Synthetic data for guided wave ultrasound monitoring of rail tracks generated using statistical modelling and variational auto-encoders

D.A. Ramatlo[1], D.N. Wilke[1], P.W. Loveday[2]
[1] University of Pretoria, Department of Mechanical and Aeronautical Engineering,
Pretoria, 0002, South Africa
e-mail: **dineo.ramatlo@up.ac.za**

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

Abstract
Monitoring rail tracks may involve the complicated propagation of ultrasonic guided waves over several hundred metres. The measured response is not only affected by dispersion, attenuation and scattering of the waves from discontinuities in the rail but also due to complex interactions of the rail with the ballast foundation, sleepers, clips, and cables. Physics-based models are limited in capturing the latter complexities when predicting guided wave propagation. This paper investigates the fusion potential of statistical and deep learning data-driven generative models to augment physics-based modelled responses with experimentally measured reflections for application in guided wave monitoring. Towards this, a virtual study is first conducted where a prototype physics-based model's response of a welded rail track is augmented with an additional simulated weld reflection. Future work will extend this towards augmenting prototype responses with synthetic damage signals towards the development of digital twins to maintain rail tracks.

1 Introduction

One of the attractive properties of guided wave ultrasound (GWU) is the ability to interrogate or monitor large volumes of a structure from a single transducer location, especially when compared with the conventional ultrasonic inspection. This feature is exploited in a permanent monitoring system for railway tracks which is currently being used to monitor a heavy haul line in South Africa [1,2]. Although seven rail breaks have been reported since the installation of the system in 2016, several major flaws have been recorded as false alarms [2] as the system was designed to detect complete breaks. Defects such as cracks can be detected by designing systems that can distinguish structural features from damage, using knowledge of how modes interact with different geometrical features [3]. To do this, ultrasonic signals containing reflections from benign structural features such as welds and potentially small reflections from growing damage are required [3]. However, obtaining monitoring data for different damage scenarios is virtually impossible for railway lines since detected defects in sections of an operational rail track are immediately removed and replaced with new rail. Laboratory damage experiments are also not possible due to end reflections from short sections of rail dominating the response. Modelling and simulation thus become increasingly important to generate synthetic signals for unavailable damage scenarios.

The data collected from a rail in the field contain reflections from aluminothermic welds joining sections of rail together. These reflections are characterized by a time and location-dependent, complicated behaviour introduced by the complex nature of guided wave propagation and varying environmental and operational conditions (EOCs). The multiple propagating modes are dispersive and are subject to attenuation due to material damping and damping introduced by the ballast foundation and sleepers. EOCs, such as

temperature, may vary along the length of the rail and with time. Similarly, the foundation of the ballast could also vary at different locations in the rail, and the welds used to join lengths of rail have distinct wave reflection characteristics.

Synthetic data for GWU monitoring can be generated from a digital twin model consisting of a digital replica of a physical inspection system [4]. If this digital replica can account for the complexities discussed above, it would be possible to model and simulate realistic damage signals. One important aspect to consider when generating synthetic data is the inclusion of damage signatures in the experimental measurements collected from the field. Ideally, it should not be possible to distinguish the *fake* damage reflection from the *real* weld reflections.

The authors of this paper have recently demonstrated a physics-based framework to model and simulate ultrasonic inspections in 1D waveguides [4,5]. The modelling framework can account for various complex features such as scattering from discontinuities, dispersion and attenuation that are apparent in the inspection data from a real system. However, the framework does not capture the coherent noise from EOCs and other complex features. Data-driven approaches have the potential to account for this coherent noise and ultimately augment the simulated damage with experimental reflections.

This paper investigates data-driven generative modelling approaches based on statistical and deep learning to augment idealistic physics-based modelled responses with virtual experimentally measured reflections for welded railway lines. The statistical model uses principal component analysis (PCA), and the deep learning model is based on variational auto-encoders (VAE). The input for the two investigated approaches is a virtual experimental response of a welded rail simulated with the modelling framework in [5]. The generative models augment this signal with an additional simulated reflection from an extra weld. This paper explores the potential of linear statistical models and quantifies the additional benefit of nonlinear deep learning models for this task. In future, the two data-driven approaches will be used to model and simulate the inclusion of coherent noise from EOCs and other complex features.

The virtual experimental response is presented in Section 2 so that the nature of these signals can be understood. The principal component analysis and variational auto-encoder modelling approaches are described in Sections 3 and 4, respectively. Modelling results obtained from the two approaches are presented in Section 5, and conclusions and recommendations for future improvements are discussed in Section 6.

2 Virtual experimental response

2.1 Physics-based modelling and simulation of the virtual experimental responses

A field experiment was performed on a UIC60 rail in an operational heavy-haul rail track. The rail was made up of 240m long sections welded together by aluminothermic welding. The field experiment was performed by exciting the guided waves propagating in the head of a rail with a piezoelectric transducer located at a distance of approximately 78m from the nearest weld. Figure 1 shows the setup for this experiment. The transducer was driven by a 17.5 cycle Hanning windowed tone burst voltage signal with a centre frequency of 35kHz. The excited guided waves were transmitted in both directions along the rail. The experimental response was constructed by measuring the reflections from welds using the transducer. The layout of the section of rail and the target mode shape with energy concentrated in the head of the rail is shown in Figure 1.

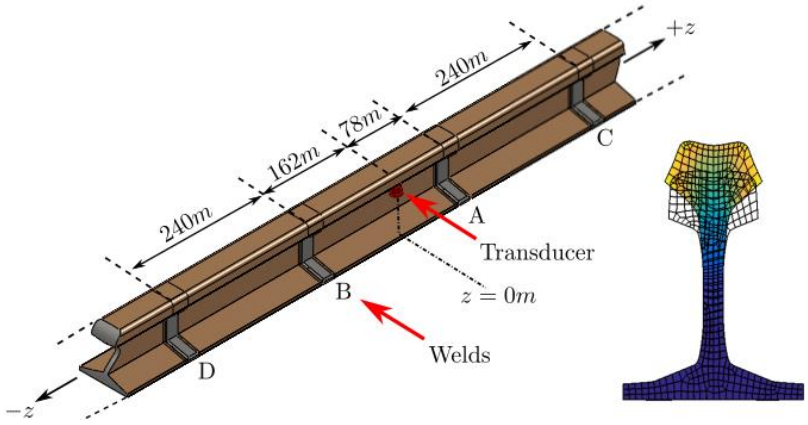

Figure 1: Setup of the rail and the target mode shape with energy concentrated in the head of the rail.

