# What Can Generative Modelling Do for Interpolation of Extremely Sparse Wind Farm Seismic Data?

*Tiexing Wang[1], Arash JafarGandomi[1], Rob Telling[1], Jing Sun[2]*

*1.Shearwater Geoservices 2. Delft University of Technology*

**Summary**

Traditionally, site surveys for wind farms are conducted using a grid of ultra-high resolution (UHR) 2D seismic lines, which provide data along specific profiles to support site investigations by mapping stratigraphic units across a survey area and aiding in the identification and localization of shallow hazards. There is also a need for re-siting wind farms and building subsurface models, which requires seismic data beyond the 2D UHR lines. However, these 2D UHR lines leave significant gaps between them, limiting the potential to make greater use of UHR data by filling these gaps through interpolation. This is a challenging task, as the gaps can range from hundreds of meters to several kilometers. Encouraged by the significant success of modern deep learning techniques in various seismic processing tasks, we investigate this specific challenge of extreme sparsity in UHR seismic data using generative modelling (specifically diffusion models) and classical supervised learning (based on U-Net and ResNet). The training and validation data are from UHR data acquired in the German North Sea planned for offshore wind development. The preliminary results show that when irregular missing data is up to 90%, the diffusion model outperforms U-Net and ResNet in terms of the peak signal-to-noise ratio (PSNR).

**Introduction**

Traditionally, ultra-high resolution (UHR) 2D seismic lines are used to support site investigations by mapping stratigraphic units across a survey area and aiding in the identification and localization of shallow hazards (Reveron, 2023). However, there is also a need for re-siting wind farms and building subsurface models. We need more information not only along the lines but also between the lines. The gaps between the 2D UHR lines are usually significant, ranging from a few hundred meters to kilometers. This sparsity, with up to 90% to 95% of traces missing irregularly, is becoming an obstacle to making greater use of UHR 2D seismic data. Consequently, the interpolation problem in wind farm surveys is significantly more challenging, requiring advanced techniques to fill in the large gaps in data and accurately model the subsurface.

In recent years, especially since 2017, deep learning methods have become widely used in geophysics, including for the interpolation problem. Many studies and results have shown that networks developed for computer vision, such as ResNet and U-Net, are also highly suitable for seismic data interpolation. Many research results have proven that these models can be effective in handling various types of missing data (Wang et al., 2019; Mandelli et al., 2019; Hlebnikov, 2022). Furthermore, in the past two years, deep generative modelling gained more attention in the seismic community and has been attempted on various processing tasks (Liu and Ma, 2024; Kaur et al., 2021; Durall et al., 2023). However, research on interpolating extremely sparse wind farm seismic data through deep learning still needs to be carried out.

In this work, we studied the specific challenge of interpolating extremely sparse seismic data in real-field UHR wind farm seismic surveys. The training dataset and test dataset are from the 2D UHR seismic data acquired in the German North Sea planned for offshore wind development. By manually decimating 90% of the traces in each image, we tested interpolation performance by using two deep learning strategies: deep generative modelling and classical deep supervised learning. Specifically, a diffusion model was implemented as the generative model and compared with a U-Net and a ResNet as benchmarks with PSNR analysis. Preliminary results suggest that with up to 90% irregular missing data, the diffusion model may have an advantage over U-Net and ResNet in terms of peak signal-to-noise ratio (PSNR).

