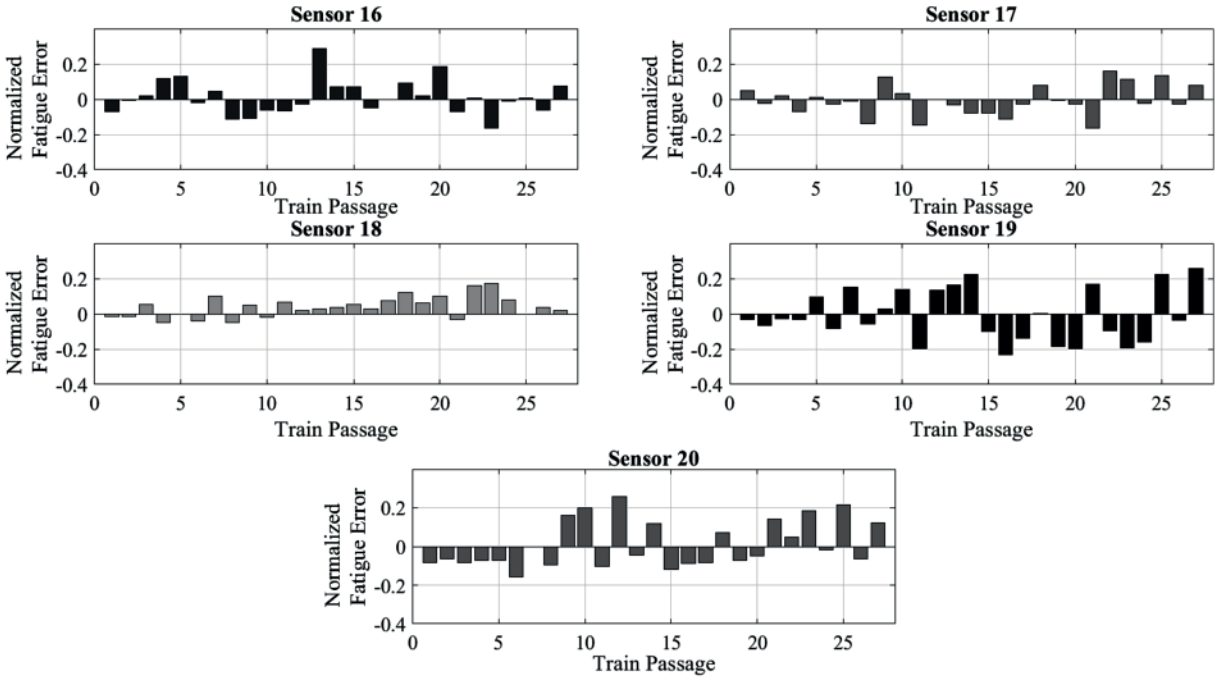

Fig. 26. Classification using POMS—normalized fatigue error for Class 1 ("light" train passage, Fig. 3).

normalized fatigue error can potentially be further reduced by removing environmental variability in POMs in subsequent follow up study. However, no environmental data was collected during the field monitoring. Correlation between estimated and measured signals was also examined using TRAC, as shown for Class 4 in Fig. 21. TRAC values between 0.9 and 1.0 were observed for all unmeasured location estimations, indicating strong correlation between estimated and measured strain signals. To account for both amplitude discrepancies and measured signal overall amplitudes, Fig. 22 presents calculated R^2 values for all Class 4 train passages at all unmeasured locations, with values ranging from 0.7 to 1.0. Higher R^2 values between 0.9 and 1.0, indicative of stronger correlation, are observed at Sensors 18 and 20. Fig. 19.

Low normalized fatigue errors for all classes and TRAC and R^2 results demonstrate that strain estimation accuracy with SVD improves using low dispersion POMs achieved via k-means clustering of strain time-history RMS.

