Feature detection in guided wave ultrasound measurements using simulated spectrograms and generative machine learning.

Isaac I. Setshedi[a,*], Daniel N. Wilke[a], Philip W. Loveday[b]

[a]*Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, 0002, South Africa*

[b]*School of Mechanical, Industrial and Aeronautical Engineering, University of the Witwatersrand, Johannesburg, 2000, South Africa*

Abstract

Guided wave ultrasound field measurements capture reflections from aluminothermic welds in the rail track at unknown distances from a transducer. The measured guided wave signals are complex and challenging to interpret due to multiple dispersive modes propagating in the rail, changing environmental conditions, and noise. Data-driven machine learning techniques have been applied to complex signal-processing problems and have shown significant potential in learning and resolving complex problems. This study aims to understand the implications of using various simulated data sets for application within an appropriate learning framework to capture the underlying features of the field measurements and maximise the performance of these techniques. The use of simulated spectrograms, including variations of signal attributes (attenuation, positions of welds, noise, and mode reflection coefficients) generally observed in experimental measurements, allows the reflections of individual modes to be isolated or combined for training. This allows us to present the training data in three distinct forms. The first dataset has the highest mode reflection information density per sample and consists of simulated spectrogram data with multiple reflections of modes from multiple welds, like experimentally obtained spectrograms. The second training dataset consists of spectrograms with multiple mode reflections; however, only for a single weld reflection per spectrogram. The third training set contains a reflection of a single mode from a single weld in each spectrogram. The data-driven models applied are principal component autoregression and variational auto-encoders. The reconstruction error and latent space interpretability were considered as metrics for the algorithms' ability to learn using test sets, i.e., unseen data. The results show that datasets with sufficient feature variation and higher mode reflection information density better construct the test set of simulated and experimental spectrograms. However, training using the third dataset shows more interpretable latent variables for an artificial growing defect attached to the rail. Furthermore, data-driven machine learning methods trained using simulated spectrogram data are useful for reconstructing and learning features from experimental measurements, provided that the training data have representative mode feature variation and noise.

Keywords: Rail, Ultrasonic Guided Waves, Spectrograms, Simulation, Principal Component Analysis, Variational Auto-Encoder

1. Introduction

Railway systems are critical transportation infrastructure components, demanding robust maintenance and safety measures. The early detection of rail breaks and defects is essential, and guided-wave ultrasound (GWU)-based monitoring systems have emerged as valuable tools in this context [1]. These systems employ guided waves that predominantly concentrate energy in the rail head, enabling long-distance inspection and the ability to identify rail breaks and defects, including cracks. To enhance the capabilities of GWU systems, this research explores incorporating a pulse-echo mode of operation for detecting, locating, and monitoring defects before catastrophic rail breakage.

This paper employs generative models, which are designed to model the underlying distribution of input data and generate new samples that resemble the training data, to learn underlying properties of spectrograms. Generative machine learning models typically involve an encoder-decoder process, as illustrated in Figure 1. The encoder extracts features from input data, known as latent variables, while the decoder reconstructs or approximates these features in the output space.

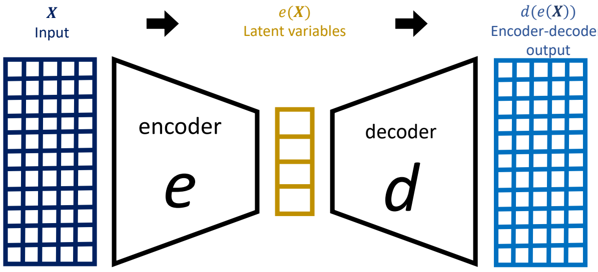


Figure 1: Illustration of the encoder-decoder process for a generative model.

The goal is to minimise information loss in this process, preserving GWU coherent features while eliminating signal noise. The goal of the generative networks is to conserve maximum feature information in the encoding process and minimise reconstruction error of the inputs in decoding latent variables. Furthermore, the generative encoder-decoder model can be written as:

$$(e^*, d^*) = \underset{(e,d)\in E\times D}{\arg\min}\, \epsilon(\mathbf{X}, d(e(\mathbf{X}))), \tag{1}$$

where e and d are encoder and decoder parameters of encoder E and decoder D subspaces, respectively, and $\epsilon(\mathbf{X}, d(e(\mathbf{X})))$ is an error function of inputs $\mathbf{X}$ and reconstructions $d(e(\mathbf{X}))$.

---

[*] Corresponding author.

Email addresses: `u10319965@tuks.co.za` (Isaac I. Setshedi), `nico.wilke@up.ac.za` (Daniel N. Wilke), `Philip.Loveday@wits.ac.za` (Philip W. Loveday)

Generative models can be categorised into variance and interpretation-driven approaches. Variance-driven models use the patterns observed in the variance of the data and capture data fluctuations and variability. These models are more likely to be sensitive to variations in the input features. On the other hand, the interpretation-driven approaches aim to find sources of variability and capture the latent structures responsible for generating the data. These models seek to disentangle the distinct sources of variation.

Current research is limited in terms of the exploration of the training methodologies involving simulated data and learning properties of experimental measurements, and there is a gap in the literature regarding the impact of selecting specific types of simulated data for training purposes. Therefore, field-measured spectrograms are used as a baseline for training and learning underlying features in GWU measurements. This is the most common use of Principal Component Analysis (PCA) for application to ultrasonic measurements [2, 3, 4, 5]. PCA transforms high-dimensional data into uncorrelated principal data components, significantly reducing dimensionality. However, PCA is a linear model, leading to entangled non-linear features, making it challenging to model and interpret such features systematically.

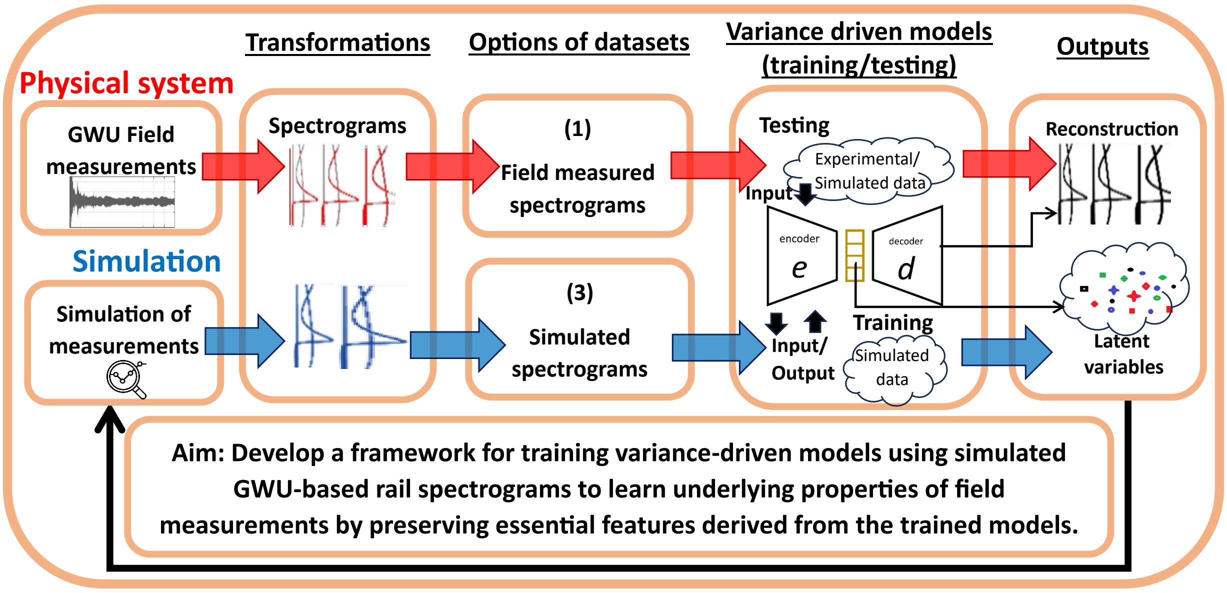
The study also investigates a potential solution in the form of Variational Auto-Encoder (VAE) networks [6], designed to capture non-linear relationships within data. VAE networks are generally used as a dimension-reduction technique to compute the latent space of data and have been used for guided wave defect detection [7]. The VAE networks typically consist of an encoder network that parametrises a posterior distribution of latent random variables for input data. This results in a prior distribution in the latent space decoded to an output distribution. The work covered in this paper focuses on both the prior representations and reconstruction.

Machine learning techniques have played an important role in the use of GWU-based data for health monitoring. Notably, discriminative learning approaches have been employed to classify input features into predefined classes. Rautela et al. [8] have used deep discriminative learning models to detect cracks in 1D waveguides, investigating a range of deep learning networks for damage detection. Similarly, Lomazzi et al. [9] and Zhang et al. [10] have applied discriminative machine learning methods to diagnose damage using guided wave ultrasound. Loveday et al. [11] proposed processing steps for simulating the growth detecting of rail defects using phased array processing, dispersion compensation, signal stretching, scaling, and reordering. However, these signal processing methods rely on an accurate model of the rail, and the application of machine learning in this domain is critical for advancing non-destructive testing.

Unsupervised machine learning methods have also emerged in this domain, as demonstrated by Rautela et al. [8], Silva et al. [12], and Bourdais et al. [13]. Silva et al. [12], in particular, using deep PCA to extract underlying feature distributions, enhancing the interpretability of GWU data. Such unsupervised approaches have improved health monitoring systems through experimental data from healthy systems. Furthermore, Ramatlo et al. [14] present a novel approach for generating synthetic ultrasonic inspection data for welded railway tracks, focusing on using of a VAE to simulate damage signatures in inspection signals.

The novelty of this study is the development of a framework for training generative variance-driven models using simulated GWU-based rail spectrograms to learn underlying properties

of experimental data, addressing a current gap in the domain of machine learning and non-destructive testing and evaluation. The approach involves preserving essential features derived from variance-driven models trained on simulated spectrograms to encode underlying properties of field-measured spectrograms.
Figure 2 also provides an illustrative methodology of the study, from the physical GWU experiments to the transformation of raw data into experimental spectrograms. Training is performed on an untrained variance model using simulated data to enhancing its capacity to capture data variability and patterns. The trained model is then evaluated using experimental and previously unseen simulated data to enable spectrogram reconstruction and the extraction of latent parameters encapsulating underlying signal characteristics. This study places dual emphasis on reconstruction and understanding the latent space, offering the potential for advancements in railway track maintenance, safety enhancement, and data-informed decision-making.



Aim: Develop a framework for training variance-driven models using simulated GWU-based rail spectrograms to learn underlying properties of field measurements by preserving essential features derived from the trained models.

Figure 2: Schematic representation of the methodology for this study.