The virtual experimental response (which will serve as our baseline signal) was simulated using the modelling framework presented in [5]. The framework models the excitation, propagation and scattering of guided waves from discontinuities. The excitation of guided waves in the rail using a piezoelectric transducer was modelled using a hybrid model which couples a 3D FEM model of a transducer and a 2D semi-analytical finite element (SAFE) model of the waveguide. Propagation in a section of the rail having a constant cross-sectional area was modelled using a SAFE method which discretizes the displacement field across the cross-section of the waveguide using 2D finite elements. The propagation of guided waves along the waveguide length is achieved by applying analytical variations in the direction of propagation. The scattering of guided waves from reflectors is modelled using a second hybrid model, which couples a 3D FEM model of the reflector with two SAFE models to represent the semi-infinite incoming and outgoing rails on either side of the reflector. The modelling framework accounts for dispersion and attenuation in the rail. Figure 2 shows a spectrogram of the virtual experimental response and the equivalent envelope of the distance domain signal. The response was validated using the measured experimental response in reference [4].

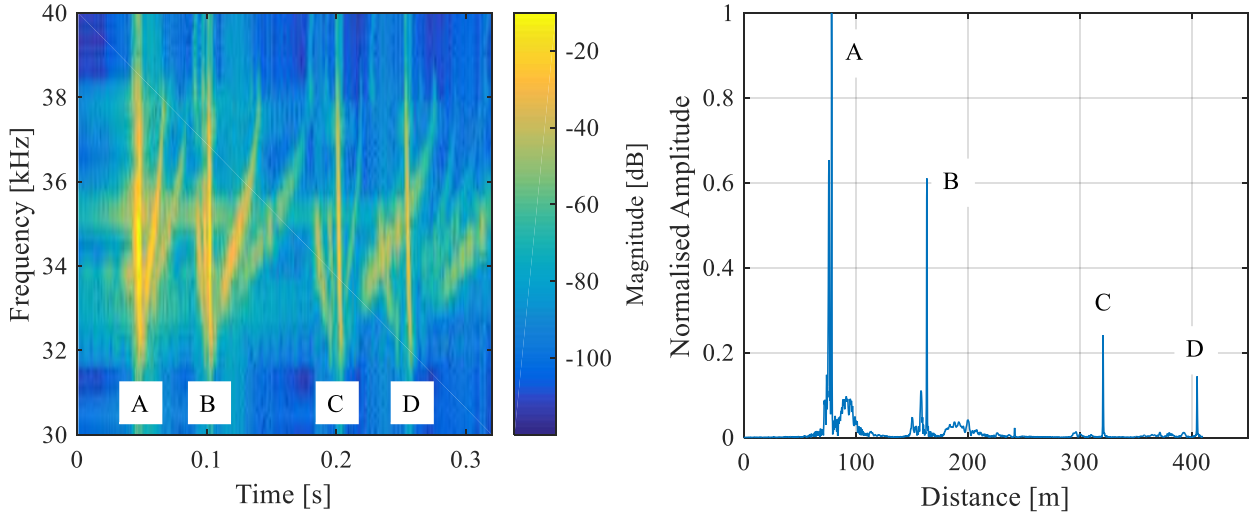

Figure 2: Spectrogram of the baseline for the virtual experimental response and the equivalent distance domain signal.

The spectrogram shows that the virtual experiment contains many modes of propagation that are reflected from each weld. The energy of the reflections decays exponentially with time due to attenuation in the rail. The mode with energy concentrated in the head of the rail was strongly excited and appeared to be almost non-dispersive. This mode is identified by a vertical trace in the time-frequency domain spectrogram and

the highest amplitude for each weld in the distance domain signal. Other modes reflecting from welds are also evident. These modes are very dispersive as their propagation velocity differs as a function of frequency. Some of these modes are coupled in pairs implying that a particular incident mode is reflected as a different mode from the weld. Another important feature of the virtual experimental response is that some modes reflected from different welds overlap with each other.

2.2 Description of the input data for PCA and VAE

One important aspect to consider when generating synthetic data is the inclusion of damage signatures in the experimental measurements collected from the field. Ideally, it should not be possible to distinguish the *fake* damage reflection from the *real* weld reflections. In this paper, we consider the case of a *fake* weld reflection and *real* virtual weld reflections. Given the virtual experimental response with reflections from four welds (Figure 2), this paper aims to use PCA and VAE to fuse a new *fake* weld reflection at a given distance. Both PCA and VAE require two sets of data as inputs to solve this task. The first set of input data is the virtual experimental response containing reflections from four welds and random distances where the reflection for the fifth weld should be added in each replica. The virtual experimental response thus serves as a baseline signal for the data-driven models. The second set of input data is the corresponding output for each pair of the baseline signal and the given distance value. This set of data will thus contain reflections from five welds, and the PCA and VAE models will need to learn the mapping from the baseline signal with four welds to the target signal with five welds. Figure 3 shows examples of target signals containing the reflections from the fifth weld, which was added at a random position.