4.2.2. Estimated strains using POMs, classification by SOM

Figs. 23 to 30 show results of strain estimation using POMs with SOM classification. Fig. 23 shows a portion of estimated and measured strain signal time-histories at unmeasured location Sensors 16–20 for the first

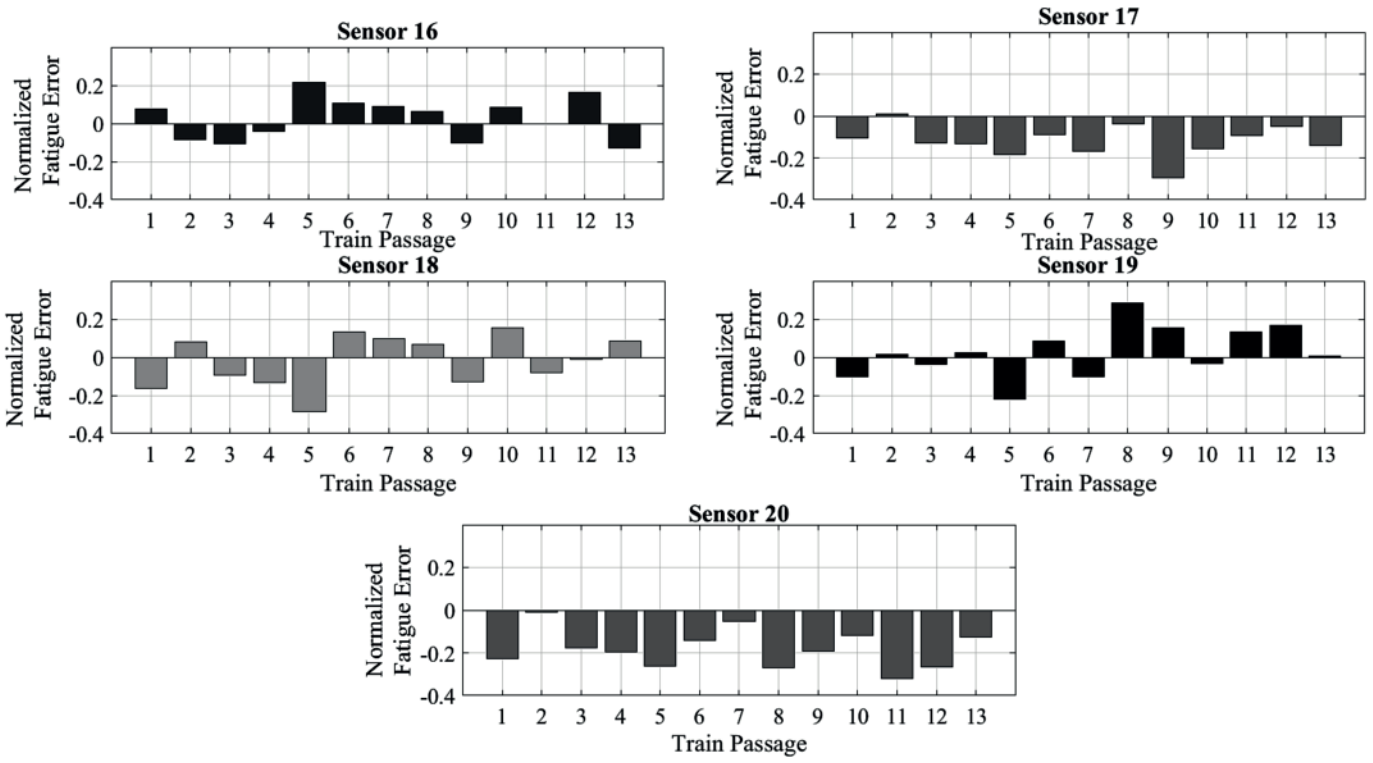

Fig. 27. Classification using POMS—normalized fatigue error for Class 2 ("moderate" train passage, Fig. 3).