The primary goal is to learn and reconstruct field spectrograms using only simulated data, as it allows for selecting specific features for training. The simulated datasets are based on characteristics from field measurements, and we evaluate the most appropriate type of simulated training using three datasets with varying information densities.
The paper is structured as follows: Section 2 discusses the PCA model and VAE using Convolutional Neural Networks (CNNs), designed for processing spectrograms in this study. Section 3 provides details about the field measurements performed on a UIC-60 rail. The measurements include a transducer attached to the rail head and a defect attached to the rail head. Temperature variations were also measured near the rail. Section 3 also discusses

the implementation of the Semi-Analytical Finite Element (SAFE) method to generate time-frequency dispersion curves representing different propagation modes along the rail. Two types of features in the GWU rail measurements are identified: coherent and incoherent, both feature types are considered in training to enhance the models' robustness. Three simulated training datasets (A, B, and C) are generated with varying levels of feature density (number of dispersion curves per sample). The datasets aim to investigate the influence of the density of coherent features on machine learning model training. Section 4 utilises PCA and VAE models to analyse the simulated datasets, and investigate optimal settings for training using the models, including the percentage of spectrograms with incoherent features and the number of latent variables. In section 5, the PCA and VAE models are trained separately with each dataset (A, B, and C) and tested on experimental data, to accurately reconstruct spectrograms and capture underlying properties such as temperature variations and varying defect amplitude.

2. Model architectures

Generative machine models, encompassing PCA and VAE, can play pivotal role in unravelling complex patterns inherent in structural data. The essence of auto-association in this generative machine models lies in capturing intrinsic structures within data, facilitating a self-representation mechanism. The PCA and VAE model can be powerful tools for dimensionality reduction, revealing dominant modes of behaviour within the structural datasets. The coherent features extracted by the models enable the identification and interpretation of critical information, laying the groundwork for a more nuanced understanding of data.

2.1. PCA

PCA aids in capturing the most significant features in a dataset by projecting it onto a lower-dimensional subspace. Furthermore, PCA can be used to identify critical modes of behaviour within data. PCA involves the computation of eigenvectors and eigenvalues from the covariance matrix of the input data. Given a dataset X of n observations and m features, the covariance matrix Σ is calculated as:

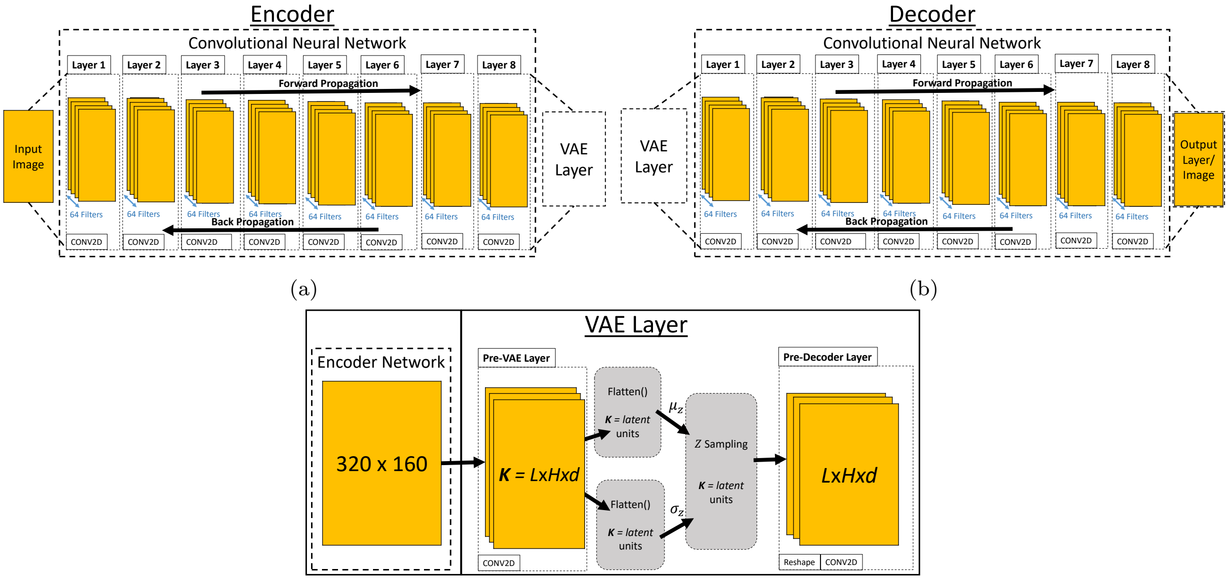
$$\Sigma = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T$$

Here, X_i represents the i-th observation, and $\bar{X}$ is the mean of the dataset. The eigenvectors $\mathbf{v}$ and corresponding eigenvalues λ of Σ satisfy the equation:

$$\Sigma\mathbf{v} = \lambda\mathbf{v}$$

2.2. VAE

The VAE is implemented with CNNs for this study, as shown in Figure 3. CNNs are deep neural network architectures designed to process data having a grid-like topology, like the spectrograms used in this study. This is preferred over fully connected networks as CNNs scale better on large images than fully connected networks. The goal of neural networks is to approximate a map of some input to an output/feature space. In our case, we map grayscale GWU rail spectrograms to a spectrogram with similar output features. The VAE networks feedforward propagate input spectrograms through hidden layers to an approximate output spectrogram. Backpropagation propagates information backwards through the network from the network to the weights of the hidden layer. The obtained gradient information is then used with a stochastic gradient descent-based optimisation algorithm. In this continuous forward and backward propagation process, the learning parameters (weights and biases) in all the layers are tuned to a value that minimises the cost function. In general, this makes neural networks powerful in approximating real-valued continuous functions. Our hidden layers employ two-dimensional (2D) Convolutional Neural Network (CNN) layers. CNNs are known for faster and more accurate image data training than a fully connected neural network (FCN). This is due to the sparse connections and parameter sharing in CNNs.



(a) (b)

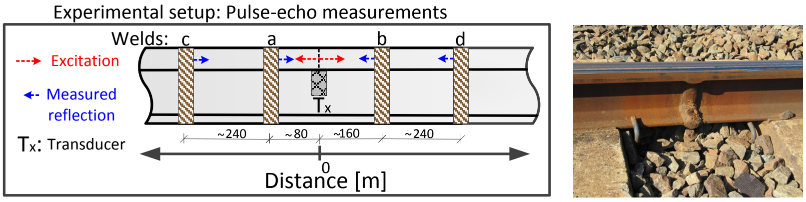
(c)

Figure 3: The VAE network architecture including (a) encoder, (b) decoder and (c) VAE layer

The CNN encoder and decoder architectures, as shown in Figures 3(a) and 3(b), respectively, are inspired by the U-net and VGG-11 networks, renowned for their high accuracy in large image recognition [15]. VGG networks are characterised by their depth and simplicity, employing small 3 x 3 convolutional filters, which make them easy to implement and adapt for various applications. In our adaptation of the VGG-11 network, we made specific modifications. First, we eliminated the Fully Connected Network (FCN) layers and replaced traditional pooling with 1 x 2 convolutional strides. The network begins with the encoder, which takes input images of dimensions (320, 160, 1). The input images are experimental and simulated spectrograms discussed in Sections 4 and 5. The encoder employs a series of convolutional layers with 2-pixel strides to reduce the spatial dimensions of the image while increasing the number of filters. To introduce Variational Auto-Encoder (VAE) capabilities, we replaced the FCN layers in the VGG with VAE layers, as shown in Figure 3(c). These VAE layers take the mean μ_z and variance σ_z values as inputs and sample points from the latent space, following a normal distribution. These sample points represent the encoded representation of the input image. Following the last layer of the encoder network, we introduced simple CNN transpose layers to the decoder network. These layers take the sampled point from the latent space as input and apply transposed convolutional layers. Throughout the network, both the encoder and decoder layers maintain 64 filters per layer. The final layer of the decoder applies a sigmoid activation function to ensure that the pixel values in the output image fall within the range of 0 to 1. The VAE model processes input images using the encoder network and employs the decoder network to generate the output image. The reconstruction loss metric used during training is based on the mean-squared error between the input and output images.

3. Data source for datasets

Our GWU measurements were taken on a continuously welded rail track, rail sections are welded together in the field using aluminothermic welding, and a weld cap remains in place. The weld cap produces a reflection of ultrasonic guided waves, which may be used to characterise the propagation modes. Figure 4(a) illustrates a transducer attached to the head of the rail and the presence of aluminothermic welds at regular distances and represents the experimental setup used in this work.



(a)

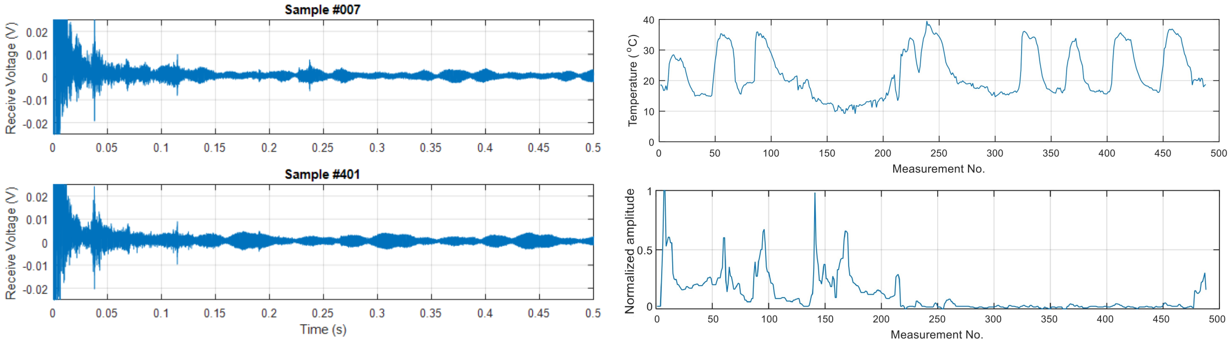
(b)

Figure 4: (a) Illustration of the experimental setup with estimated weld distances from the transducer location. (b) Photograph of an aluminothermic weld.

An artificial defect was included by attaching a mass to the head of the rail when simulating a defect that would lead to time-varying reflections. The expectation was that this glue joint would degrade relatively quickly, causing the mass to fall off, thereby creating a dynamic reflection scenario detectable by the monitoring system. The attached mass was designed to generate a displacement reflection coefficient similar to that of a weld. This choice was based on previous predictions and modelling, which suggested that a small transverse crack in the rail head would reflect similarly to the mass. Loveday et al. [11] provide further details of the monitored rail section and the artificial defect used in this work. Figure 4(b) also indicates that weld reflections can originate on either side of the transducer since the excited waves propagate forward and backward. The distances between the welds can be up to 240m, but these distances are not accurately known. Environmental conditions also significantly impact the wave propagation characteristics of rail [16].

