Seismic wave field prediction using encoder-decoder networks: from learning transfer functions to Virtual Seismic Arrays

Dissertation

with the aim of achieving a doctoral degree

at the Faculty of Mathematics, Informatics and Natural Sciences

Department of Earth System Sciences

at University of Hamburg

submitted by

Jana Klinge

Hamburg, 2025

Department of Earth System Sciences	
Date of Oral Defense:	09.07.2025
Reviewers:	Prof. Dr. Céline Hadziioannou
	Prof. Dr. Conny Hammer
Members of the examination commission:	Prof. Dr. Céline Hadziioannou
	Prof. Dr. Conny Hammer
	Prof. Dr. Christine Thomas
	Prof. Dr. Katharina-Sophie Isleif
	Prof. Dr. Bernd Leitl
Chair of the Subject Doctoral Committee	
Earth System Sciences:	Prof. Dr. Hermann Held
Dean of Faculty MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

The prediction of seismic wave fields between stations using machine learning offers great potential for geophysical monitoring, particularly in remote areas or in regions with sparse sensor coverage. This thesis introduces a novel encoder-decoder deep learning architecture that successfully learns the transfer function between seismic stations. By learning the complex signal transformations, this method enables accurate predictions of how seismic signals alter as they travel from one station to another. Notably, high quality predictions are achieved using only two days of data consisting solely of ambient seismic noise. The method's robustness in a range of scenarios is demonstrated via validation at a seismic exploration site with a variety of noise sources. The network shows particular strength in capturing phase-related features, which is crucial to its performance in seismic wave prediction. A systematic parameter study reveals important insight about the variables affecting model performance and points out areas for future development.

Virtual Seismic Arrays are introduced as a powerful proof of concept, extending the approach from individual station pairs to entire seismic arrays. By training the algorithm on all station pairs within an array, a set of predictive models is obtained that collectively form the Virtual Seismic Array. This enables the reconstruction of full-array recordings from a single reference station, even after physical sensors are no longer present. In the secondary microseism frequency band, beamforming analysis validates the effectiveness of Virtual Seismic Arrays by showing a high degree of agreement between the original and predicted waveforms.

This novel application of encoder-decoder networks for modelling transfer functions has the potential to enhance seismic monitoring, while reducing the need for continuous sensor coverage. By reconstructing signals at multiple stations from a single reference station, the approach enables ongoing array functionality in remote regions while reducing costs and maintaining array capabilities. These improvements are beneficial in industries like advanced seismic instrumentation and ultra-precision manufacturing where even small vibrations have significant impact on results. This is particularly beneficial in projects like the Einstein telescope, where the sensitivity of gravitational wave detections depends on reducing seismic disturbances.

Zusammenfassung

Die Vorhersage seismischer Wellenfelder zwischen seismischen Stationen mithilfe Künstlicher Intelligenz eröffnet neue Möglichkeiten für geophysikalische Messungen, insbesondere in abgelegenen oder sensorarmen Regionen. Diese Arbeit stellt eine neuartige Encoder-Decoder-Architektur vor, die erfolgreich die Übertragungsfunktion zwischen seismischen Stationen erlernt. Durch das Erlernen von Signaltransformationen ermöglicht die Methode präzise Vorhersagen, wie sich komplexe seismische Signale zwischen zwei Stationen verändern. Besonders hervorzuheben ist, dass Vorhersagen von hoher Qualität bereits mit nur zwei Tagen an Daten gelingen, die ausschließlich aus natürlichem Umgebungsrauschen bestehen. Die Robustheit der Methode wird an einem seismischen Explorationsstandort mit verschiedenen Quellen validiert. Das Netzwerk zeigt besondere Stärken bei der Erfassung phasenbezogener Merkmale – ein entscheidender Faktor für die Genauigkeit der Wellenvorhersage. Eine systematische Parameterstudie liefert zudem wichtige Erkenntnisse über Faktoren, die die Leistung des Ansatzes beeinflussen und zeigt Punkte für zukünftige Weiterentwicklungen auf.