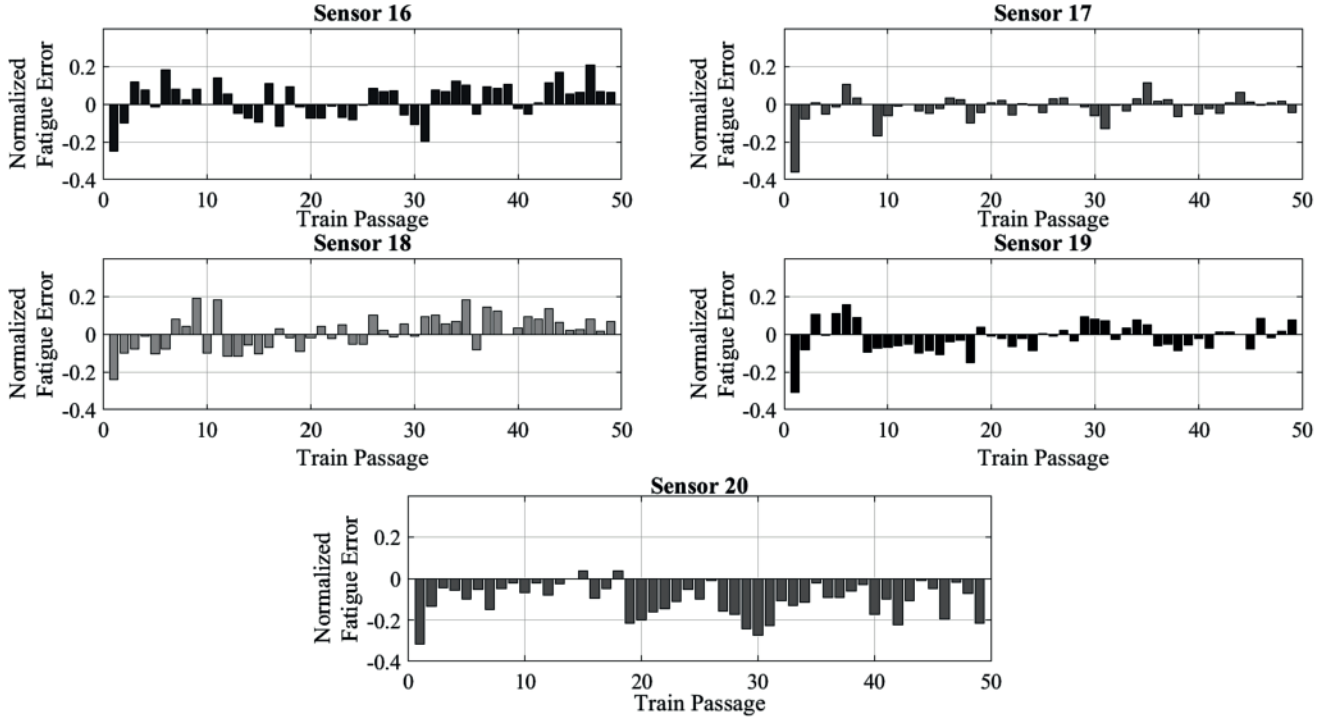

Fig. 28. Classification using POMS—normalized fatigue error for Class 4 ("extremely heavy" train passage, Fig. 3).

train passage in Class 3. Estimated signals appeared to match measured signals at each sensor location reaonsably well, but amplitude differences were observed. Fig. 24 shows that normalized Class 3 fatigue error ranged between 0.4 and −0.4, with its error histogram shown in Fig. 25. Normalized fatigue errors for other classes were similar as shown in Figs. 26 to 28. Fig. 29 details TRAC correlation for Class 3, with values ranging between 0.87 and unity, indicating good correlation. R^2 values between 0.82 and unity were observed for all unmeasured locations for Class 3 train passages as shown in Fig. 30.