3.1. Underlying features

A pulse-echo transducer was attached under the head of the UIC60 rail to mainly excite modes with energy concentrated in the head of the rail to collect field measurements at a sampling frequency of 100 kHz. The transducer was excited by a 17.5-cycle Hanning-windowed tone burst centred at 35 kHz to produce propagating modes, which interact with welds and an artificial defect at unknown distances. These signals were generated within a data acquisition system and then amplified to reach a peak-to-peak maximum amplitude of 200 V. Measurements, shown in Figure 5, were acquired at 30-minute intervals continuously over 14 days. However, to ensure the accuracy of the data collected, a system delay was implemented to delay measurements when the train was passing.



(a) (b)

Figure 5: (a) Pulse-echo measurements performed on a UIC60 rail with an artificial defect attached to the rail head. (b) Temperature measured in a cabinet placed near the rail and the normalised amplitude of a reflection from an artificial defect attached to the head of the rail.

The temperature of the rail varied over time, and the dependence of the elastic modulus and speed of sound in the rail on temperature is well understood [16]. Variations in temperature cause variation in the elastic modulus of the rail and, therefore, in the speed of sound. Furthermore, the slight variation in the rail temperature results in modes propagating at different speeds and other unknown changes in the received signals. Temperature measurements were taken using a temperature sensor in a cabinet near the rail, as shown in Figure 5(b). It should be noted that these temperatures estimate the overall temperature variation of the rail, and the rail may be unevenly heated along the length. The artificial defect reflection can be seen in the time signal measurement in Figure 5(a) and the strength of the reflection from the artificial defect varied over time as shown in Figure 5(b). Sample numbers 7 and 401 are measured signals for the most significant and minor defect reflection signals, respectively. The reflection from the defect did not decrease monotonically as expected. It is speculated that local resonances involving the mass and the glue joint stiffness cause some large reflections as the glue joint deteriorates.

3.2. Spectral representation

The resulting time-domain measurements are transformed into spectrograms using the STFT, which is used in signal processing to analyse non-stationary signals [17] using the transformation:

$$STFT\{x(t)\}(\tau,\omega) = X(\tau,\omega) = \int_{-\inf}^{\inf} x(t)w(t-\tau)e^{-j\omega t}dt \tag{2}$$

where w, $x(t)$, τ and ω represent the windowing function, time signal, spectral time, and frequency. All our measurements range from $t = 0$ to $t = 0.6$ seconds, $\omega = 32kHz$ to $\omega = 42kHz$ and $w = 512$, which result in 320 by 160-pixel spectrograms. Pixel amplitudes are normalised to a range of 0 and 1 to avoid scaling errors in the machine learning models. The time-domain measurements in Figure 5(a) can be represented in the time-frequency domain as spectrograms, as shown in Figure 6, which contain energy representations of reflections from the many aluminothermic welds on the rail.

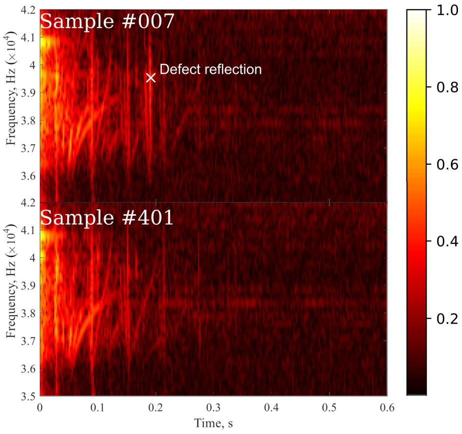


Figure 6: Field measurement samples represented by spectrograms.

Spectrograms or image data allow for comparison with established non-linear ML image-driven approaches. We deliberately choose spectrograms over raw time-series data for two reasons. Firstly, spectrograms provide the raw time-series data after non-linear transformations have been applied to the lower dimensional windowed raw time-series data; therefore, improving the linear methods' setting. Secondly, research on image data using non-linear machine methods is vastly more mature than raw time-series data. It is important to note that the transformation from time-series to time-frequency spectrograms results in losses in information, such as phase information, however, training using spectrograms is computationally more efficient and allows easier information extraction.

3.3. Simulation framework

Our GWU-based measurements are difficult to label and intractable to translate available labels between simulated and experimental data. We, therefore, must consider a generative approach to training. We identify modes propagating along the rail from field measurements to simulate the training samples, as depicted in Figure 7. A transducer attached to the rail head measures the propagating modes as time signals. These time signals can be better visualised in the time-frequency domain using a Short-Time Fourier Transform (STFT) [17], with the results depicted as spectrograms. The energy bands on the spectrograms represent the times of arrival of different modes of propagation reflected from different welds in the rail. We simulate the propagating modes using the Semi-Analytical Finite Element (SAFE) method [18, 19, 20, 21, 22]. The method can efficiently compute dispersion curves by assuming a constant cross-section along the rail, given that material properties are known.

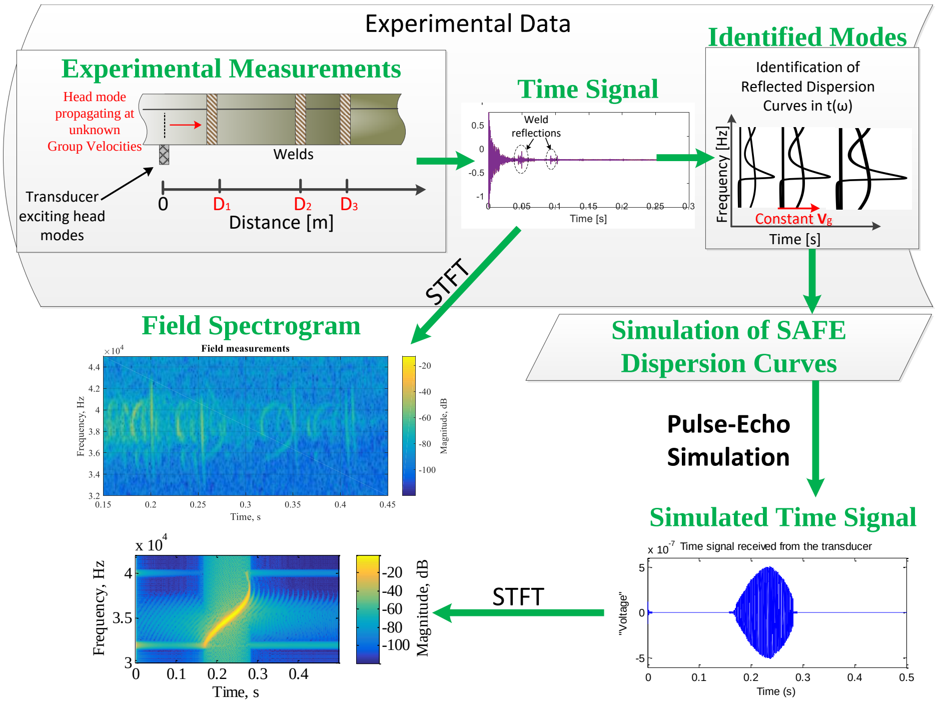


Figure 7: Illustration of the isolation and generation of simulated time-frequency spectrograms of modes that are measured in field pulse-echo experiments.

A simple modelling framework was adopted to model weld reflections, where dispersion curves computed using the SAFE method propagated a Hanning-windowed tone burst to various distances assumed to be aluminothermic welds with selected reflection coefficients. The propagated signals are reflected by the welds where the incident modes are reflected using reflection coefficients (including mode conversion) selected to approximate the reflections observed in experimental data. It should be noted that this simple modelling framework uses only a SAFE model and no 3D finite element model as used in the more sophisticated modelling framework developed by Ramatlo et al. [23, 24]. Using the STFT the computed simulated time signals can be transformed into spectrograms, as shown in Figure 8.

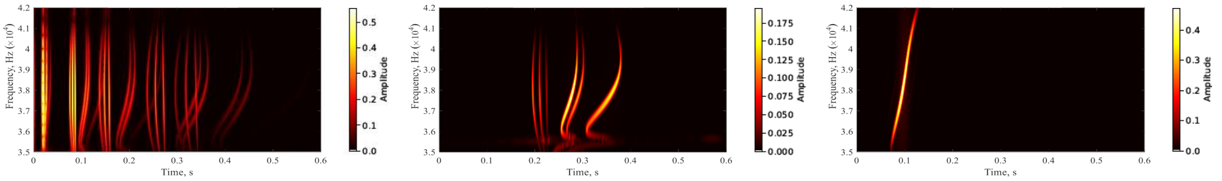
3.4. Ultrasonic guided wave features

Our GWU rail measurements generally consist of two kinds of features, coherent and incoherent. To accurately train machine learning models and simulate guided measurements that are like to the experimentally measured signals, these features of the experimental signals are considered:

- Coherent features include individual dispersive modes that propagate and are reflected by welds back to the pulse-echo transducer. At the welds, propagating modes are both reflected and transmitted. If modes are reflected, they may return to the transducer as the incident mode or as different modes due to scattering [25]. Modes that return to the transducer as different modes are called coupled or converted modes. The railhead transducer measures at least six different modes of propagation. The measured modes are shown in a spectrogram of a field measurement in Figure 6. Superposition of other non-dominant modes of propagation reflected by other reflectors in the rail, which are seen in field measurements, can exhibit the same characteristics as the excitation waveform [26]. We can account for the signals by simulating spectrograms with the multiple modes of the rail. Another essential feature of the measured signals is the decay or attenuation of the measured energy over time or distance, which is evident in the field-measured spectrogram shown in Figure 6.
- Incoherent features are not correlated to the measured signal and do not influence the measured signal. However, this random noise must be included in the training to enhance the robustness of the trained machine-learning models.

3.5. Simulated training data

We simulated three time-series datasets of guided wave measurements with the experimental measurements' wave propagation and dispersion properties. Each sample from dataset A (with a sample shown in Figure 8(a)) has multiple mode reflections (including mode-converted reflections) from five simulated welds in a rail waveguide. The samples from dataset B (with a sample shown in Figure 8(b)) have a single weld reflection with multiple modes of propagation. The samples from dataset C (with a sample shown in Figure 8(c)) also have a single weld simulated and only one mode is reflected per sample. The aim of using these samples is to investigate whether feature density (the number of dispersion curves per sample) in our spectral data influences the training of machine learning models.



(a) (b) (c)

Figure 8: Samples from datasets with (a) multiple weld reflections (dataset A), (b) single welds (dataset B) and (c) isolated propagating modes (dataset C).