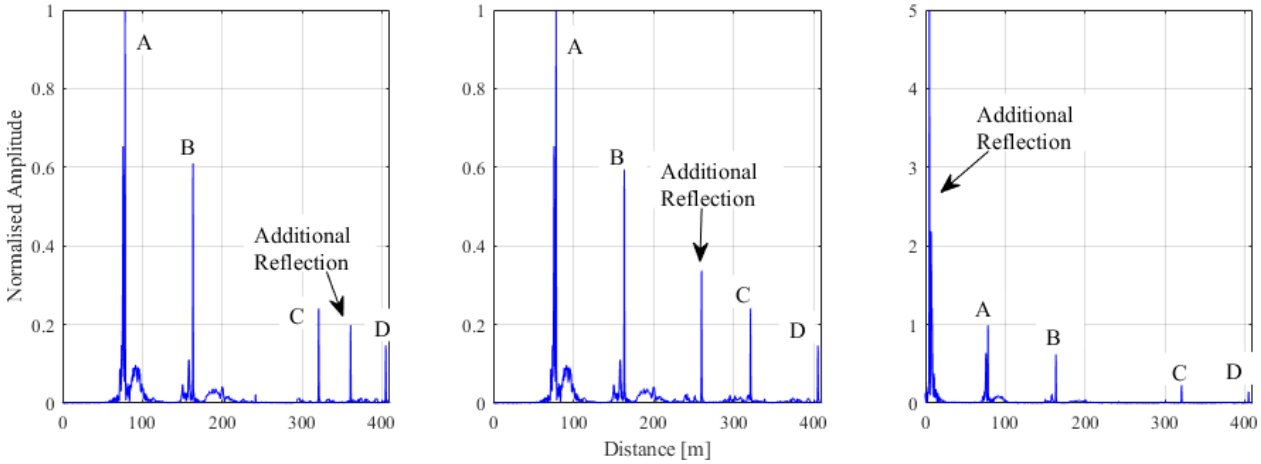

Figure 3: Examples of target signals containing fused reflections at distances of 357m, 257m and 5m away from the transducer, respectively.

To train the two models, $n = 1204$ observations of $m = 2400$ −dimensional data were generated by modelling and simulating the response for a rail with five welds. The positions for welds A, B, C and D were fixed according to Figures 1 and 2. The fifth weld was located at distinct positions between the transducer and each of the farthest welds on the two sides of the transducer. The input data were normalized according to the peak of the reflection for weld A.

3 Statistical modelling using PCA

The input data described in Section 2 was pre-processed by calculating the difference between the target signals with five weld reflections and the baseline signal with four weld reflections. Residual plots for observations in Figure 3 are plotted in Figure 4.

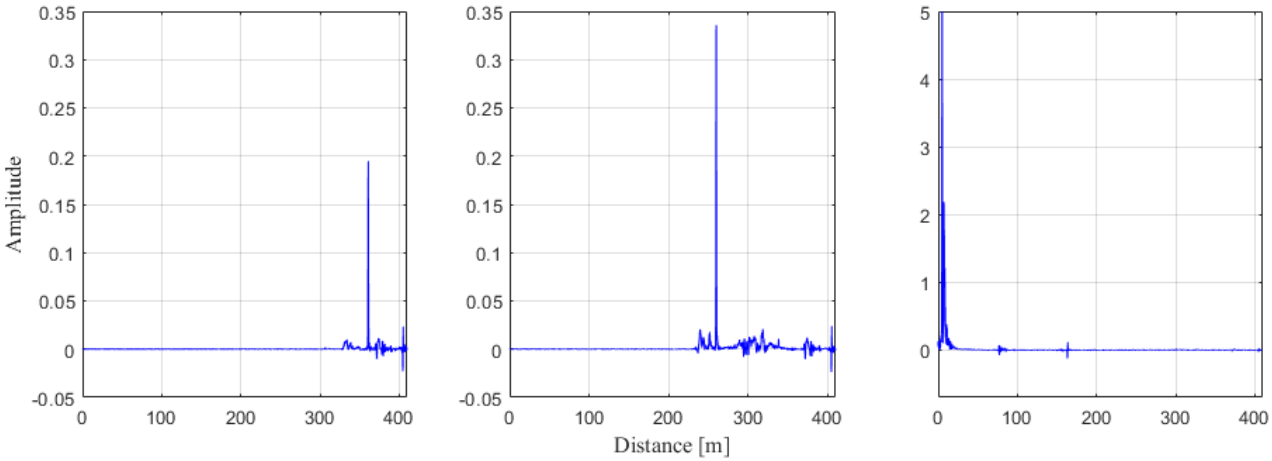

Figure 4: Residual signals for PCA input data.

Principal component analysis is a dimensionality reduction procedure that transforms a set of high-dimensional input observations with correlated variables into a set of low-dimensional linearly uncorrelated variables [6]. Given a set of $n = 1204$ residual signal observations (Figure 4) $\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \cdots, \tilde{\boldsymbol{x}}_n$ where each observation is represented by a vector of length m, the input dataset can be represented by a matrix $\tilde{\boldsymbol{X}}_{n\times m}$ where

$$\tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{x}}_1 \quad \tilde{\boldsymbol{x}}_2 \quad \cdots \quad \tilde{\boldsymbol{x}}_n]^T \tag{1}$$

The deviation of each $\tilde{\boldsymbol{x}}_i$ from the average is defined as:

$$\boldsymbol{\Phi}_i = \tilde{\boldsymbol{x}}_i - \boldsymbol{\mu} \tag{2}$$

where $\boldsymbol{\mu}$ is the average vector of all observations. PCA is a transformation of $\boldsymbol{\Phi}_i$ into a new vector $\boldsymbol{\Upsilon}_i$ by:

$$\boldsymbol{\Upsilon}_i = \boldsymbol{U}^T \boldsymbol{\Phi}_i \tag{3}$$

where $\boldsymbol{U}$ is the $m \times m$ orthogonal matrix whose rows are the principal components. The matrix $\boldsymbol{U}$ can be calculated based on the principle of covariance:

$$\boldsymbol{C} = \frac{1}{n}\boldsymbol{A}\boldsymbol{A}^T \tag{4}$$

where $\boldsymbol{C}$ is the $m \times m$ symmetric covariance matrix and $\boldsymbol{A} = [\boldsymbol{\Phi}_1 \quad \boldsymbol{\Phi}_2 \quad \cdots \quad \boldsymbol{\Phi}_n]$. The diagonals of $\boldsymbol{C}$ are the variances, and the off-diagonal terms are the covariances of the problem. The columns of the transformation matrix $\boldsymbol{U}$ are the eigenvectors of the covariance matrix $\boldsymbol{C}$. PCA thus solves an eigenvalue problem:

$$\lambda_j \boldsymbol{u}_j = \boldsymbol{C}\boldsymbol{u}_j, j = 1,2,\cdots,m \tag{5}$$

where λ_j is one of the eigenvalues of **C** and $\boldsymbol{u}_j$ is the corresponding eigenvector.