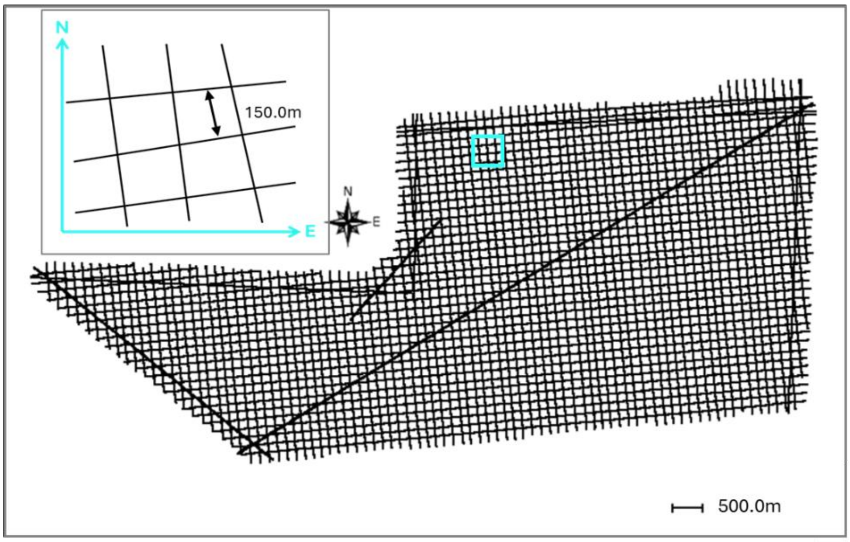


*Figure 1 survey map and problem description. The figure shows the survey map of the 2D UHR seismic lines. As the zoom-in diagram on the top left shows, the UHR data leaves a big gap between the lines.*

## Challenge Description

In this example, we utilized the data acquired in the German North Sea planned for offshore wind development. The survey map is given in Figure 1. 2D multichannel seismic was shot in a grid across a total area of 44 km², with lines running approximately east-west and north-south, leading to 140m-150m spacing between lines. The streamer was towed at a nominal depth of 0.7m, had 72 channels spaced at 1m. Suppose our desired interpolated grid has a 10m spacing. Given the UHR geometry and the distance between the 2D UHR lines with 1m channel spacing, the missing ratio is roughly 90%. We used this number to decimate our data and test the interpolation performance of deep generative modeling and deep supervised learning.

## Method

Deep generative modelling and classical deep supervised learning with convolutional neural networks (CNNs) were chosen for this study due to their complementary strengths in handling complex data patterns and their proven effectiveness in various applications. Deep generative models, such as diffusion models, which consist of a forward diffusion and a parametrized reverse process, excel in capturing the underlying data distribution and generating new, pseudo realistic data samples. This capability is potentially useful in scenarios where data is extremely sparse, as generative models can help to synthesize missing data and provide a more complete image.

On the other hand, classical deep supervised learning methods implemented with CNNs like U-Net or ResNet have shown remarkable performance in tasks requiring precise predictions based on labelled training pairs. These CNN models are adept at learning from examples and can effectively interpolate missing seismic data by leveraging their architectures designed for detailed feature extraction. Such CNN architectures, especially U-Net, are also commonly employed in diffusion models, as we chose for this work.

For an in-depth and detailed explanation of the above mentioned deep learning models, we refer the reader to Ho et al. (2020) for diffusion model, Ronneberger et al. (2015) for U-Net and He et al. (2016) for ResNet.

## Experiment

In this experiment, we compare the methods considering the optimal conditions for the stable diffusion model, U-Net, and ResNet. We prepare a set of images from the UHR data to train the networks. A total of 10,000 images (128x128) are selected from the dataset, and we manually set 90% of the traces to zero to mimic the sparse data scenario. After training, 200 images from the same dataset are used to test the performance of the networks. Note that the test images are not used during the training process, but they belong to the same dataset used for training, i.e., they belong to the same data distribution. After training, we calculate the peak signal-to-noise ratio (PSNR) and root mean square error (RMSE) of the 200 test images to evaluate the performance of the networks. The PSNR (in dB) is defined as

$$PSNR = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE_{I,K}}}\right)$$

where the $MAX_I$ is the maximum amplitude of the original image $I$ and $MSE_{I,K}$ is the mean squared error between the original image and inferenced image from the network.

## Results

Figure 2 illustrates the comparison of seismic data interpolation results obtained from three different deep learning models. Figure 2(a) shows the input seismic data with missing traces. Figure 2(b) shows the ground truth seismic data, while Figure 2(c), (d), and (e) display the interpolated results from the diffusion model, U-Net, and ResNet, respectively. Figure 2(f), (g), and (h) highlight the differences between the ground truth and the interpolated results from the diffusion model, U-Net, and ResNet, respectively. Upon visual inspection, the output from the diffusion model more closely resembles the ground truth seismic data compared to the outputs from U-Net and ResNet, particularly in terms of continuity and coherence of features. This outperformance is supported by the lower difference shown in Figure 2(f). This slight improvement indicates the diffusion model's greater potential to effectively handle high levels of data sparsity.