The first training dataset uses simulated spectrograms like those obtained experimentally. This dataset has multiple weld reflections and multiple propagation modes in each spectrogram sample. Training with this dataset allows the ML models to present the highest-density information of propagating modes per training sample. The learning and reconstruction of test data depend on how well the trained model breaks down wave propagation characteristics presented in this high information-density dataset. The second and third training datasets have less information density of propagating modes per training sample than the first. For both, only a single weld reflection is considered. The second training dataset considers multiple propagating modes per training sample, while the third training dataset dataset uses only one propagating mode per training sample. To ensure that the properties in our experimental data are accurately represented in SAFE simulated spectrograms the following properties are considered and varied:

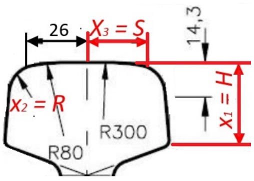
- Reflection distances are unknown and reflected signals from different reflectors may overlap. As shown in Figure 6, energy reflections of very dispersive modes from one weld may be received simultaneously with the energy of other modes from other weld reflections. This interaction between reflections is considered only when generating dataset A.
- The amplitude of the reflected signal is dependent on the distance and the propagation mode. Due to attenuation in the rail, the reflected energy received from welds reduces with the distance to weld. Furthermore, very dispersive modes tend to have low propagation energy relative to less dispersive ones. To account for attenuation in simulation for this study, the received signal is multiplied by a signal intensity coefficient, which was characterised by an exponential decay with distance, described by the formula:

  $$I = I_0 \cdot e^{-2\alpha x} \tag{3}$$

  where I represents the intensity coefficient of each received reflected signal at a given distance x, $I_0 = 1$ is the initial intensity, α is the attenuation coefficient, and x is the distance to the reflector. This attenuation coefficients were selected to provide similar attenuation to that observed in experimental measurements.
- Reflection coefficients for all modes and reflectors are unknown. Furthermore, topologies of welds are not considered for our study. Therefore, welds are treated as reflectors with unknown but varied properties.
- The method by which the rail is worn is unknown. Contact between the rail and train wheels constantly changes the head profile of the rail, and the profile is also periodically ground to prevent shelling [27, 28, 29, 30]. Therefore, to account for the possible wear profiles, the SAFE models were generated using various rail profiles. Setshedi et al. [31] found that at least three dimensions are required to describe these wear characteristics. These three dimensions of the UIC60 rail are shown in red in Figure 9(a). Geometric parameters X_1, X_2 and X_3 (described in mm) enable a reduction in the head height, increased interior radius and a shaper exterior radius respectively.

- The elastic modulus and Poisson's ratio of the rail are not known precisely. Furthermore, these properties can be affected by environmental changes. The elastic modulus is affected by temperature changes and the temperature of the rail may vary over its length. The SAFE model was therefore computed using a wide range of steel elastic properties to account for the various elastic properties caused by environmental changes.

To generate SAFE dispersion curves for various properties, Latin Hypercube Sampling (LHS) [32] was employed to uniformly sample the space of possible geometrical and material parameters. 5,000 SAFE dispersion curve samples were generated using the `LHSdesign` function in MATLAB for geometry and elastic variables: H, R, S, ν (Poisson's ratio) and c (speed of sounds in the waveguide). The computed dispersion curves can propagate tone bursts to various distances to generate simulated signals. The upper and lower bounds of the simulation variables are shown in Figure 9(b).



(a)

Variables	**H**	**R**	**S**	ν	c **[m/s]**	α	x **[m]**
Lower bound	27.5	13	26	0.25	4800	0	10
Upper bounds	37.5	21	36	0.35	5200	0.5	1000

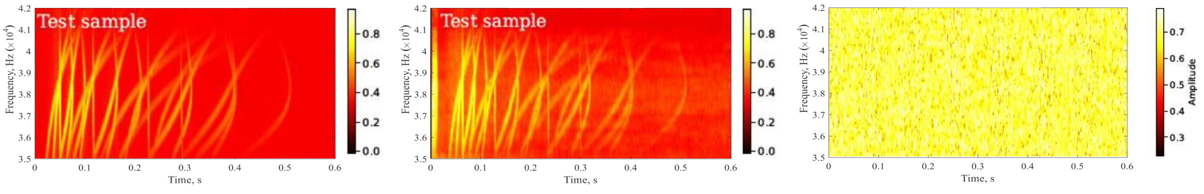
(b)

Figure 9: (a) Standard UIC60 rail head profile highlighting parameters used to describe wear and (b) ranges for the variables used to generate simulated data.

Together with simulated datasets, spectrograms with incoherent noise are also included in the training sets to ensure that our trained models are robust and do not break when applied to experimental measurements. The random noise samples are generated by adding Gaussian random noise to time-domain signals converted to spectrograms using the STFT. A sample spectrogram with incoherent noise is shown in Figure 10(c). These spectrograms will be referred to as noise spectrograms.

4. Model training approach and hyperparameter selection

A dataset with simulated spectrograms containing multiple reflectors, with sample spectrograms shown in Figure 10, was generated to study the structure of reconstruction and latent space. The simulated spectrograms were generated such that the reflected energies of modes are like those presented in experimental spectrograms, with the least dispersive mode having the highest energy. They also contain dispersion curves with intersecting modes from different reflectors and distances to welds assumed to be unknown. Gaussian random noise was added to the time-series signals and decay in the energy measured for the measured reflections by the transducer was also simulated using Equation 3 in the dataset to ensure that the spectrograms are similar to field measurements. Therefore, in each sample, four variations that are seen in the field measured spectrograms were simulated by randomly varying positions of weld reflections (that are known to change due to temperature variations in the rail [31]), reflection coefficients of modes, attenuation of received signals and adding Gaussian noise to each signal.



(a) (b) (c)

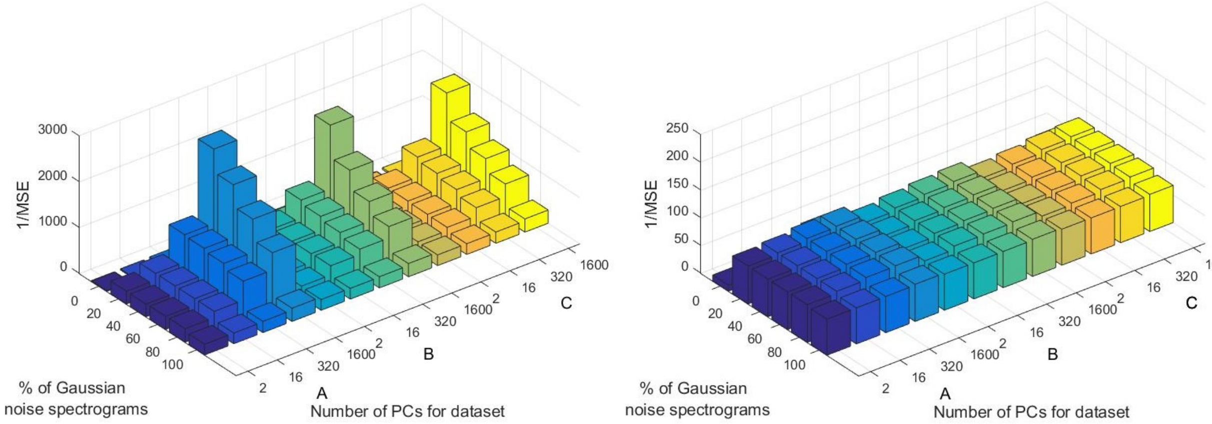
Figure 10: Samples of the five-reflections spectrograms generated from SAFE models: (a) without Gaussian noise added and (b) with Gaussian noise added to each spectrogram, and (c) a sample of a Gaussian noise spectrogram included in the training batch.

4.1. PCA

The settings for training the models were optimised by experimenting with different percentages of noise spectrogram samples and varying numbers of latent variables. Additionally, PCA models were trained on the dataset to extract useful coherent features from experimental measurements. The Mean Squared Error,

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{X}_i)^2, \tag{4}$$

where X is a spectrogram of observed values of the variable being predicted, with $\hat{X}$ being the predicted values for each model, were used to estimate the best setting for training with noise spectrogram samples in the training batch.

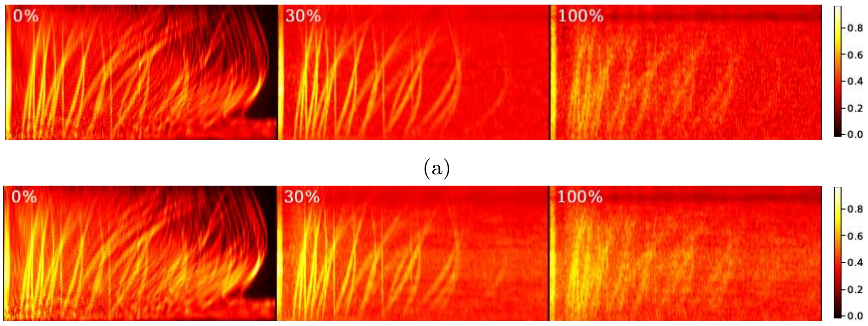


(a) (b)

Figure 11: Reciprocal of the mean squared error using the PCA models to reconstruct the error (sensitivity) of (a) spectrograms with coherent features only and (b) spectrograms with coherent features and added noise to each sample, versus the number of latent variables and the percentage of Gaussian noise spectrograms in a training batch.

Values of 1/ MSEs are computed for the PCA models on a test dataset of spectrograms with five reflections, as shown by the bar plots in Figure 11. The spectrograms with coherent features only (without noise added to each spectrogram), with a sample shown in Figure 10(a), were used for generating plots in Figure 11(a). Furthermore, spectrograms with coherent features and added noise to each sample were used to generate plots in Figure 11(b), with a sample shown in Figure 10(b). Both data sets contain coherent feature spectrograms, ensuring that the comparison is against the underlying signals present in each, aiming for a target MSE of 0. This allows for assessing the models' performance with and without the influence of added noise to each spectrogram.

The PCA model sensitivity bar plots indicate that the model is sensitive to inclusion of Gaussian noise spectrograms, with the best setting for training with Gaussian noise spectrogram at 20-40%. Dataset A shows greater sensitivity to the proportion of the Gaussian noise spectrogram and trends on the sensitivity plots indicate that the PCA model is more accurate when more latent variables are applied. This is expected as more information is passed through the latent variables. The plots also show that the model with 1600 latent units has the best reconstruction when Dataset A and 20% of noise spectrograms are used for training, with further analysis of the data showing optimal results when approximately 30% Gaussian noise spectrograms are used. This is due to the greater density of information presented on Dataset A spectrograms compared to datasets B and C. Reconstruction plots shown in Figure 12 also show that the PCA model fits on the incoherent noise on the spectrogram with added noise (shown in Figure 10(b)). Furthermore, the reconstruction plots show that the model fits dispersion curves to noise when the percentage of Gaussian noise spectrograms is 0 in the training batch.