The output matrix $\boldsymbol{\Upsilon}_i$ computed from equation 3 contains projections of the original features onto the principal components. To reduce the dimensionality of the data, we choose a smaller number ($k < \min(m, n)$) of eigenvectors having the largest eigenvalues. Such eigenvalues correspond to components with maximum variances, and the remaining $(n - k)$ components will generally contain noise. k principal components from the representation matrix $\boldsymbol{U}$ in equation 3 can thus be used to reconstruct the new vectors $\boldsymbol{\Upsilon}_i$ which are a low dimensional representation of $\tilde{\boldsymbol{x}}_i$ vectors. The reconstructed target signals $\tilde{\boldsymbol{y}}_i$ (Figure 3) with the fifth weld reflection included, can finally be constructed by adding $\boldsymbol{\Upsilon}_i$ to the distance domain baseline signal (Figure 2).

Figure 5 plots the cumulative explained variance as a function of principal components for the problem considered in this paper. The plot shows that at least 600 components are required to retain almost 98% of the features in the input dataset.

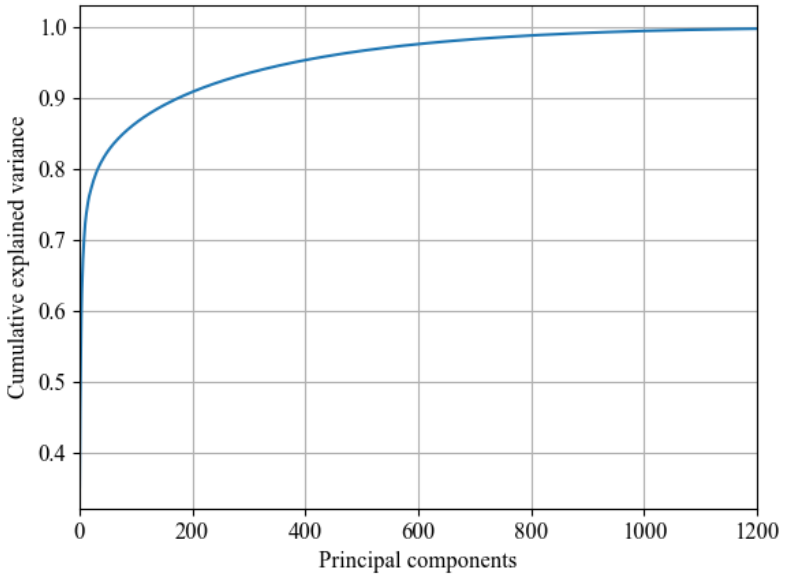

Figure 5: Cumulative explained variance as a function of principal components.

The first 600 PCA components were selected for dimensionality reduction. PCA can be used to generate new response signals containing new reflections at given positions by sampling from the latent space and mapping from 600 components back to the original dimension of the input data. To visualize how the latent space varies as a function of the position of the new reflection, the 600 components were projected on a 2D manifold representation using the Uniform Manifold Approximation and Projection (UMAP) technique, Figure 6.

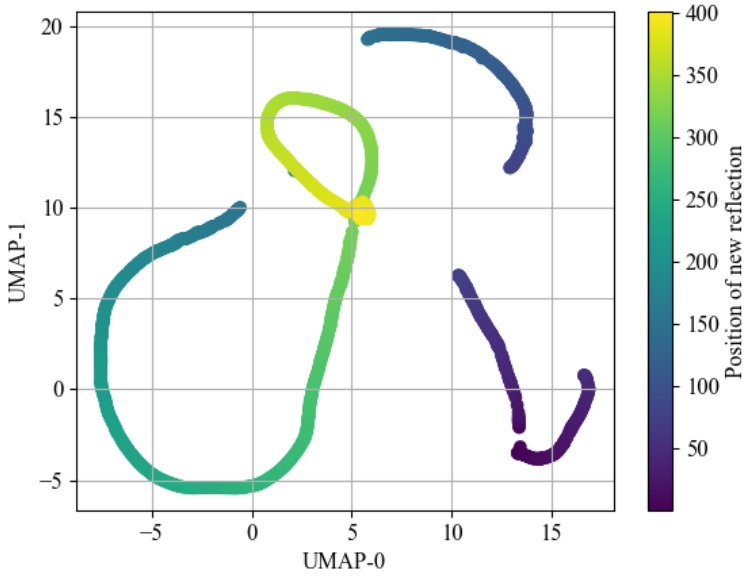

Figure 6: Cumulative explained variance as a function of principal components.

4 Variational auto-encoder

VAE is a generative model, and like PCA, it is also based on the principle of dimensionality reduction. A VAE consists of an encoder network and a decoder network connected through a latent space $\boldsymbol{z}$. The encoder compresses the input data to a lower-dimensional space which maps the data to a continuous latent vector $\boldsymbol{z}$. The decoder uses the latent variable to generate data which maps $\boldsymbol{z}$ to the reconstructed data approximating the input.

Given a set of m distance domain response signals with reflections from four welds, vectors $\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_n}$ where $i = 1,2, \cdots, n = 1204$ and distances $\boldsymbol{d} = (d_1, d_2, \cdots, d_n)^T$, we want to force the introduction of a new reflection from the fifth weld at a distance of d_i meters from the transducer location in the response signal $\boldsymbol{x_i}$. The n replicas of the distance domain baseline response in Figure 2 form the main inputs for our VAE. The second meaningful inputs are the distances where the fifth weld should be added. The VAE should be trained to approximate the target signals $\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_n}$ with new reflections added according to the distance d_i for each $\boldsymbol{y_i}$.

The encoder receives as input 3D sequences resulting from the concatenation of n baseline signals and the embeddings of the distances d_i. The encoder consists of a stack of 3 sequential 1D convolutional layers connected to 2 dense layers to approximate the mean and the variance of the 2D latent distribution. The decoder samples from the 2D latent distribution and forms 3D sequences by using a *RepeatVector* layer. The generated sequences are then concatenated back with the original d_i embeddings and are passed through a stack of two dense layers to increase dimensionality. A *TimeDistributed* layer is then applied to train the VAE to add new weld reflections, and the output is passed through a stack of 3 1D deconvolution layers to continue increasing the dimensionality back to a length of m, to approximate the reconstruction $\tilde{\boldsymbol{y}}_{\boldsymbol{i}}$ of the target distance domain signals $\boldsymbol{y_i}$.