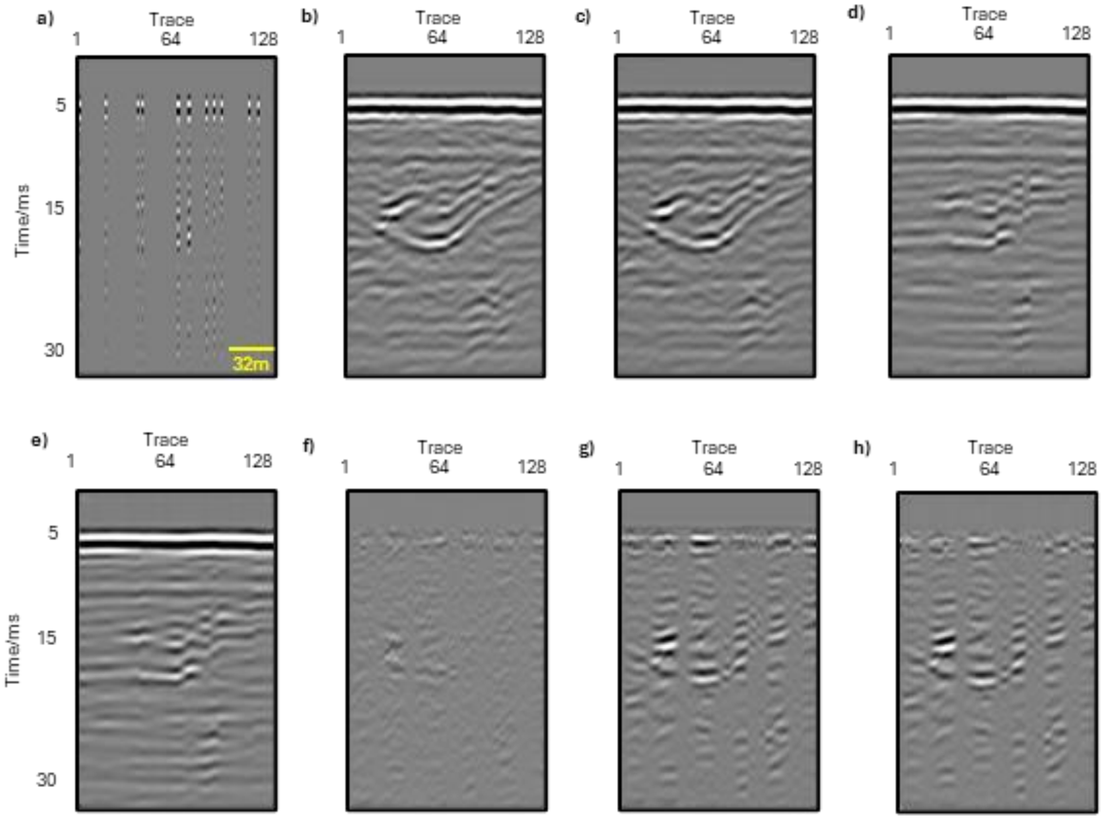


*Figure 2 a) Decimated seismic image, b) ground truth with 1m trace spacing, c-e) interpolated image via diffusion model, U-Net, ResNet, f-h) differences between the ground truth and interpolated image.*