(a)

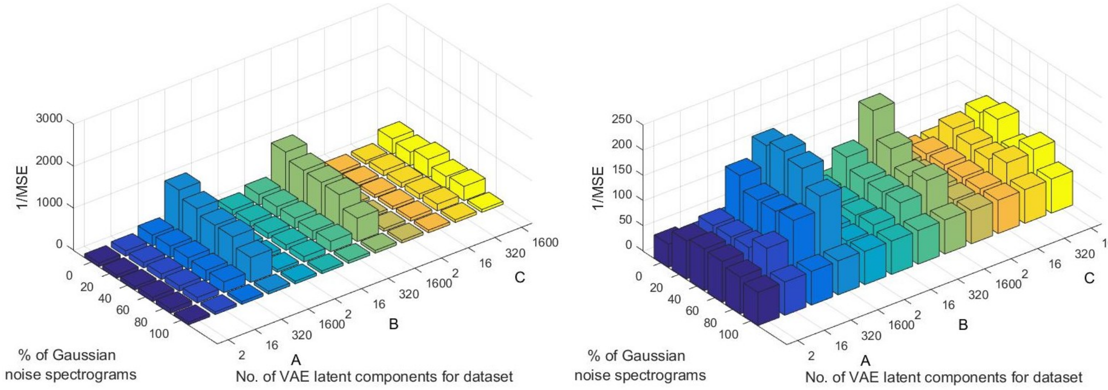
(b)

Figure 12: Reconstructions of the samples in Figures 10(a) and 10(b) using a PCA model trained with dataset A and 1600 latent components, under varying percentages of noise Gaussian spectrograms in the training batch: (a) shows the reconstruction of the sample spectrogram in Figure 10(a), and (b) shows the reconstruction of the sample spectrogram in Figure 10(b). The horizontal axes represent time in seconds, while the vertical axes denote the frequency of the reconstructed spectrograms.

4.2. VAE

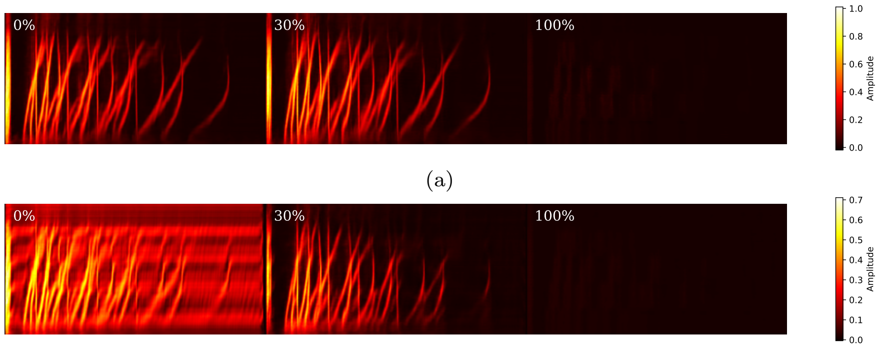
A VAE model was trained using the three datasets investigated in this paper, similar to the PCA model. Both sensitivity bar plots of PCA shown in Figure 11 and the VAE shown in

Figure 13 indicate that the models are sensitive to including Gaussian noise spectrograms, with the best setting for training with Gaussian noise spectrogram at 20-40%. The sensitivity of the models to the proportion of Gaussian noise spectrograms varied between datasets, with dataset A showing greater sensitivity to the inclusion of Gaussian noise spectrograms in both the PCA and VAE models. The VAE model with 1600 latent variables was found to have the best reconstruction when trained on dataset A and 30% Gaussian noise spectrograms. This is due to the non-linear nature of the VAE model and the greater density of information presented on dataset A spectrograms compared to datasets B and C. Overall, the results suggest that for all datasets, the best setting for training is 1600 latent variables and 30% Gaussian noise spectrograms.



(a) (b)

Figure 13: Reciprocal of the mean squared error using the VAE models to reconstruct the error (sensitivity) of (a) spectrograms with coherent features only and (b) spectrograms with coherent features and added noise to each sample, versus the number of latent variables and the percentage of Gaussian noise spectrograms in a training batch.



(a)

(b)

Figure 14: Reconstructions of the samples in Figures 10(a) and 10(b) using a VAE model trained with dataset A and 1600 latent components, under varying percentages of Gaussian noise spectrograms in the training batch: (a) shows the reconstruction of the sample spectrogram in Figure 10(a), and (b) shows the reconstruction of the sample spectrogram in Figure 10(b). The horizontal axes represent time in seconds, while the vertical axes denote the frequency of the reconstructed spectrograms.

4.3. Reconstuctions and latent variables

To assess the performance and generalization capability of our models, tests are performed on simulated data in this section and experimental data in the next section, focusing on the detection of propagation modes. The models under evaluation were trained with a composition of 70% coherent feature spectrograms and 30% Gaussian noise spectrograms, utilizing 1600 latent variables for optimal performance.

The reconstruction results obtained from the PCA models, as shown in Figure 15, reveal consistent performance across three distinct training datasets (A, B, and C). Furthermore, PCA tends to reconstruct noise prevalent in the test spectrograms, especially when employing a higher number of latent variables.

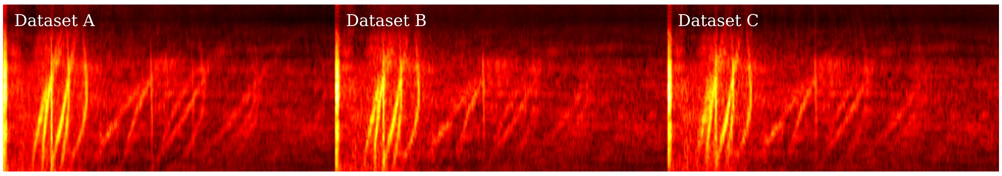
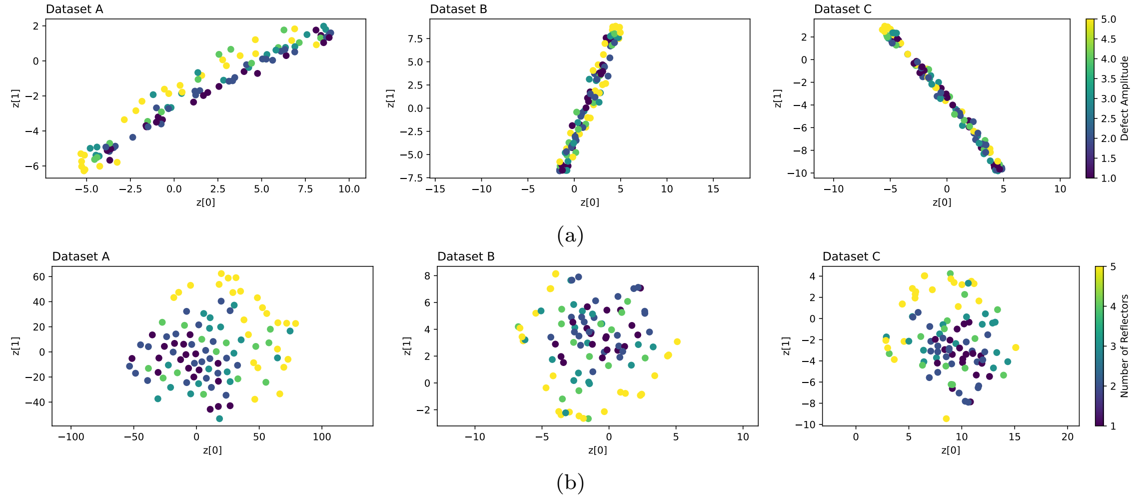


Figure 15: Spectrogram samples reconstructed from PCA models trained using a composition of 70% feature spectrograms and 30% Gaussian noise spectrograms. Each model used 1600 latent variables for the three distinct datasets. The horizontal axes represent time in seconds, while the vertical axes denote the frequency of the reconstructed spectrogram.

The silhouette score [33] is incorporated to assess the effectiveness of the latent representations. This metric, quantifying the compactness and separation of clusters, facilitates a rigorous evaluation. Positive silhouette scores indicate well-defined clusters, while negative scores suggest overlapping or unstructured latent space. By incorporating the silhouette score into our analyses of PCA and VAE models, we demonstrate how the quantitative assessment of their clustering performance on generated latent representations can be performed.



(a)

(b)

Figure 16: t-SNE reduction of (a) 4 and (b) 1600 latent variables recovered from test simulated samples for models trained using PCA.

The latent distributions, visualized using t-distributed stochastic neighbor embedding (t-SNE) [34], are shown in Figure 16. The t-SNE plots show a two-dimensional representation of 4 and 1600 PCs computed from the PCA models' latent variables. The structures captured by all training instances are evident visually that an increase in the number of latent variables from 4 to 1600 enhances the structure of the latent spaces. Silhouette scores further quantify this improvement. For PCA models with 4 PCs, silhouette scores are $-0.104, -0.107$, and -0.101 for datasets A, B, and C, respectively. Increasing the latent variables to 1600 results in improved scores of $0.242, 0.203$, and 0.197 for datasets A, B, and C, confirming the enhanced structural representation in the latent space.

Like the PCA models, reconstruction results of simulated spectrograms in Figure 17 and the t-SNE visualised latent variable plots shown in Figure 18 show similar performances for the three simulated training datasets. However, the VAE models can reconstruct underlying dispersion curve features from spectrograms and discard noise, compared to the PCA models. Therefore, the VAE models are able to separate coherent and incoherent features from the spectrograms. The PCA and VAE models exhibit comparable performance in capturing the structural characteristics of latent variables within this test dataset.

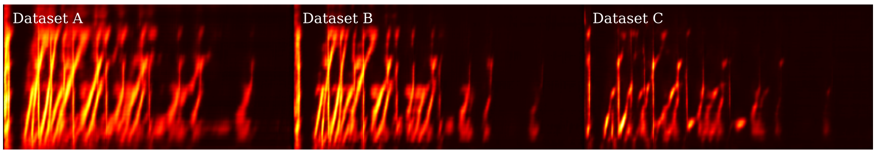


Figure 17: Spectrogram samples reconstructed from VAE models trained using a composition of 70% feature spectrograms and 30% Gaussian noise spectrograms. Each model used 1600 latent variables for the three distinct datasets. The horizontal axes represent time in seconds, while the vertical axes denote the frequency of the reconstructed spectrogram.