Die Implementierung Virtueller seismischer Arrays erweitert die Methode von einzelnen Stationspaaren auf vollständige seismische Arrays. Durch das Training des Algorithmus auf allen Stationspaaren innerhalb eines physischen Arrays entstehen eine Reihe von Vorhersagemodellen, die zusammen das Virtuelle Seismische Array bilden. Damit lassen sich vollständige Array-Messungen allein auf Basis einer einzigen Referenzstation rekonstruieren - selbst wenn keine physischen Sensoren mehr vorhanden sind. Im Frequenzbereich der sekundären Mikroseismik zeigt eine Beamforming-Analyse eine hohe Übereinstimmung zwischen den vorhergesagten und originalen Wellenformen und validiert damit die Funktionalität Virtueller Seismischer Arrays.

Der Einsatz von Encoder-Decoder Netzwerken zum Erlernen von Übertragungsfunktionen kann seismischen Messungen erheblich verbessern und gleichzeitig den Bedarf an flächendeckender Sensorabdeckung verringern. Der Ansatz ermöglicht es, Signale an mehreren Stationen mithilfe einer einzigen Referenzstation zu rekonstruieren. Dadurch kann die Funktionalität des Arrays auch in abgelegenen Regionen kosteneffizient aufrechterhalten werden. Diese Verbesserungen sind besonders wertvoll in Bereichen wie der hochpräzisen seismischen Instrumentierung und

der Ultrapräzisionsfertigung, wo selbst kleinste Vibrationen signifikante Auswirkungen auf die Ergebnisse haben können. Ein besonderes Beispiel ist das Einstein-Teleskop, dessen Empfindlichkeit bei der Detektion von Gravitationswellen stark von der Reduktion seismischer Störungen abhängt.

List of Publications

Study I

Klinge J., Schippkus S., Walda J., Hadziioannou C., Gajewski D. (2025). Predictive modelling of seismic wave fields: learning the transfer function using encoder–decoder networks, *Geophysical Journal International*, 240(3), 1611-1621. https://doi.org/10.1093/gji/ggaf004
Published in *Geophysical Journal International*

Study II

Klinge J., Schippkus S., Hadziioannou C. (2025). Unlocking the potential of single stations to replace seismic arrays
Submitted to *Geophysical Research Letters*

Table of Contents

Abstract ..... **i**
Zusammenfassung ..... **iii**
List of Publications ..... **v**
Contents ..... **vii**
List of Figures ..... **ix**
1. Introduction **1**
1.1. Previous studies 3
1.2. Machine learning 8
1.3. Research objectives 12
1.4. Approach and outline 13
2. Study I: Encoder-decoder networks for seismic transfer **15**
3. Study II: Introduction of Virtual Seismic Arrays **29**
4. Optimizing network performance – A parameter study **49**
4.1. Parameter testing 49
4.2. Discussion and Conclusion 73
5. Summary and Conclusions **75**
References **81**
Acknowledgements **92**

List of Figures

1.1 Illustration of beamforming 4
1.2 Backazimuth vs. slowness plots from GRF array for four months in winter 1995/96 6
1.3 Conceptual illustration of a transfer function between two seismic stations 7
1.4 Schematic illustration of an encoder-decoder architecture 10
2.1 Simplified visualisation of the network architecture .... Study I / 19
2.2 Geometry of the experiment and synthetic setups .... Study I / 20
2.3 Results of the model analysis for two synthetic data scenarios .... Study I / 22
2.4 Results of the model analysis for three field data scenarios .... Study I / 24
3.1 Gräfenberg array beamforming .... Study II / 35
3.2 The Virtual Seismic Array .... Study II / 37
3.3 Beamforming results for the single dominant regime .... Study II / 39
3.4 Beamforming results for the changing regime .... Study II / 40
3.5 Beamforming results for the unseen regime .... Study II / 42
4.1 Map view of selected stations. 49
4.2 Relationship between model performance and interstation distance 52
4.3 Frequency-dependent model performance across interstation distance 53
4.4 Results for different scalings using RMSE and cross-correlation coefficient (CC) across the entire target time-series 58
4.5 Impact of network depth on the network performance 60/61
4.6 Impact of the number of training epochs on the model performance 64
4.7 Comparison of performance of different final activation functions 67
4.8 Results for data with varying numbers of periods at 100 Hz 69
4.9 Results for data with varying numbers of periods at 25 Hz 71