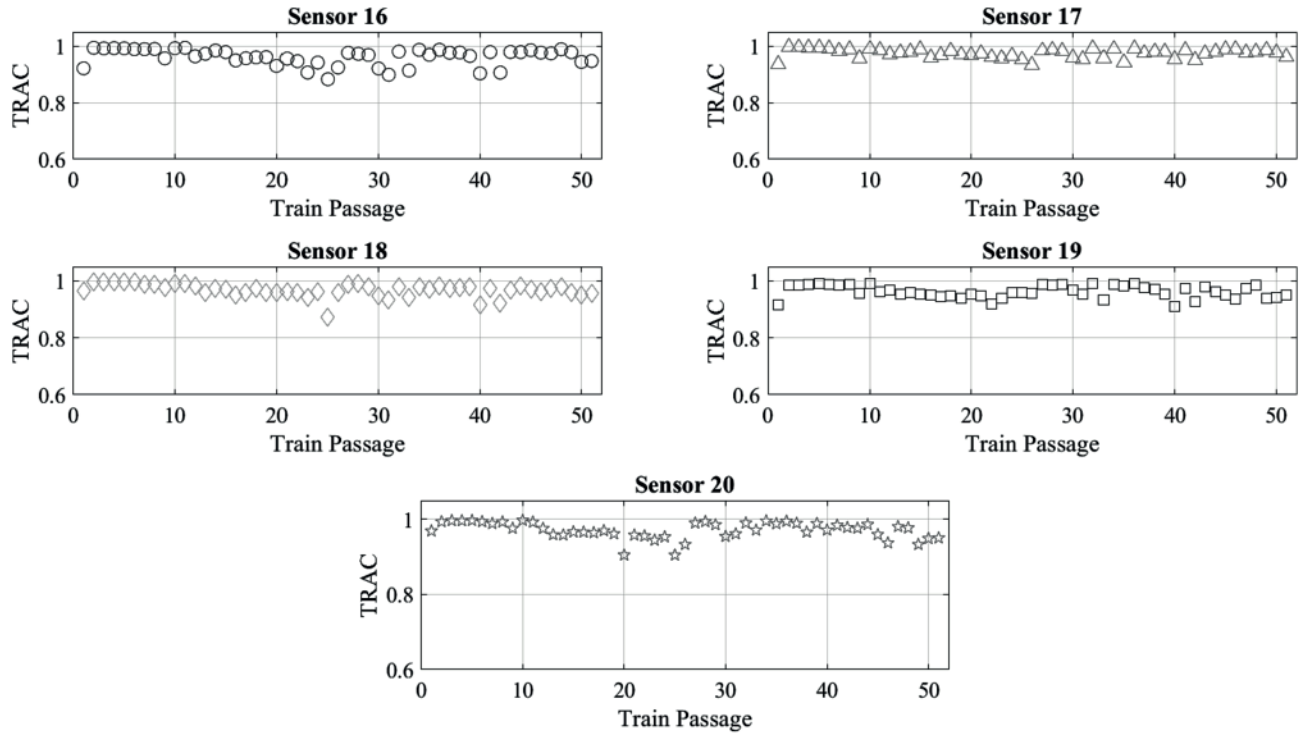

Fig. 29. Classification using POMS—TRAC values for Class 3 ("heavy" train passage, Fig. 3).