The VAE is trained by minimizing the reconstruction loss, which is the mean squared error between the original target $\boldsymbol{y_i}$ from Figure 3 and the approximation $\tilde{\boldsymbol{y}}_{\boldsymbol{i}}$.

5 Results

5.1 Reconstructed responses

The distance domain results reconstructed using PCA and VAE are plotted and compared to the original target responses in Figures 7 to 10. The fifth weld was added at distances of 30m, 120m, 200m and 350m, respectively. The amplitude responses are also plotted on a log scale. Figure 7 shows that both the VAE and PCA could fuse the fifth weld reflection at a distance of 30m from the transducer. The two approaches also yielded a reasonable reconstruction of the amplitude, though the VAE overestimated the amplitudes for the four original welds.

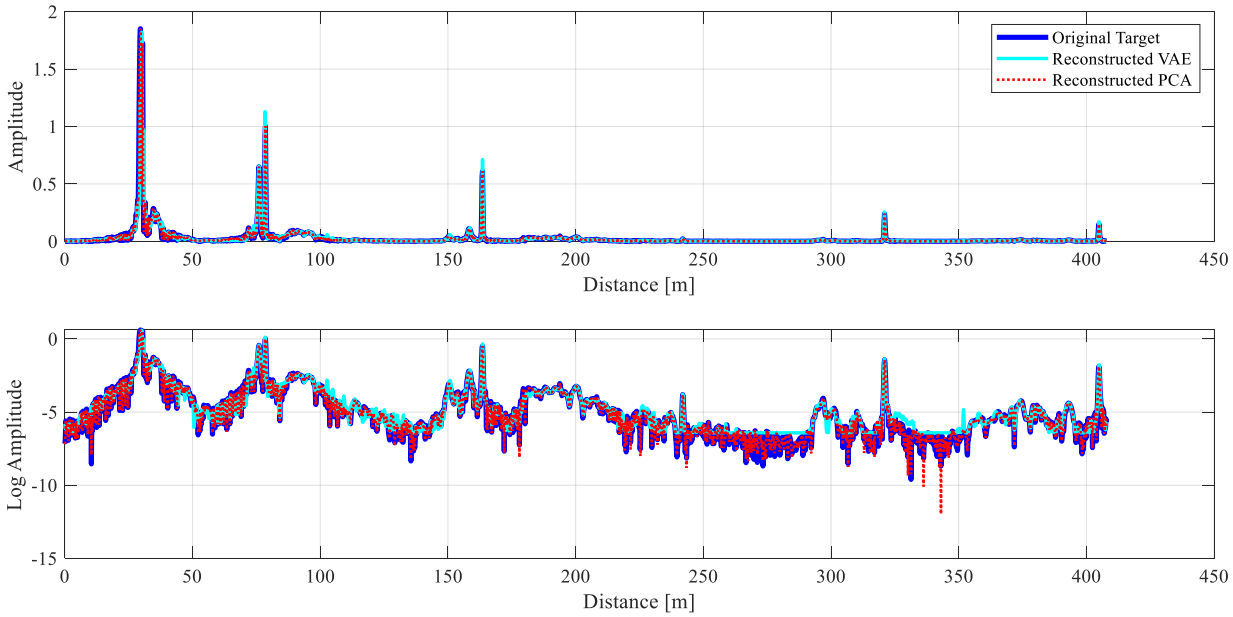

Figure 7: Reconstructed response with fused reflection at 30m.

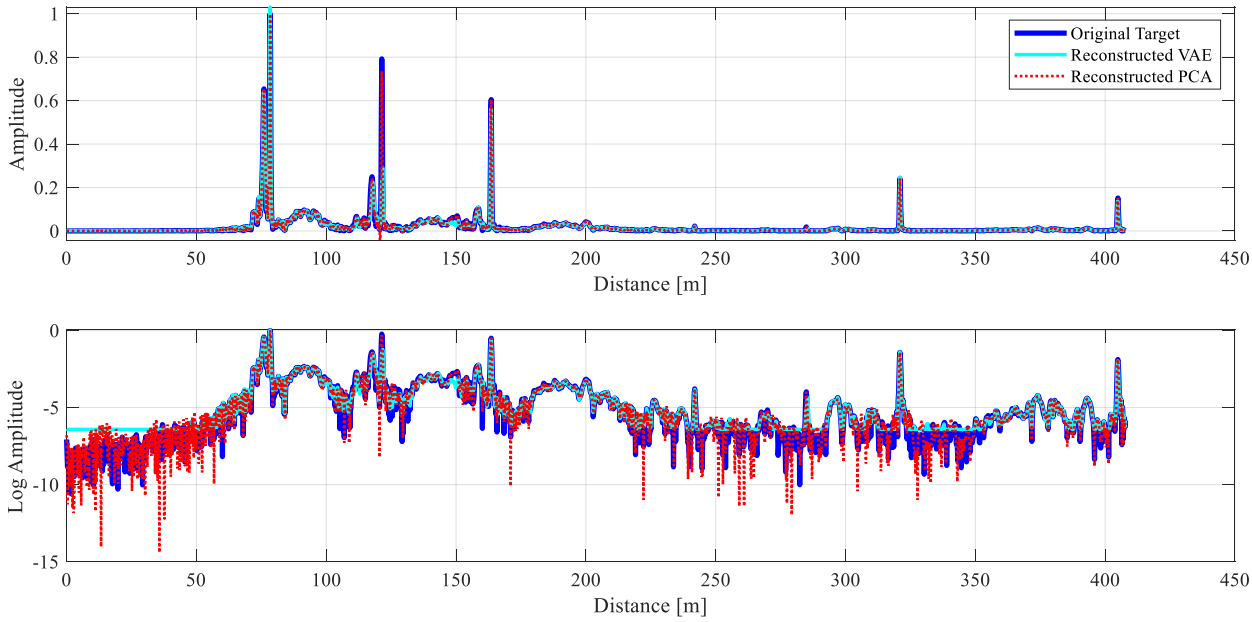

Figure 8: Reconstructed response with fused reflection at 120m.