1. Introduction

Seismic signals offer an important perspective on the Earth's dynamic processes and give insights into different kinds of ground movements resulting from both natural and human activities (Chapman, 2004). The most widely recognized events are earthquakes and volcanic eruptions, but there is a far wider variety of seismic sources. Beside anthropogenic activities like traffic or energy production (Liu, 2017; Dias et al., 2020) and mass movements such as landslides and rockfalls (Jongmans and Garambois, 2007; Zimmer and Sitar, 2015), natural phenomena like winds or ocean waves (Walker and Hedlin, 2009; Hillers et al., 2012) also contribute to the overall seismic activity that is observed.

Each of these sources emits signals in various frequency bands that travel through the Earth as seismic waves (Sato et al., 2012). Seismic stations, which are equipped with seismometers, detect and measure these waves and allow seismologists to analyse specific wave attributes such as arrival times, polarization, amplitude, and frequency content that provide insights into their propagation mechanisms and characteristics (Bormann et al., 2012; Cheng et al., 2014). While individual stations already provide important data, seismic arrays further improve signal detection by incorporating a set of closely spaced seismometers. Seismic arrays can range from a handful of sensors to hundreds or thousands, which allows collecting seismic data over a broader area (Barker et al., 1996; Schweitzer et al., 2012). This increased coverage improves detection accuracy and spatial resolution and allows for a more precise analysis of seismic wave characteristics. Array processing techniques, such as beamforming, improve data interpretation by combining data from multiple sensors and help to determine the direction of arrival or apparent velocity of incoming waves (Rost and Thomas, 2002; Chen et al., 2002). The transfer function further describes how waves propagate through different media and interact with various geological features. It quantifies the relationship between signals at different locations and across seismic stations.

Another key aspect of seismic data analysis is the distinction between seismic signals and seismic noise (Bormann and Wielandt, 2013). Typically, seismic signals are discrete events, like earthquakes and explosions that often show distinct waveforms and clear patterns. In contrast, background vibrations, also referred to as seismic noise, originate from a variety of natural and human-made sources like traffic or ocean waves (known as microseism). Although seismic noise can complicate data interpretation, especially when trying to analyse seismic events or detect weak signals (Schimmel and Paulssen, 1997; Gaci, 2014), it is not always detrimental to seismic analysis. In fact, noise signals are widely employed in many different applications as well, including source localization (Shapiro et al., 2006; Chiariotti et al., 2019), imaging (Shapiro et al., 2005; Yang and Ritzwoller, 2008; Ritzwoller et al., 2011) and monitoring of subsurface structure changes (Snieder et al., 2002; Wegler and Sens-Schönfelder, 2007; Campillo et al., 2008).

Seismic observations are essential for advancing our understanding of the dynamic processes of the Earth. While individual seismic stations and seismic arrays offer powerful configurations for the efficient collection of high-quality data, their full potential is affected by various challenges. When seismic stations go offline or malfunction, it becomes problematic to sustain continuous measurements. This results in data gaps and affects the integrity of data analysis. This problem is especially notable in regions, where the deployment of stations is difficult, and where repairs or maintenance poses challenges. This can lead to long-lasting failures or shutdowns of stations, which affects the completeness of data and the interpretation of seismic activity in these regions. The spatial coverage of seismometers is further restricted by their uneven distributions, particularly in remote areas where it is difficult to set up stations. This leads to localized observation gaps that impact our understanding of seismic activity in those areas. Each factor adds to the difficulties of guaranteeing comprehensive data quality independently. In addition to challenges posed by the stations and their coverage, resource management has an impact on the efficiency of seismic networks. The limited availability of equipment, for example when stations are already in use in other regions, makes it difficult to allocate resources for new deployments. Additionally, both the equipment itself and the logistical difficulties, come with significant costs. Resolving these challenges would therefore greatly increase seismic data acquisition and lead to more reliable and accurate data.