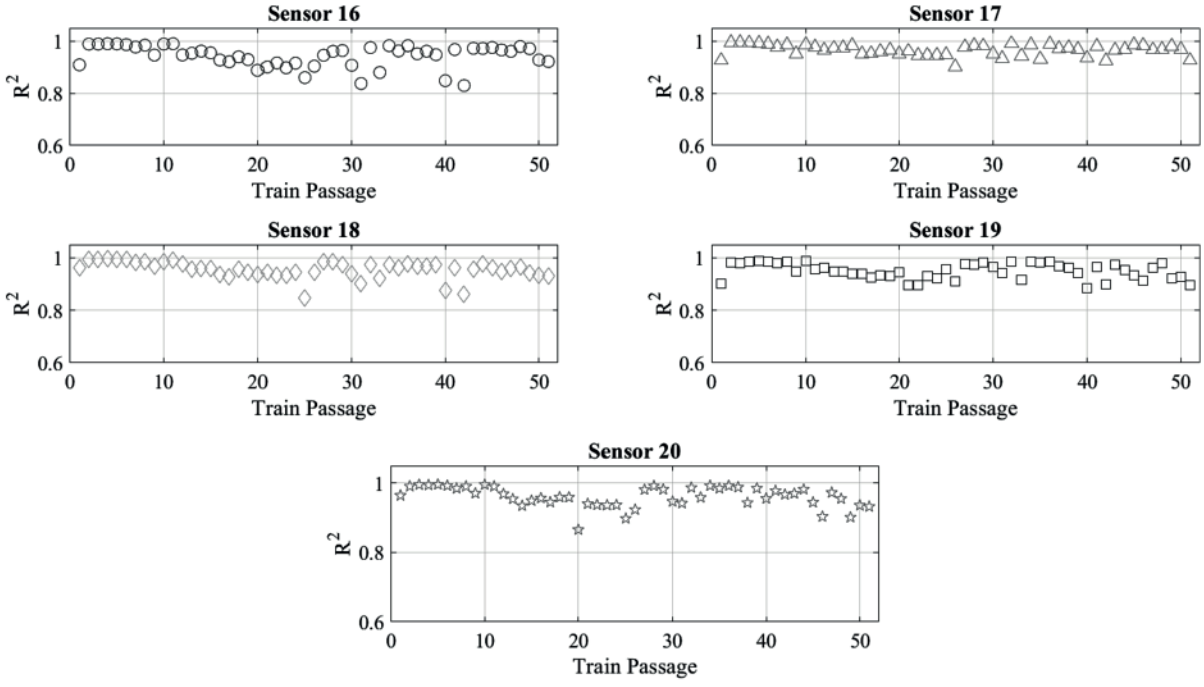

Fig. 30. Classification using POMS—coefficient of determination for Class 3 ("heavy" train passage, Fig. 3).

5. Conclusion

This study attempted to improve strain estimation using novel, data-driven SVD by reducing POM variability. Data was supplied from measured response of an in-service, riveted, steel, truss, two track, railroad bridge whose response to train passages of varying loads, speeds, and directions was monitored. Left singular vectors (i.e., POMs) determined from SVD of measured response data were stored in a snapshot matrix. This information was utilized to examine the effectiveness with which strain signals were predicted at "unmeasured" locations, which were actually measured. Comparisons were completed to determine the efficacy and robustness of using a data driven SVD using a larger data set of trains of varying load, speed, and direction for strain estimation than earlier studies [22]. POM dispersion of the larger data set was also investigated to ascertain its effect on the accuracy of estimated strain response. It was observed that the higher the POM dispersion, the less accurate the estimated strains.

To further reduce POM variability caused by varying train speeds,

loads, and directions (i.e., transverse location), two clustering methodologies were selected and examined to ascertain their ability to effectively categorize strain time-histories into classes moving in similar directions and having similar loads. The first methodology involved classification by k-means clustering using RMS of strain time-histories. The second methodology involved neural net clustering classification by SOM using strain time-history POMs. Each developed class strain responses from trains traveling in similar directions and carrying similar loads. Strain estimation was performed for each class/cluster and three quality measurement types were utilized to assess performance and accuracy of each methodology.

Results showed that a data driven SVD method could predict suitable strain time signals at unmeasured locations for most train passages in all classes, with a normalized fatigue error between 0.4 and −0.4 (40 % and 40 %), an improvement over results reported in the literature [1]. Examined clustering methodologies appeared to perform equally well. It was also observed, however, that the accuracy of strain estimation using SVD largely depended on data set quality. The placement and configuration of the sensors, their sensitivity, and their resistance to environmental conditions can all have a significant impact on the quality of the measured response data set. For instance, sensors that are improperly positioned on a bridge or those that are susceptible to high temperature ranges may impair response measurements, which could have an impact on the accuracy of strain estimation. Investigations focused on further improving the accuracy of virtual sensing by including environmental variability or choosing an optimal sensor placement strategy or exploring adoption of other machine learning techniques for time series analysis are recommended.