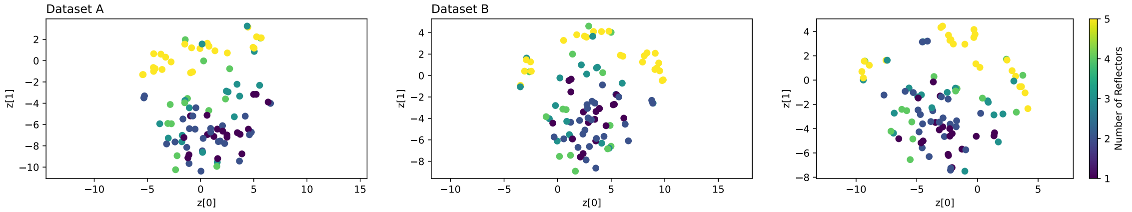
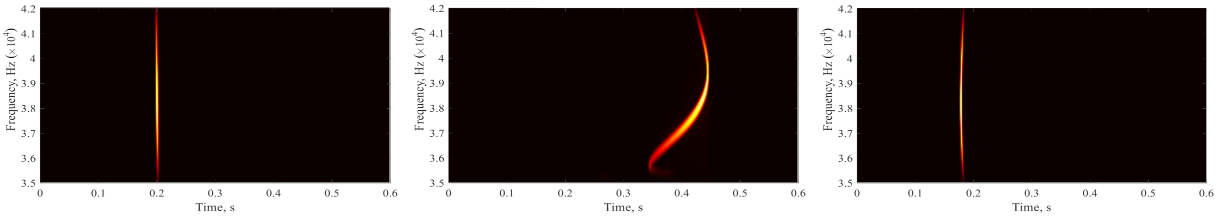


Figure 18: t-SNE reduction of VAE 1600 latent variables recovered from simulated test samples to visualise the number of reflectors in spectrograms.

4.4. Mode shapes and distances to reflectors

The measurements captured in experiments produce reflections from aluminothermic welds with unknown mode shapes, arrival times and attenuation. We use frequency-domain simulated single mode spectrograms to study the effect on the variance captured by our trained models. We show this, by studying the effect of mode shapes and arrival times in the latent space on our trained PCA and VAE models. Figure 19 shows the three main modes of propagation that are observed in the experimental spectrograms. In our test sample spectrograms, the simulated sets of modes are propagated to various distances and attenuated.



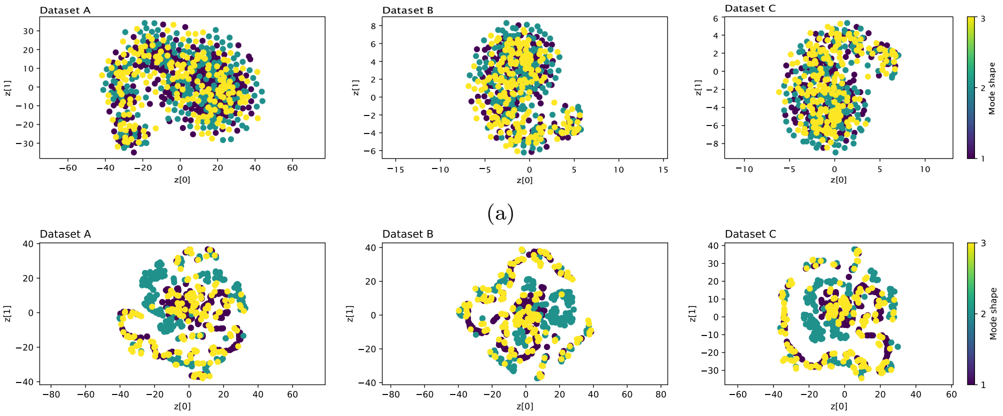
(a)

(b)

(c)

Figure 19: Samples of the three main modes of propagation observed in experimental measurements: (a) Mode 1, (b) Mode 2 and (c) Mode 3.

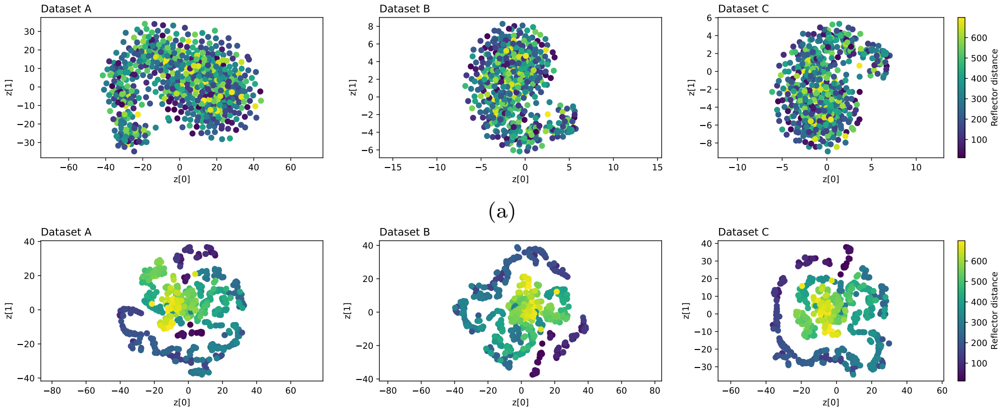
To study the effect of types of mode shapes and times of arrivals on the latent space, we employ the better-performing PCA and VAE models, which all have 1600 latent variables and 30% noise.



(a)

(b)

Figure 20: t-SNE visualisation of the mode shape type using 1600 latent variables: (a) PCA (top row) and (b) VAE (bottom row).



(a)

(b)

Figure 21: t-SNE visualisation of the reflector distance using 1600 latent variables: (a) PCA (top row) and (b) VAE(bottom row).

The latent space demonstrations in Figures 20 and 21 show that VAE models perform better than PCA models in interpreting the latent space. The VAE models better separate times of arrival or distances to welds of propagated modes. Furthermore, the VAE results also show that the first and third modes are interpreted as the same. This result is expected as these modes are very similar in shape. Distances to simulated weld reflectors are only disentangled to a greater degree by the VAE models.

4.5. Summary of results

Both PCA and VAE models were tested to optimise the training settings, assessing various noise percentages and latent variables. The PCA models showed sensitivity to including Gaussian noise spectrograms, performing best with 20-40% of Gaussian noise spectrograms and 1600 latent variables, particularly on Dataset A. Meanwhile, VAE models demonstrated enhanced performance, displaying the best reconstructions at 1600 latent variables and 30% noise on Dataset A. The study found that the VAE models performed better, distinguishing between coherent and incoherent features from spectrograms, effectively reconstructing dispersion curve features while discarding noise. Furthermore, analyses using t-SNE visualisation of the latent space revealed the VAE models' superior capability in interpreting reflector distances and mode shapes, outperforming the PCA models. These results underscore the significance of dataset parameter selection in optimising the training settings, particularly the proportion of Gaussian noise spectrograms to spectrograms with coherent features, offering insights into the performance differences between linear and non-linear models in capturing underlying features and disentangling mode shapes and reflector distances.

5. Experimental study

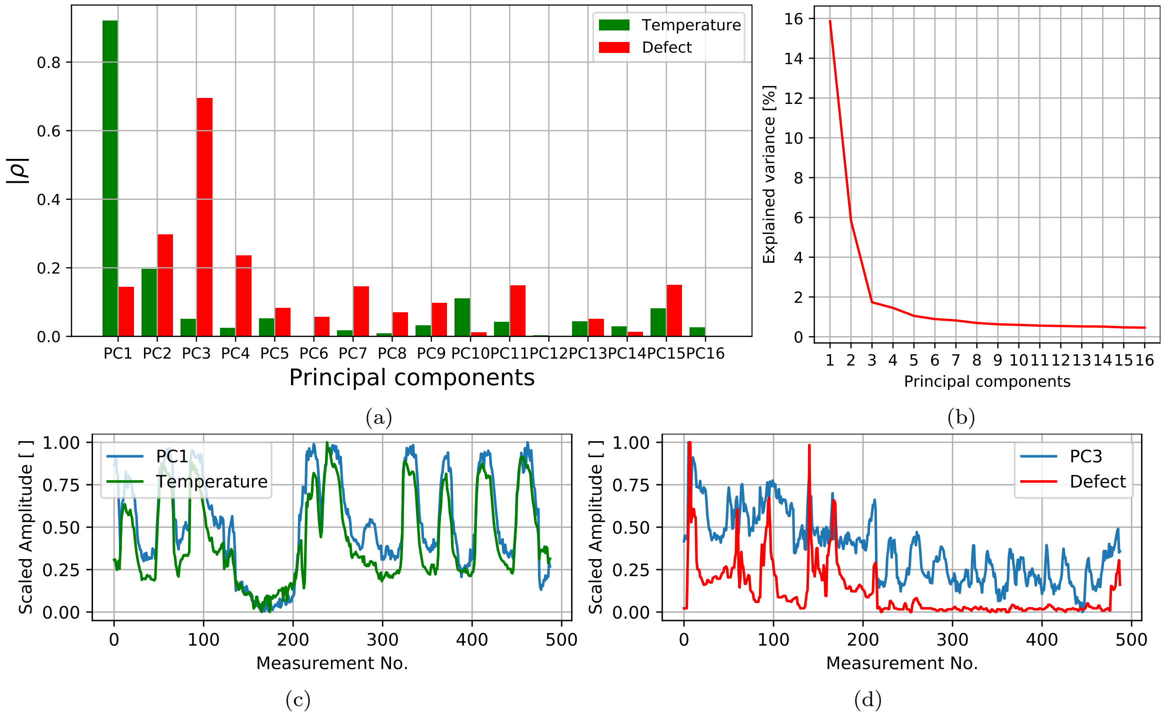
This section uses field-measured and simulated spectrogram data for training using the PCA and VAE models. In most instances, there is a limitation in field-measured data due to limited accessibility to testing sites and equipment. Our field measurements are limited to 488 spectrograms, which can be problematic for training using deep machine learning models. The measurements are analysed using PCA and VAE models, and both the recovered principal components or latent variables and reconstructed spectrograms are studied. By studying the two spaces, we can investigate the feasibility of using PCA and VAE models to study the effect of temperature change and growing defects in the rail.

5.1. Training using experimental data

Plots in Figure 22 show a comparison between the first 16 recovered principal components and their correlation to temperature and artificial defects. The absolute values of correlation coefficients $|\rho|$ were computed using the formula:

$$|\rho_{X,Y}| = \left|\frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}\right|, \tag{5}$$

where X is the feature of interest (temperature or growing defect), Y is the recovered principal vector recovered from applying the PCA model, and σ_X and σ_Y are the standard deviations of X and Y. The computed coefficients and explained variance plotted in Figures 22(a) and 22(b), show that temperature variance dominates the measured signals, while the variation of the defect is captured less accurately by the PCA model. These results are as expected as PCA is a variance-driven algorithm and temperature changes are expected to influence the overall variance of the signals. In contrast, the growing defect is expected to be a source component that influences fractional parts of the measured signals.