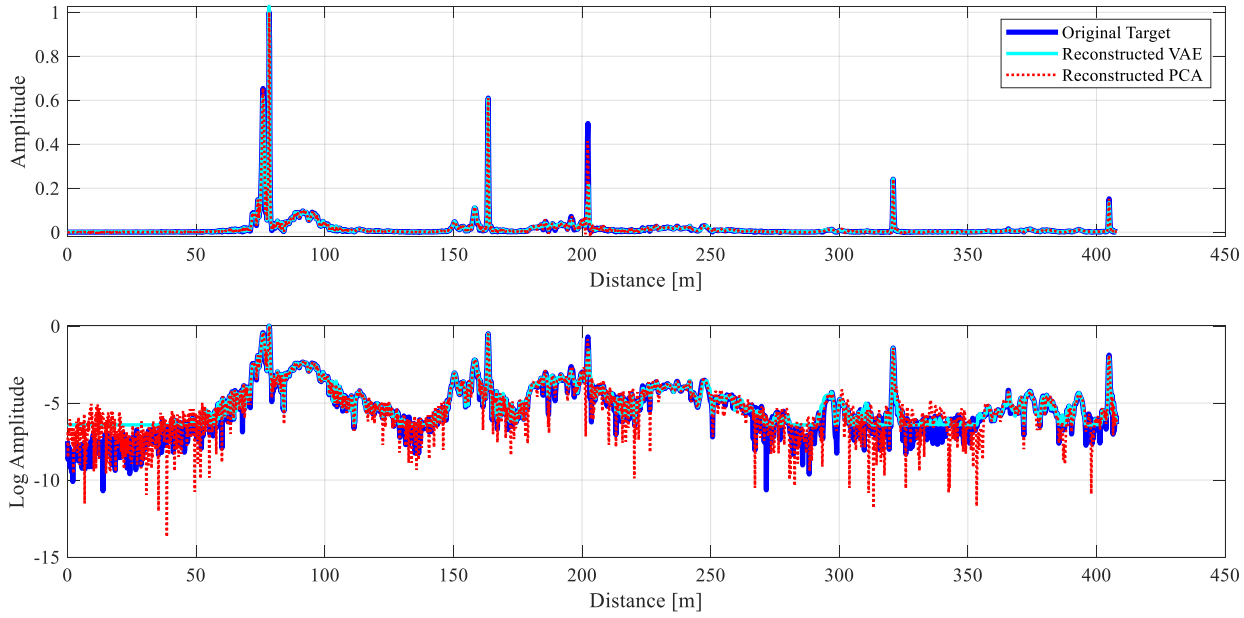

Figure 9: Reconstructed response with fused reflection at 200m.

In Figure 8, we notice that PCA continues to yield a good approximation for the new reflection at 120m while VAE struggles to yield a better amplitude. However, the position of the new reflection was well predicted. This trend in VAE results is also evident in Figures 9 and 10, when new reflections were fused at 200m and 350m, respectively. The VAE seem to be clipping the parts of the signal containing noise. In Figures 9 and 10, we also notice that as the new reflections move away from the transducer, the PCA method starts to yield bad approximations of the amplitude and starts to clip away noise information in the signal. Figure 11 compares the predicted amplitude and error for the fifth reflection at different positions for 1204 observations used as input data. This plot suggests that PCA outperformed the VAE approach. It is worth mentioning that the reconstruction ability of the two approaches looks promising.

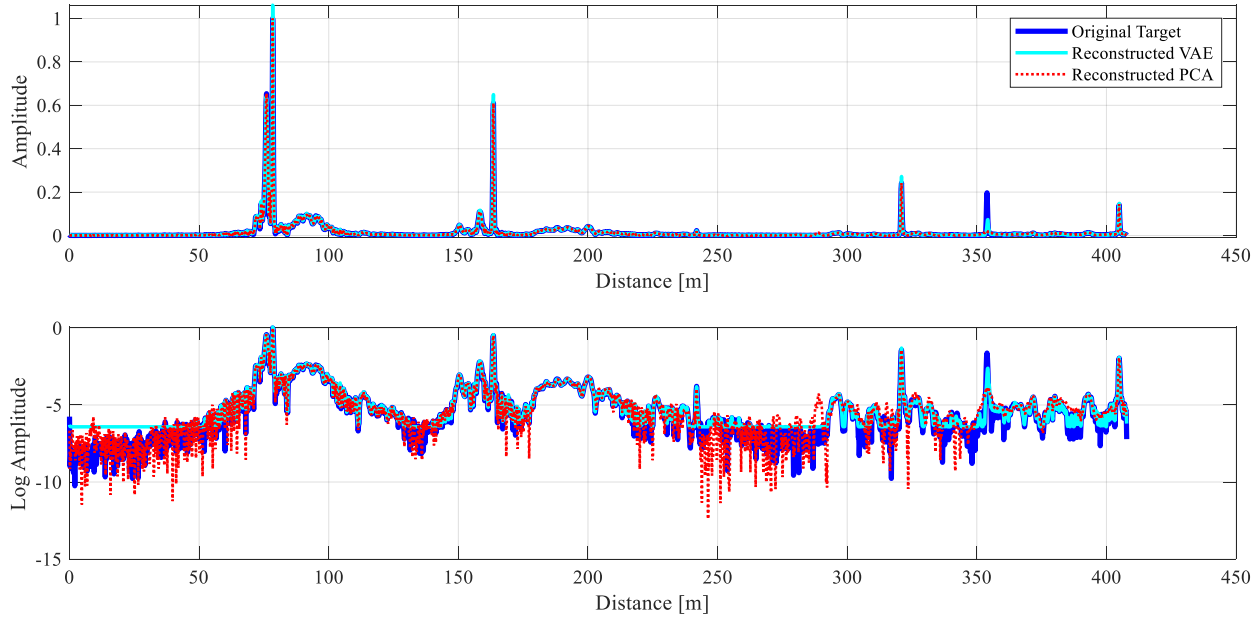

Figure 10: Reconstructed response with fused reflection at 350m.