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.

Acknowledgments

The authors would like to acknowledge support provided by NSF Award #1762034 BD Spokes: MEDIUM: MIDWEST: Smart Big Data Pipeline for Aging Rural Bridge Transportation Infrastructure (SMARTI). The authors also gratefully acknowledge the assistance, access, computing resources, data, and expertise provided by the University of Nebraska's Holland Computing Center, Union Pacific Railroad Company, and Bridge Diagnostics Inc. in association with this project.

References

[1] Tarpø M, et al. Expansion of experimental mode shape from operational modal analysis and virtual sensing for fatigue analysis using the modal expansion method. Int. J. Fatigue 2020;130:105280.
[2] Kammer DC. Test-analysis model development using an exact modal reduction. Int. J. Anal. Exp. Modal Anal. 1987;2(4):174–9.
[3] Papadimitriou C, et al. Fatigue predictions in entire body of metallic structures from a limited number of vibration sensors using Kalman filtering. Struct. Control Health Monit. 2011;18(5):554–73.
[4] Smyth A, Wu M. Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring. Mech. Syst. Sig. Process. 2007;21(2):706–23.
[5] **H. Jo and B. F. Spencer Jr, "Multi-metric model-based structural health monitoring," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2014*, 2014, .**
[6] Palanisamy R, et al. Experimental validation of Kalman filter-based strain estimation in structures subjected to non-zero mean input. Smart Struct. Syst. 2015;15(2):489–503.
[7] Maes K, et al. Dynamic strain estimation for fatigue assessment of an offshore monopile wind turbine using filtering and modal expansion algorithms. Mech. Syst. Sig. Process. 2016;76:592–611.
[8] **E. N. Chatzi and A. W. Smyth, "The unscented Kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing," *Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures*, vol. 16, *(1)*, pp. 99-123, 2009.**
[9] Mariani S, Corigliano A. Impact induced composite delamination: state and parameter identification via joint and dual extended Kalman filters. Comput. Methods Appl. Mech. Eng. 2005;194(50–52):5242–72.
[10] Lourens E, et al. An augmented Kalman filter for force identification in structural dynamics. Mech. Syst. Sig. Process. 2012;27:446–60.
[11] Ching J, Beck JL, Porter KA. Bayesian state and parameter estimation of uncertain dynamical systems. Prob Eng Mech 2006;21(1):81–96.
[12] **J. Graugaard-Jensen *et al*, "Modal based fatigue monitoring of steel structures," in *Structural Dynamics EURODYN 2005: Proceedings of 6th International Conference on Structural Dynamics, Paris, France, 4-7 September 2005*, 2005, .**
[13] Avitabile P, Pingle P. Prediction of full field dynamic strain from limited sets of measured data. Shock Vibrat 2012;19(5):765–85.
[14] Aenlle ML, et al. Stress estimation in a scale model of a symmetric two story building. in *Proceedings of 5th International Operational Modal Analysis Conference*. 2013.
[15] Skafte A, et al. Experimental study of strain prediction on wave induced structures using modal decomposition and quasi static ritz vectors. Eng. Struct. 2017;136:261–76.
[16] Nabuco B, et al. Nonlinear strain estimation based on linear parameters. in *International Conference on Offshore Mechanics and Arctic Engineering*. 2018.
[17] Iliopoulos A, et al. Fatigue assessment of offshore wind turbines on monopile foundations using multi-band modal expansion. Wind Energy 2017;20(8):1463–79.
[18] **H. P. Hjelm *et al*, "Determination of stress histories in structures by natural input modal analysis," in *Proceedings of IMAC-XXIII, A Conference & Exposition on Structural Dynamics. Society for Experimental Mechanics*, 2005, .**