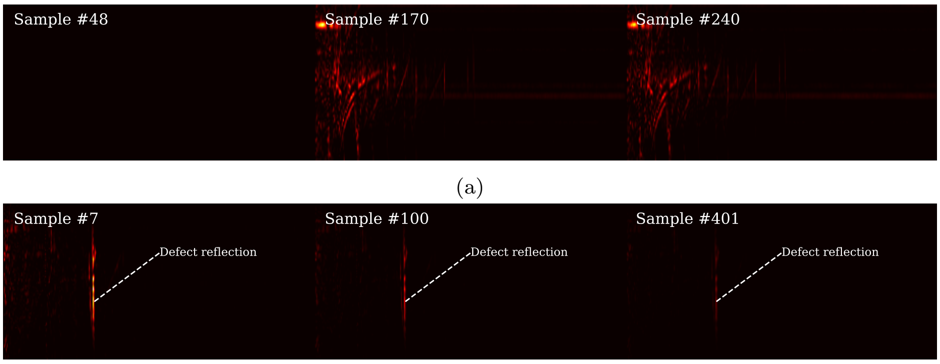


(a) (b)

(c) (d)

Figure 22: Plots showing the correlation between (a) the first 16 PCs and temperature and artificial defect amplitude variation, (b)PC1 and temperature, and (c) PC3 and (d) defect amplitude.

The latent variable or vectors, PC1 and PC3, shown in Figures 22(c) and 22(d) are studied further in reconstruction. The reconstructions of the PCs of interest are computed using the inverse transform and zeroing the weights of other principal components that are not of interest. By computing the inverse transform in this manner, we can reconstruct coherent features that are varied by the PCs of interest as shown in Figure 23. When the inverse transforms of PC1 are computed for various samples (samples number 48, 170 and sample 240, which are at high, mid and low temperatures, respectively), based on the vector seen in Figure 22(c), it is evident from Figure 23(a) that temperature influences the overall variance of the signal. The inverse transforms shown in Figure 23 indicate that the temperature decrease is negatively correlated to the overall energy received by the transducer over time. Reconstruction of PC3 shown in Figure 23(b) (for samples 7, 100 and 401, which are at high, mid and low defect reflection amplitudes, respectively) shows the varying defect reflection captured by PC3, however, with other coherent features. The other captured features result in an overall low correlation of PC3 with the varying defect reflection as seen in Figures 22(a) and (d).



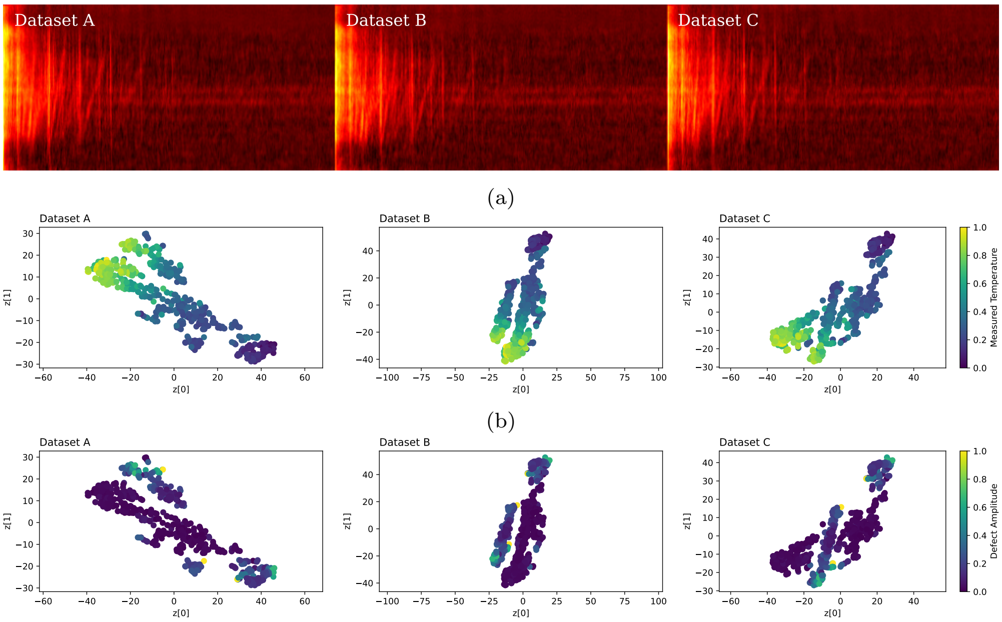
(a)

(b)

Figure 23: The PCA inverse transform of sample spectrograms using only (a) PC1 and (b) PC3. The horizontal axes represent time in seconds, while the vertical axes denote the frequency of the reconstructed spectrograms.

5.2. Training using simulated data

The trained models are also evaluated using a dataset of pulse-echo measurements conducted on the UIC60 rail and with sample spectrograms reconstructed in Figure 24.



(a)

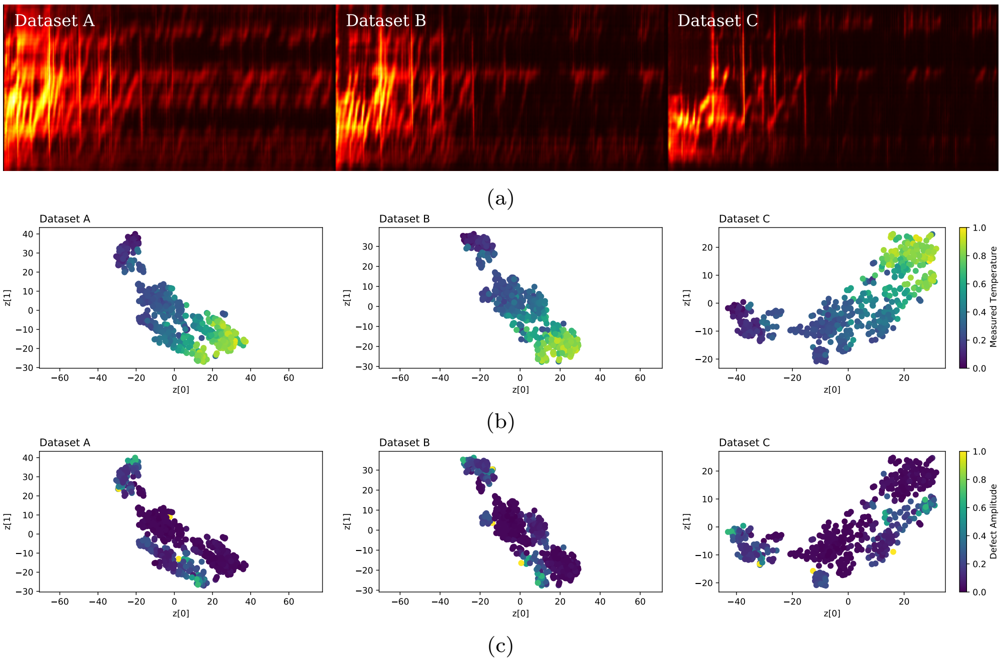
(b)

(c)

Figure 24: (a) Sample spectrograms reconstructed (with the horizontal axes representing the time in seconds, while the vertical axes denote the frequency of the reconstructed spectrograms), (b) t-SNE visualisation for temperature and (c) a growing defect using a PCA model trained using 1600 latent variables and three simulated datasets.

All the training instances accurately reconstruct the field spectrograms, with sample reconstructions shown in 24(a). However, the PCA model has poor extraction coherent features and denoising of the spectrograms. The variance of the temperature measurements taken in a cabinet near the rail (shown in Figure 5(b)) is expected to be captured for all the training instances. Latent variables of the model are interpreted using t-SNE, and all instances of training capture temperature variations accurately, as shown in Figure 24(b). Furthermore, the latent variables also capture the variation of the growing defect, as shown in 24(c).

The VAE model was also trained on simulated datasets and then used to analyse experimental data. The model was trained using a combination of 70% feature spectrograms, 30% noise spectrograms, and 1600 latent variables, as this was found to be the optimal setting through experimentation with the simulated data. Figure 25(a) shows reconstructions of an experimental sample using this setting for the three simulated datasets used for training, demonstrating that the VAE model can extract underlying coherent features in the spectrograms. The reconstructions using datasets B and C result for both models (ICA and VAE) resemble the multiple weld reflections rather than the isolated propagating weld reflections or modes as shown in Figures 8(b) and 8(c). This is expected, as the trained models minimise reconstruction errors in each test sample by overlaying learned modes from datasets B and C. This involves fitting individual modes or welds to the test spectrogram and then superimposing these fitted modes into a single spectrogram, resulting in the complete reconstruction of test spectrograms containing multiple modes.



(a)

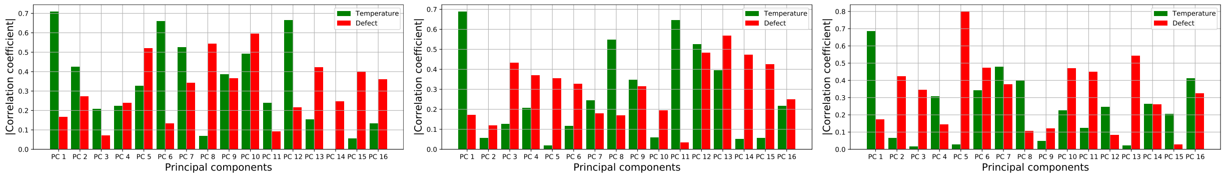
(b)

(c)

Figure 25: (a) Sample spectrograms reconstructed (with the horizontal axes representing the time in seconds, while the vertical axes denote the frequency of the reconstructed spectrograms), (b) t-SNE visualisation for temperature and (c) a growing defect using a VAE model trained using 1600 latent variables and three simulated datasets.

Comparing the t-SNE visualized latent variables of the VAE models (shown in Figures 25(b) and 25(c)) to the t-SNE variables of the PCA models (shown in Figures 24(b) and 24(c)), it can be seen that the latent spaces have similar distributions. This shows that the temperature and growing defect amplitude features are similarly mixed in the latent variables of both the PCA and VAE models.

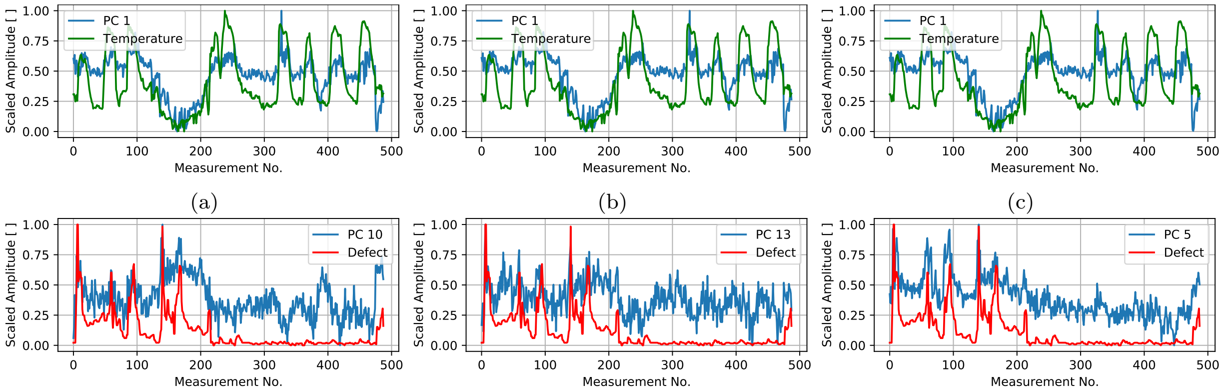
To investigate the effectiveness of the training method proposed in this paper further, we analyse the latent space using 16 latent variables on the field measurements. This demonstrates the advantage of the training method used in this paper (training using simulated data) compared to extracting latent variables or PCs directly from field measurements (training using field measurements). After training the PCA model using 16 latent variables and the three simulated datasets, the field measurements were used as test samples to compute the correlation of latent variables or PCs with the measured temperature and defect amplitude variation, as shown on the plots in Figure 26.