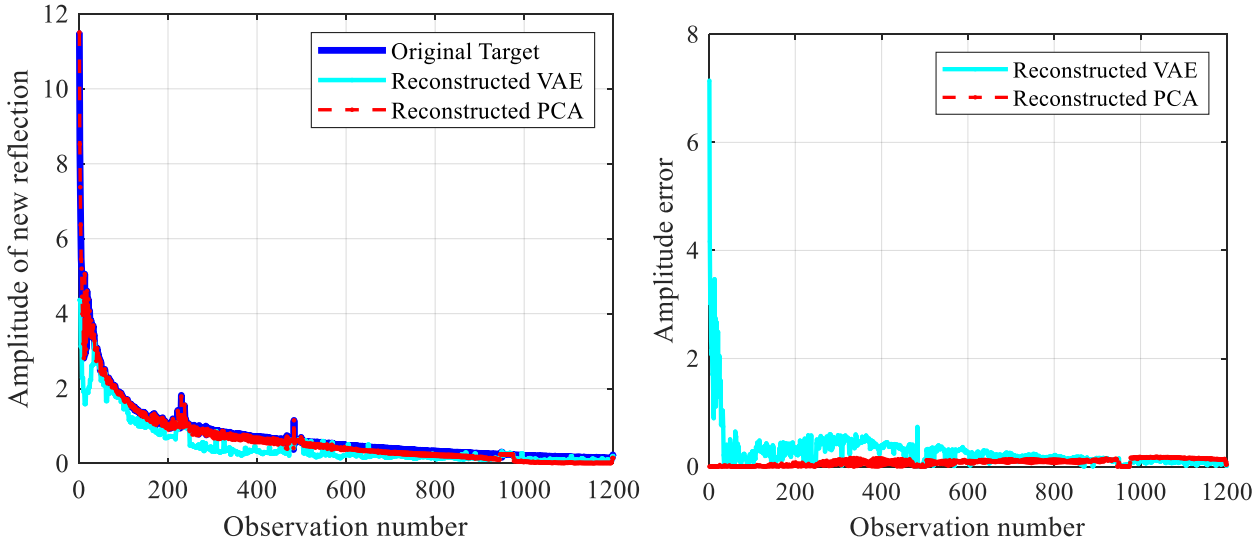

Figure 11: The predicted amplitudes for the fused reflection and amplitude error for n=1204 observations in the input data.

5.2 Generating new responses using the PCA and VAE models

5.2.1 Fusing a new reflection close to another reflection

A possible application of our PCA and trained VAE models is using them to generate new data. Figure 12 shows the results when new reflections were fused close to welds A and B, respectively, at 90m and 158m. The results are compared to the baseline signal (Figure 2) containing only the four original welds. It is evident that in both cases, the PCA and VAE were able to fuse reflections at those positions. Though the amplitudes are not well approximated, the two approaches were able to capture the dispersion nature of the reflection as well as multiple propagating modes that reflected from the new weld.

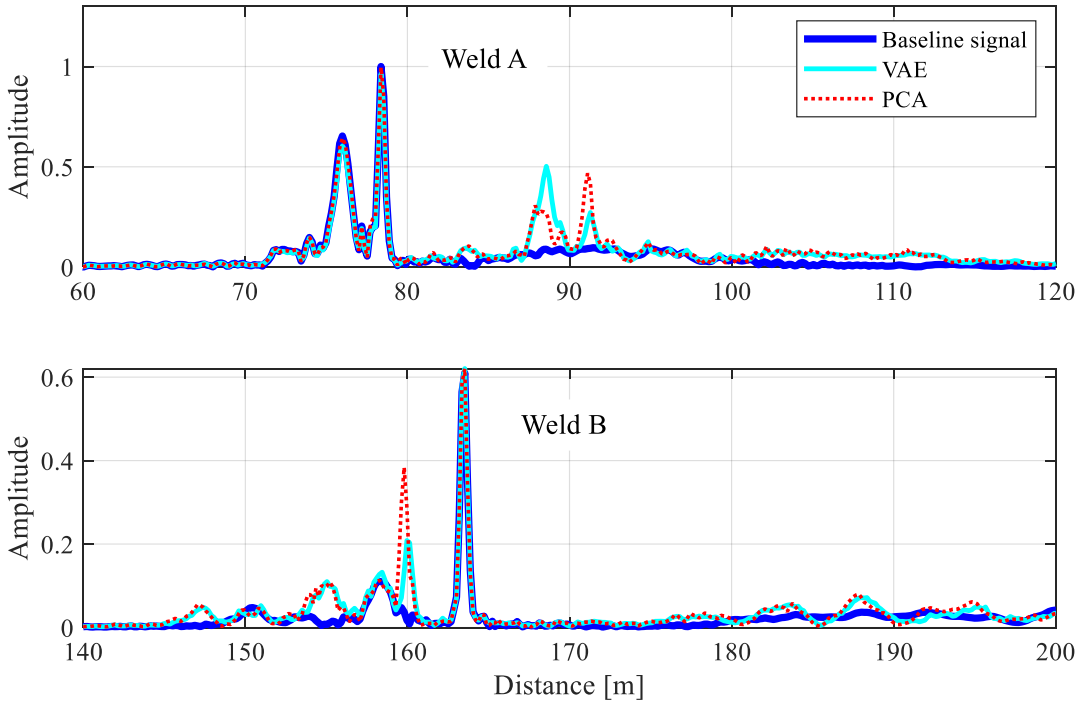

Figure 12: New reflections fused at 90m and 158m, respectively.

5.2.2 Overlapping reflections

In Figure 13, new reflections were fused on top of existing reflections from the four original welds. The results show that the VAE predicted larger amplitudes when new reflections were overlapped with original reflections from the four respective welds. For weld B, there was an increase in amplitude for the mode appearing before 160m though the amplitude of the mode with energy concentrated in the head decreased. On the other hand, though the PCA was able to fuse a reflection overlapping with weld A, it did not perform well for reflections overlapping with welds C and D. This bevahiour was also noticed from the results in Section 5.1.