[19] Nabuco B, et al. Reliability analysis of offshore structures using OMA based fatigue stresses. in *International Conference on Offshore Mechanics and Arctic Engineering*. 2017.
[20] Kullaa J. Virtual sensing of structural vibrations using dynamic substructuring. Mech. Syst. Sig. Process. 2016;79:203–24.
[21] Lourens E, Fallais D. Full-field response monitoring in structural systems driven by a set of identified equivalent forces. Mech. Syst. Sig. Process. 2019;114:106–19.
[22] **S. Eftekhar Azam *et al*, "Experimental validation and numerical investigation of virtual strain sensing methods for steel railway bridges," *J. Sound Vibrat.*, vol. 537, pp. 117207, 2022. Available:** https://www.sciencedirect.com/science/article/pii/S0022460X2200400X. **DOI: https://doi.org/10.1016/j.jsv.2022.117207.**
[23] Kanungo T, et al. An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. 2002;24(7):881–92.
[24] Kohonen T. The self-organizing map. Proc. IEEE 1990;78(9):1464–80.
[25] Klema V, Laub A. The singular value decomposition: Its computation and some applications. IEEE Trans. Autom. Control 1980;25(2):164–76.
[26] Autonne L. Sur les groupes linéaires, réels et orthogonaux. Bull. Soc. Math. France 1902;30:121–34.
[27] Eckart C, Young G. A principal axis transformation for non-Hermitian matrices. Bull. Am. Math. Soc. 1939;45(2):118–21.
[28] Sylvester JJ. A new proof that a general quadric may be reduced to its canonical form (that is, a linear function of squares) by means of a real orthogonal substitution. Messenger Math. 1889;19:1–5.
[29] Schmidt E. Zur theorie der linearen und nichtlinearen integralgleichungen. in Integralgleichungen Und Gleichungen Mit Unendlich Vielen UnbekanntenAnonymous; 1989.
[30] Weyl H. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). Math. Ann. 1912;71(4):441–79.
[31] Stewart GW. On the early history of the singular value decomposition. SIAM Rev. 1993;35(4):551–66.
[32] Golub GH, Reinsch C. Singular value decomposition and least squares solutions. *Linear Algebra*Anonymous 1971.
[33] Chopra AK. Dynamics of Structures 2007.
[34] Rageh A. "Optimized health monitoring plans for a steel. Double-Track Railway Bridge," 2018.
[35] Eftekhar Azam S, Rageh A, Linzell D. Damage detection in structural systems utilizing artificial neural networks and proper orthogonal decomposition. Struct. Control Health Monit. 2019;vol. 26, (2):e2288.
[36] Sajja PS, Akerkar R. Bio-inspired models for semantic web. in *Swarm Intelligence and Bio-Inspired Computation*Anonymous. 2013.
[37] Kohonen T. The self-organizing map. Neurocomputing 1998;21(1–3):1–6.
[38] Kohonen T, et al. Engineering applications of the self-organizing map. Proc. IEEE 1996;84(10):1358–84.
[39] **K. Du and M. N. Swamy, *Neural Networks in a Softcomputing Framework*. 2006.**
[40] Vesanto J, et al. Self-organizing map in matlab: the SOM toolbox. in *Proceedings of the Matlab DSP Conference*. 1999.
[41] **A. Almarnaess, "Fatigue handbook: offshore steel structures," 1985.**
[42] **M.M. Pedersen "Introduction to metal fatigue," And no.: Technical Report ME-TR-11 2018 91.**
[43] Chen N, Wang G, Soares CG. Palmgren–miner's rule and fracture mechanics-based inspection planning. Eng. Fract. Mech. 2011;78(18):3166–82.
[44] **Obando SE, et al. Prediction of forced response on ancillary subsystem components attached to reduced linear systems. Dyn. Coupled Struct. 2014;1:51–72.** https://doi.org/10.1007/978-3-319-04501-6_5.
[45] **P. C. Hansen, V. Pereyra and G. Scherer, *Least Squares Data Fitting with Applications*. 2013.**
[46] Tarpø M, et al. Operational modal analysis based prediction of actual stress in an offshore structural model. Procedia Eng. 2017;199:2262–7.