(a) (b) (c)

Figure 26: Plots showing the correlation between the first 16 PCs and the measured temperature and artificial defect amplitude for models trained using simulated datasets (a) A, (b) B and (c) C.

Based on the correlation plots, it is evident that the temperature and defect variations are mixed in many of the PCs, and less disentangled compared to PCs extracted directly from field measurements, as shown in Figure 22. We expect an interpretation-driven approach to show less mixing of components.



(a) (b) (c)

(d) (e) (f)

Figure 27: Plots the measured temperature and the best-correlated PC for models trained using simulated datasets (a) A, (b) B and (c) C, and the measured defect amplitude and the best-correlated PC for models trained using simulated datasets (d) A, (e) B and (f) C.

Further analysis of the plots shown in Figure 26 and the PCs that are best correlated to the temperature, as shown in Figures 27(a) to 27(c), show that training using simulated data does not provide the best setting for detecting temperature variations in the field measured data compared to directly applying the PCA model to the field measured spectrograms. However, plots of the PCs that are best correlated to the measured defect amplitude, shown in Figures 27(d) to 27(f), indicate that training using the three simulated datasets provides an improved setting for detecting the variation of the defect amplitude. Furthermore, a more significant correlation between PC5 and defect amplitude was observed for a PCA model trained using dataset C. This is expected as spectrograms in dataset C only have single dispersion curves that reconstruct source components such as the defect reflection.

6. Conclusions and future work

This study investigated machine learning models for reconstructing simulated and experimental rail guided wave-based spectrograms. Training data sets with different features are compared on different machine learning models: PCA (a linear variance-driven model) and VAE (a nonlinear variance-driven model). Our results show that the VAE model achieves better reconstruction of spectrograms and a more structured latent space than PCA models. This is expected since our VAE network has more learning layers than the PCA network. Training using spectrograms with multiple simulated reflectors shows better reconstruction than single-reflector and single-mode spectrograms. This is due to feature density in the spectrograms with multiple reflectors compared to the other spectrograms from other datasets. However, a drop in performance is realised when the ML models are trained using the multiple reflector data set with less varied features. When analysing the latent space, training using a simulated single-mode spectrogram dataset results in a better-structured latent space than other datasets. The proposed training and learning framework also showed good general performance for all three investigated datasets since the machine learning models learned all the simulated modes propagating in the rail and their variations. The results show that ML algorithms can learn and reconstruct simulated and experimental spectrograms when sufficient feature variations are contained in the simulated training data. This means that realistic rail GWU propagation features are captured in the models. The proposed framework is robust and helpful in deconstructing/learning features in experimental spectrograms where data is limited. Further investigation of the latent space and possible disentanglement of the variables can show different performances for the datasets proposed in this paper. In future work, we will use the proposed framework of training machine learning models to investigate interpretation-driven approaches and perform further feature learning of field-measured spectrograms, i.e., defect and temperature monitoring.

References

[1] F. Burger, P. Loveday, Ultrasonic broken rail detector and rail condition monitor technology, in: Proceedings of the 11th International Heavy Haul Association Conference (IHHA 2017), Cape Town, South Africa, 2017, pp. 275–280.

[2] R. Gorgin, Y. Luo, Z. Wu, Environmental and operational conditions effects on lamb wave based structural health monitoring systems: A review, Ultrasonics 105 (2020) 106114.

[3] D. A. Tibaduiza, L. E. Mujica, J. Rodellar, Damage classification in structural health monitoring using principal component analysis and self-organizing maps, Structural Control and Health Monitoring 20 (10) (2013) 1303–1316.

[4] A.-M. Yan, G. Kerschen, P. De Boe, J.-C. Golinval, Structural damage diagnosis under varying environmental conditions—part i: a linear analysis, Mechanical Systems and Signal Processing 19 (4) (2005) 847–864.

[5] M. Torres-Arredondo, J. Sierra-Pérez, G. Cabanes, An optimal baseline selection methodology for data-driven damage detection and temperature compensation in acousto-ultrasonics, Smart Materials and Structures 25 (5) (2016) 055034.

[6] D. P. Kingma, M. Welling, An introduction to variational autoencoders, Foundations and Trends in Machine Learning 12 (4) (2019) 307–392.

[7] I. D. Khurjekar, J. B. Harley, Closing the sim-to-real gap in guided wave damage detection with adversarial training of variational auto-encoders, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022, pp. 3823–3827.

[8] M. Rautela, S. Gopalakrishnan, Ultrasonic guided wave based structural damage detection and localization using model assisted convolutional and recurrent neural networks, Expert Systems with Applications 167 (2021) 114189.

[9] L. Lomazzi, S. Fabiano, M. Parziale, M. Giglio, F. Cadini, On the explainability of convolutional neural networks processing ultrasonic guided waves for damage diagnosis, Mechanical Systems and Signal Processing 183 (2023) 109642.

[10] H. Zhang, J. Lin, J. Hua, T. Zhang, T. Tong, Attention-based interpretable prototypical network towards small-sample damage identification using ultrasonic guided waves, Mechanical Systems and Signal Processing 188 (2023) 109990.

[11] P. W. Loveday, C. S. Long, D. A. Ramatlo, Ultrasonic guided wave monitoring of an operational rail track, Structural Health Monitoring 19 (6) (2020) 1666–1684.

[12] M. Silva, A. Santos, R. Santos, E. Figueiredo, C. Sales, J. C. Costa, Deep principal component analysis: An enhanced approach for structural damage identification, Structural Health Monitoring 18 (5-6) (2019) 1444–1463.

[13] F. Le Bourdais, O. Mesnil, O. D'Ameida, Machine-Learning Based Temperature Compensation For Guided WaveImaging In Structural Health Monitoring, in: 11th International Symposium on NDT in Aerospace, Nov 2019, Paris-Saclay, France (AeroNDT 2019), 2020, pp. 1–16.

[14] D. A. Ramatlo, W. D. N., L. P. W., A data-driven hybrid approach to generate synthetic data for unavailable damage scenarios in welded rails for ultrasonic guided wave monitoring, Structural Health Monitoring 0 (0) (0) 14759217231197265.

[15] V. Iglovikov, A. Shvets, Ternausnet: U-net with vgg11 encoder pre-trained on ImageNet for image segmentation (2018). arXiv:1801.05746.

[16] P. Loveday, P. Wilcox, Guided wave propagation as a measure of axial loads in rails, in: SPIE conference proceedings, Vol. 7650, 2010, pp. 7650 – 7650 – 8.

[17] P. Welch, The use of Fast Fourier Transform for the estimation of Power Spectra: A method based on time averaging over short, modified periodograms, IEEE Transactions on Audio and Electroacoustic au-15 (2) (1967) 70–73.

[18] L. Gavrić, Computation of propagative waves in free rail using a finite element technique, Journal of Sound and Vibration 185 (3) (1995) 531–543.

[19] T. Hayashi, W.-J. Song, J. Rose, Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example, Ultrasonics 41 (3) (2003) 175–183.

[20] I. Bartoli, A. Marzani, F. Lanza di Scalea, E. Viola, Modeling wave propagation in damped waveguides of arbitrary cross-section, Journal of Sound and Vibration 295 (3-5) (2006) 685–707.

[21] V. Damljanović, R. Weaver, Propagating and evanescent elastic waves in cylindrical waveguides of arbitrary cross section, The Journal of the Acoustical Society of America 115 (4) (2004) 1572–1581.

[22] P. Loveday, Analysis of piezoelectric ultrasonic transducers attached to waveguides using waveguide finite elements, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Controlultrasonics, ferroelectrics, and frequency control 54 (10) (2007) 2045 – 2051.

[23] D. A. Ramatlo, C. S. Long, P. W. Loveday, D. N. Wilke, Physics-based modelling and simulation of reverberating reflections in ultrasonic guided wave inspections applied to welded rail tracks, Journal of Sound and Vibration 530 (2022) 116914.

[24] D. A. Ramatlo, C. S. Long, P. W. Loveday, D. N. Wilke, A modelling framework for simulation of ultrasonic guided wave-based inspection of welded rail tracks, Ultrasonics 108 (2020) 106215.

[25] C. S. Long, P. W. Loveday, Prediction of guided wave scattering by defects in rails using numerical modelling, in: AIP Conference Proceedings, Vol. 1581, 2014, pp. 240–247.

[26] M. J. S. Lowe, D. N. Alleyne, P. Cawley, The Mode Conversion of a Guided Wave by a Part-Circumferential Notch in a Pipe, Journal of Applied Mechanics 65 (3) (1998) 649–656.

[27] C. Chongyi, W. Chengguo, J. Ying, Study on numerical method to predict wheel/rail profile evolution due to wear, Wear 269 (3-4) (2010) 167–173.

[28] J. Gerlici, T. Lack, Railway wheel and rail head profiles development based on the geometric characteristics shapes, Wear 271 (1-2) (2011) 246–258.

[29] P. Wang, J. Xu, K. Xie, R. Chen, Numerical simulation of rail profiles evolution in the switch panel of a railway turnout, Wear 366-367 (2016) 105–115.

[30] M. Ignesti, M. Malvezzi, L. Marini, E. Meli, A. Rindi, Development of a wear model for the prediction of wheel and rail profile evolution in railway systems, Wear 284-285 (2012) 1–17.

[31] I. I. Setshedi, P. W. Loveday, C. S. Long, D. N. Wilke, Estimation of rail properties using semi-analytical finite element models and guided wave ultrasound measurements, Ultrasonics 96 (2019) 240–252.

[32] J. C. Helton, F. J. Davis, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, Reliability Engineering and System Safety 81 (1) (2003) 23–69.

[33] A. Naghizadeh, D. N. Metaxas, Condensed silhouette: An optimized filtering process for cluster selection in k-means, Procedia Computer Science 176 (2020) 205–214, knowledge-Based and Intelligent Information and Engineering Systems: Proceedings of the 24th International Conference KES2020.

[34] M. Wattenberg, F. Viégas, I. Johnson, How to use t-sne effectively, Distill 1 (10) (2016) e2.