These challenges can be addressed by gaining an understanding of wave propagation between seismic stations. Transfer functions mathematically describe the translation of seismic waves from one station to another (McCowan and Lacoss, 1978; Yan et al., 2003). Once estimated, transfer functions are capable of reconstructing the transformations that a seismic signal undergoes before it reaches another station, even if that station is malfunctioning or offline. While they are mathematically complex, they are not always explicitly studied. Instead, they have been implicitly analysed in a number of array processing studies through investigations of signal coherence and wave propagation across arrays (Rost and Thomas, 2009; Boué et al., 2016). The following subsection offers a comprehensive review of existing literature on seismic arrays, transfer functions, and their applications in wave propagation analysis.

1.1. Previous studies

Built to detect and identify nuclear explosions and discriminate them from earthquakes (Rost and Thomas, 2002), seismic arrays are since powerful tools for the detection, localization, and analysis of seismic events and seismic noise. While they were originally tested in various geometries like cross- or L-shape (Schweitzer et al., 2012, e.g. Keenan and Dyer, 1984; Koper et al., 2009), studies showed that the aperture of seismic arrays ideally relies on the properties of the waves under investigation (Rost and Thomas, 2002; Karamzadeh et al., 2018; Schweitzer, 2021). Independent of their individual geometry, seismic arrays are installed in locations all over the world and are used for a variety of applications. These include earthquake monitoring (Spudich and Oppenheimer, 1986; Meng et al., 2014; Meng and Ben-Zion, 2017) and nuclear test observations (Gibbons and Ringdal, 2006; Selby, 2010) to analysing volcanic activity (Saccorotti et al., 2001) and induced seismicity (Majer et al., 2007; McClellan et al., 2018).

Array processing

A number of signal-processing techniques have been developed to take advantage of the potential of seismic arrays in these applications. For instance, by transforming the measured data from the time into the frequency domain, frequency-wavenumber (f-k) analysis helps to determine the direction and speed of incoming waves (Capon, 1969; Capon, 1973). Beamforming is another fundamental technique that is based on comparable ideas of spatial filtering. With the delay-and-sum approach, which is a widely applied beamforming method (Rost and Thomas, 2002; Schweitzer et al., 2012; Perrot et al., 2021), array waveforms from a target direction are aligned and summed (Figure 1.1**a**), while waveforms from other directions overlap less effectively and cancel out (Rost and Thomas, 2002; Schweitzer et al., 2012). Like this, one can locate seismic sources by estimating the direction (backazimuth) and the apparent velocity (slowness) of the incoming wave (Figure 1.1**b**). Furthermore, beamforming improves the signal-to-noise ratio (SNR), aiding in the identification of weaker signals in the data (Rost and Thomas, 2002, Schweitzer et al., 2012). Another technique for improving signals in array processing is Wiener filtering. It effectively suppresses noise while preserving the signal of interest by minimizing the difference between the estimated and desired signal (Green et al., 1966; Wang et al., 2011). Along with others, these array-processing techniques enable the extraction of greater amounts of information from seismic data, thus improving the ability to detect, locate, and characterize seismic events.

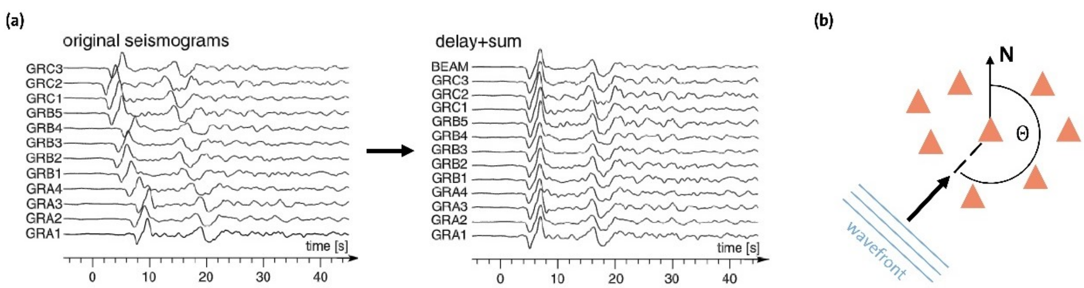


Figure 1.1 Illustration of beamforming (a) "Delay and sum" beamforming method for records of the Gräfenberg array (GRF). Left: Traces recorded by array stations. Right: Beamforming results. From Rost and Thomas (2002). (b) Horizontal angle of incidence (backazimuth $\boldsymbol{\theta}$) for wavefront from southwest (after Schweitzer et al., 2012).