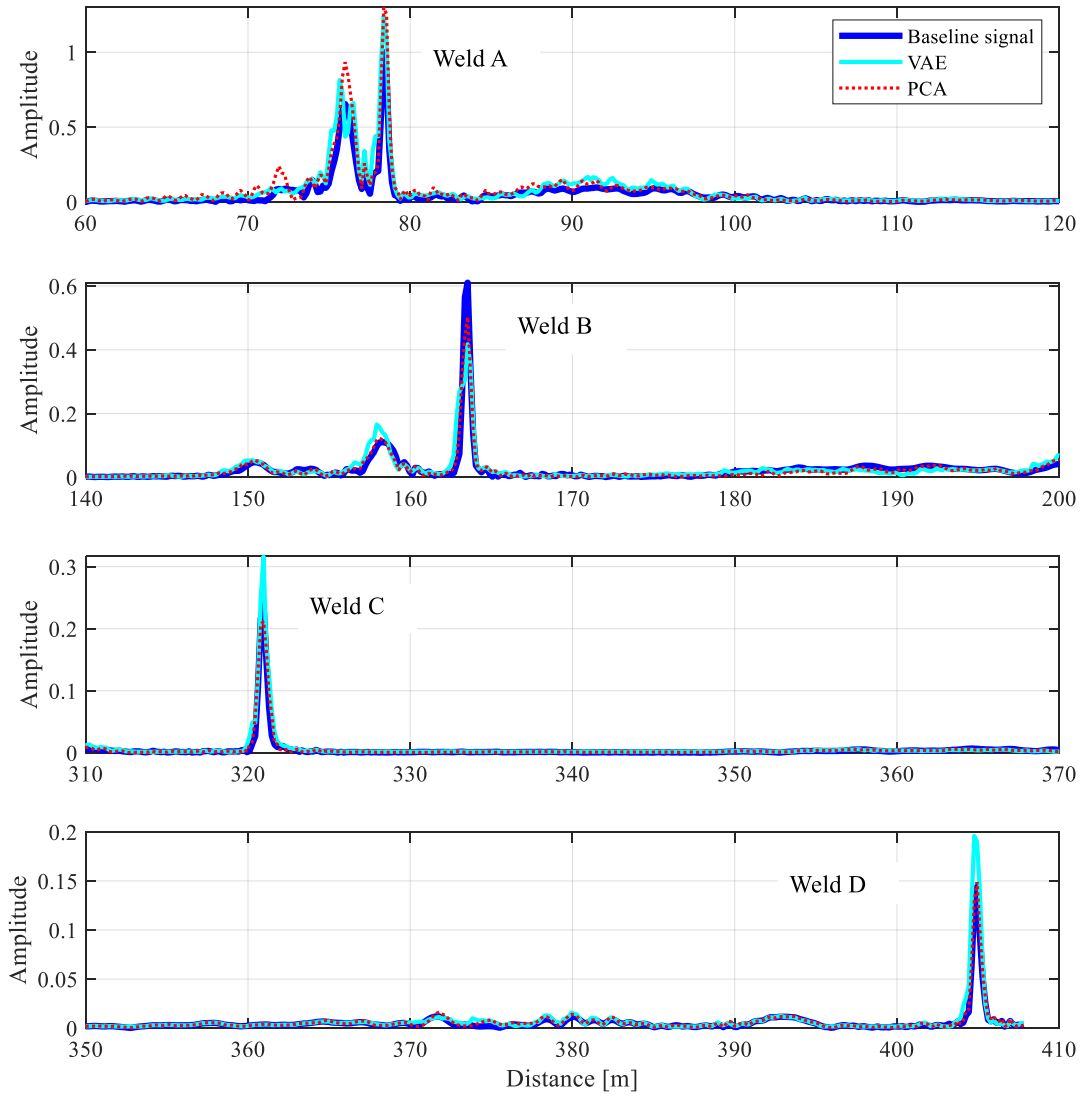

Figure 13: New reflections fused at distances overlapping with the original reflections.

6 Conclusions and future work

In this paper, we demonstrated the application of PCA and VAE to augment idealistic physics-based modelled responses with virtual experimentally measured reflections for welded railway lines. The inputs for the two investigated approaches were virtual experimental responses of a welded rail simulated with a physics-based finite element approach and random distance positions for the new welds. The generative models augment the virtual experimental response signal with additional simulated reflections from the extra welds. The results showed that both PCA and VAE have a good reconstruction ability. The two approaches were able to fuse new reflections at correct distances. The VAE was not able to capture the correct amplitudes. The VAE was able to fuse new reflections for cases when there was an overlap with an existing weld. PCA did not peform well for larger distances. In future, the two data-driven approaches will be used to model and simulate the inclusion of damage and coherent noise from EOCs and other complex features.

References

[1] P. Loveday, D. Ramatlo and F. Burger, "Monitoring of rail track using guided wave ultrasound," in *Proceedings of the 19th World Conference on Non Destructive Testing*, 2016, pp. 3341–3348.

[2] F. Burger and P. Loveday, "Ultrasonic broken rail detector and rail condition monitor technology," in *Proceedings of the 11th International Heavy Haul Association Conference*, 2017, pp. 275–280.

[3] C. Liu, J. Dobson and P. Cawley. "Efficient generation of receiver operating characteristics for the evaluation of damage detection in practical structural health monitoring applications," in *Proceedings Royal Society*, vol. 437, pp. 1–26, 2017.

[4] D. Ramatlo, C. Long, P. Loveday and N. Wilke, "A modelling framework for simulation of ultrasonic guided wave-based inspection of welded rail tracks," *Ultrasonics*, vol. 108, 2020.

[5] D. Ramatlo, C. Long, P. Loveday and N. Wilke., "Physics-based modelling and simulation of reverberating reflections in ultrasonic wave to welded rail tracks," *Journal of Sound and Vibration*, vol. 530, 2022.

[6] S. Mohammed, A. Khalid, S. Osman and R. Helali, "Usage of Principal Component Analysis (PCA) in AI Applications," *International Journal of Engineering Research & Technology*, vol. 5, no. 12, 2016.