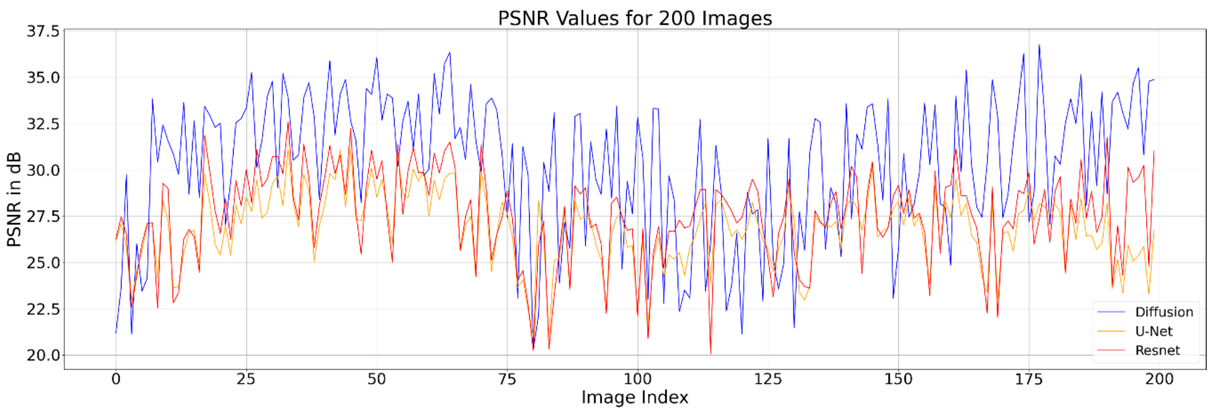


***Figure 3**. Comparison of PSNR for the test dataset using diffusion model, U-Net and ResNet.*

Figure 3 presents a comparison of PSNR values for a test dataset evaluated using three different deep learning models. The diffusion model is represented by the blue line, U-Net by the red line, and ResNet by the orange line. From the graph, it is evident that the diffusion model generally achieves higher PSNR values compared to the U-Net and ResNet, indicating better interpolation quality for most seismic images. The U-Net and ResNet exhibit similar performance, with their PSNR values frequently overlapping, although ResNet occasionally outperforms U-Net.

**Conclusions**

We have evaluated the performance of the diffusion model, U-Net, and ResNet in addressing the interpolation challenges associated with 2D UHR seismic data. The training and test datasets are derived from UHR 2D seismic data acquired from the German North Sea, intended for offshore wind development. By decimating 90% of the information in the dataset, we applied the three deep learning strategies to fill the gaps and analysed the results using difference and PSNR metrics.

Our findings indicate that the diffusion model outperforms U-Net and ResNet, suggesting it is a promising approach for future application. A possible reason is that with up to 90% to 95% of the data missing, the U-Net and ResNet models from supervised learning may not have enough reference points to accurately learn the underlying patterns, leading to increased errors and less reliable interpolations. However, though the diffusion model demonstrates promising performance in this work, the overall reliability of the new traces generated by generative models remains a subject of ongoing discussion and warrants further investigation.

**Acknowledgements**

We thank Shearwater GeoServices for permission to publish this work. We also thank Bundesamt für Seeschifffahrt und Hydrographie (BSH), the German Federal Maritime and Hydrographic Agency, for provision of data.

**References**

Durall, R., Ghanim, A., Fernandez, M. R., Ettrich, N., and Keuper, J. [2023]. Deep diffusion models for seismic processing. Computers and Geosciences, 177, 105377.

He, K., Zhang, X., Ren, S., and Sun, J. [2016]. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778.

Hlebnikov, V. [2022]. Deep learning as a tool for seismic data interpolation. PhD thesis, the University of Oslo.

Ho, J., Jain, A., and Abbeel, P. [2020]. Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.

Kaur, H., Pham, N., and Fomel, S. [2021]. Seismic data interpolation using deep learning with generative adversarial networks. Geophysical Prospecting, 69(2), 307-326.

Liu, Q., and Ma, J. [2024]. Generative interpolation via a diffusion probabilistic model. Geophysics, 89(1), V65-V85.

Mandelli, S., Lipari, V., Bestagini, P., and Tubaro, S. [2019]. Interpolation and denoising of seismic data using convolutional neural networks. arXiv preprint arXiv:1901.07927.

Reveron, J. [2023]. Correlation between geotechnical and geophysical data through seismic inversion, 84th EAGE Annual Conference and Exhibition.

Ronneberger, O., Fischer, P., and Brox, T. [2015]. U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, pp. 234-241. Springer International Publishing.

Wang, B., Zhang, N., Lu, W., and Wang, J. [2019]. Deep-learning-based seismic data interpolation: A preliminary result. Geophysics, 84(1), V11-V20