Seismic arrays for event analysis

Several studies demonstrate the power of seismic arrays and corresponding processing techniques in their works. For instance, Meng and Ben-Zion (2017) use data from a dense seismic array and develop a multi-step approach that combines waveform stacking, envelope multiplication, and beamforming to detect more small earthquakes compared to standard catalogues. Similarly, Gibbons and Ringdal (2006) can detect smaller earthquakes than traditional methods by showing that array-based waveform correlation techniques are able to identify repeating patterns also for weak signals. In order to identify various seismic signals, Lythgoe et al. (2021) use a dense nodal array in the noisy urban environment of Singapore and apply an image processing approach. They can detect signals such as previously unreported earthquake, man-made events, and ground motion caused by thunder. Kiser and Ishii (2012) increase the lateral resolution of their analysis by combining arrays. This allows the authors to study the source properties of five major earthquakes and create detailed rupture images using a backprojection method. The power of seismic arrays to study rupture also extends to real-time applications in Earthquake Early Warning (EEW) systems. Meng et al. (2014) track the extent and directivity of ruptures during large earthquakes, which provides important information for the speed and accuracy of EEW. In addition to earthquake studies, seismic processing techniques can be used to study other complex environments. For example, Nanni et al. (2022) use high-resolution imaging techniques with dense seismic arrays to detect different features of an Alpine glacier, including diffracting materials and active crevasses. Expanding these applications from glacial environments to volcanic settings further highlights the applicability of seismic arrays in diverse geological systems. Low frequency signals associated with eruptions and gas-jet activity were captured by an array of nine broadband seismometers at Stromboli volcano, giving new perspectives on the internal dynamics of the volcano (Neuberg et al., 1994).

Seismic arrays for noise analysis

Beyond these event-based studies, seismic arrays are also widely used to analyse ambient seismic noise. Given that the ambient seismic wave field consists of signals from various origins, recent studies have focused on isolating individual sources to understand better their spatial distribution and spectral characteristics. Caused by ocean wave interactions with the solid Earth, one dominant component of ambient noise at periods between 5 and 20 seconds are primary and secondary microseisms (Webb, 1992; Hillers et al., 2012). A study of Friedrich et al. (1998) employed a frequency-wavenumber analysis on data of the Gräfenberg array (GRF) in southern Germany and identified multiple generating areas of microseisms in the Atlantic Ocean, Arctic Sea, and Mediterranean Sea. Figure 1.2 shows their findings for primary (Figure 1.2a) and secondary (Figure 1.2b) microseisms for a period of four-months in winter.

Beside ocean-generated microseisms, wind and human activities are further significant sources of seismic noise. A study by Hu et al. (2019) in eastern Portugal showed that seismic noise contains identifiable signatures from wind turbines (WT) and wind gusts. With these signatures, it is accurately possible to determine the WTs operational status, whereas wind gusts remain detectable even after removing WT effects from the data. Stammler and Ceranna (2016) also investigate WT influence on background noise using the GRF array, previously mentioned in microseism studies. They found significant disturbances at stations within 5 km of WTs and detectable signals up to 15 km away, which emphasize the impact of WT noise on seismic stations. Similarly, Riahi and Gerstoft (2015) study urban seismic noise produced by traffic with an array of 5200 geophones located in Long Beach, California. The array's density allowed a high-resolution analysis of noise sources, revealing signatures from metro trains, aircrafts taking off and landing, and night-time traffic along a highway. Machinery like oil pump jacks are beyond identified as significant sources of seismic noise by Schippkus et al. (2020) using data from an industrial-scale deployment of over 10.000 seismic stations. The pump jacks generate strong and characteristic signals, particularly in the 2-20 Hz range, which most likely have to do with the machinery and the pump jacks up-and-down motion.

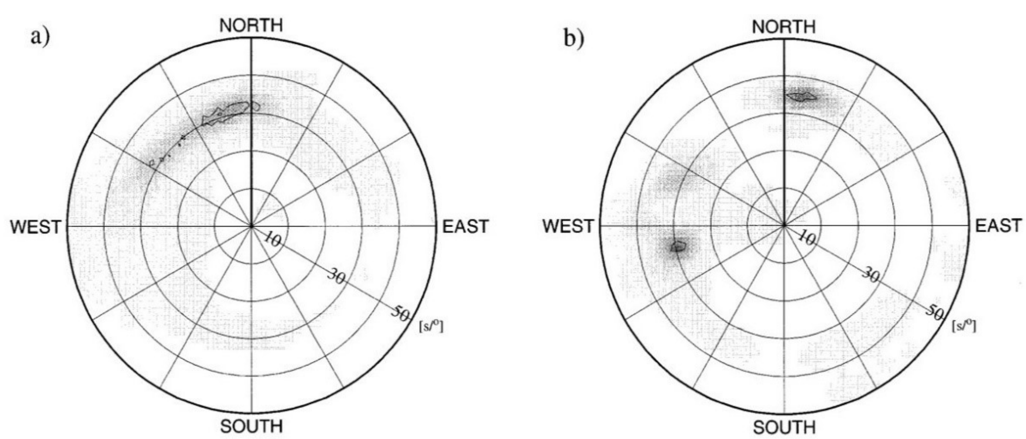


Figure 1.2 Backazimuth vs. slowness plots from GRF array for four months in winter 1995/96: a) primary microseisms: two dominant source areas at backazimuths of about 10° and 260°. b) secondary microseisms: broad azimuthal distribution from -70° to 10° (from Friedrich et al., 1998)

For seismic imaging, Ritzwoller et al. (2011) have tracked surface wavefronts from ambient noise recordings across more than 1000 stations from the EarthScope USArray to create detailed velocity maps of the crust and uppermost mantle. Similarly, Saygin and Kennett (2010) used ambient seismic noise tomography to map continental structures of the Australian crust. By cross-correlating data from over 2000 stations, they were able to estimate group velocity maps, revealing sedimentary basins and cratonic regions. While these studies use noise recordings from thousands of stations, Schippkus et al. (2018) create a high-resolution 3D shear-velocity model of the upper crust in the Vienna Basin region by using noise recordings from 63 array stations.

Transfer functions

Seismic noise, once regarded as a disturbing background sound, has evolved into an important data source for seismological studies. This advancement has significantly aided by seismic arrays, which have enhanced the understanding of ambient seismic wave fields across multiple stations. Seismic data, for instance from earthquakes and ambient noise, can be used to estimate transfer functions that represent how seismic waves travel between stations. Figure 1.3 shows a conceptual illustration of a transfer function between two seismic stations.

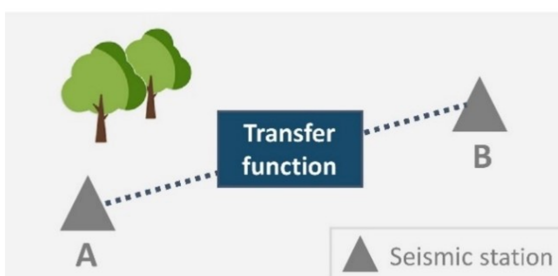


Figure 1.3 Conceptual illustration of a transfer function between two seismic stations (grey triangles). The dashed line represents the path along which seismic waves propagate.

For instance, Lim and Ahn (2023) estimate transfer functions from ambient noise data by computing surface recordings from borehole seismometer data. This method proves particularly beneficial in regions where it is not feasible to install seismometers at the surface. Working with earthquakes, Rajabi and Rajabi (2015) determine the transfer function for seismic waves traveling between a location near the earthquake and another area where vibration is anticipated. This approach is used in the context of Earthquake Early-Warning systems to perform real-time prediction of earthquakes at distant locations. With a known transfer function, it is further possible to calculate both amplitude and phase responses at any frequency, where the amplitude response describes changes in the strength of different frequency components, and the phase response gives information on their timing (McCowan and Lacoss, 1978). Transfer functions are correlated with, yet distinct from, other fundamental concepts in seismology. The instrument response, for example, describes the connection between ground motion and the recorded seismogram (Walden and White, 1998; Havskov and Alguacil, 2016). In contrast, the Green's function defines how a delta source at a given location affects the response at another location (Snieder, 2004; Sabra et al., 2005; Denolle et al., 2013).

1.2. Machine learning

Machine learning (ML), a subfield of artificial intelligence, uses algorithms and statistical models to let computers learn from data, which enhances our ability to handle complex tasks and effectively process big datasets. The field of machine learning has evolved significantly since its beginnings, with rapid growth in the 21st century through increased computing capabilities, the availability of big data, and algorithmic advancements. Machine learning is widely applied in a number of industries, such as healthcare (Ahmad et al., 2018; Shailaja et al., 2018), finance (Dixon et al., 2020; Ahmed et al., 2022), environmental science (Zhong et al., 2021), and manufacturing (Wuest and Thoben, 2016; Morariu et al., 2020). It offers a wide range of applications, including predictive analytics and image recognition (such as in self-driving cars and facial recognition: Fujiyoshi et al., 2019; Guo et al., 2024), natural language processing (used in virtual assistants: Imrie and Bednar, 2013; Duguleană et al., 2020), and automated decision-making (as seen in fraud detection: Wihlborg et al., 2016; Araujo et al., 2020). These applications employ various ML methods, including decision trees and neural networks, to process data, and facilitate informed decision-making. ML has become an essential resource in scientific areas such as seismology due to its ability to draw conclusions from large datasets and address complicated problems.

Machine learning in seismology

Machine learning opens up new opportunities and provides powerful approaches for seismic data analysis that can improve existing methods in seismology. Conventional array processing techniques rely on established physical approaches, whereas ML algorithms can learn relationships directly from the data. This allows the identification of patterns that may be challenging to recognize and understand using traditional approaches. A detailed overview of progress and challenges regarding the use of ML in earthquake seismology is provided by Mousavi and Beroza (2023), together with recommendations for further investigation. Mousavi and Beroza (2023) emphasize that developments in concepts, algorithms, and computation have greatly improved earthquake monitoring, forecasting techniques, and the compilation of more comprehensive catalogs. Kong et al. (2018) and Kubo et al. (2024) offer further reviews of ML in seismology.

Neural networks are a type of machine learning models that mimic the function of the human brain. Consisting of layers of connected units known as "neurons", each layer processes information and passes it to the next, gradually building upon the previous layer (Gurney, 2018). Deep learning, a subfield of ML, uses these neural networks with multiple layers to analyse complex data (Mousavi and Beroza, 2022). The training of these networks is based on large datasets that include information relevant to the particular problem the network is designed to solve. Thereby, this data is effectively processed by the network's architecture through specific parameter choices or its depth, allowing it to learn complex patterns and extract information from the data. Like this, neural networks are able to handle diverse seismological problems, from signal detection and phase picking (Zhu and Beroza, 2019; Zhu et al., 2019; Pardo et al., 2019) to tasks like waveform prediction and subsurface characterization (An et al., 2001; W. Zhu et al., 2019; Jozinović et al., 2020).

Encoder-decoder networks

The encoder-decoder network is one type of neural network architecture. These networks work by extracting features from input data and turning them into relevant outputs, which makes them particularly helpful for tasks where the lengths of the input and output sequences can vary. Encoder-decoder networks consist of two main elements: an encoder that processes the input data to extract essential information and compress them into a meaningful representation, and a decoder, which takes this condensed representation and converts it into the intended output sequence. A schematic illustration of this encoder-decoder architecture and its components is shown in Figure 1.4. In addition to their widespread use in natural language processing (Badola and Gupta, 2021; Ao et al., 2022), computer vision (Badrinarayanan et al., 2017; Chen et al., 2018), and speech recognition (Chiu et al., 2018; Toshniwal et al., 2018), encoder-decoder networks have also proven to be powerful tools in seismology.

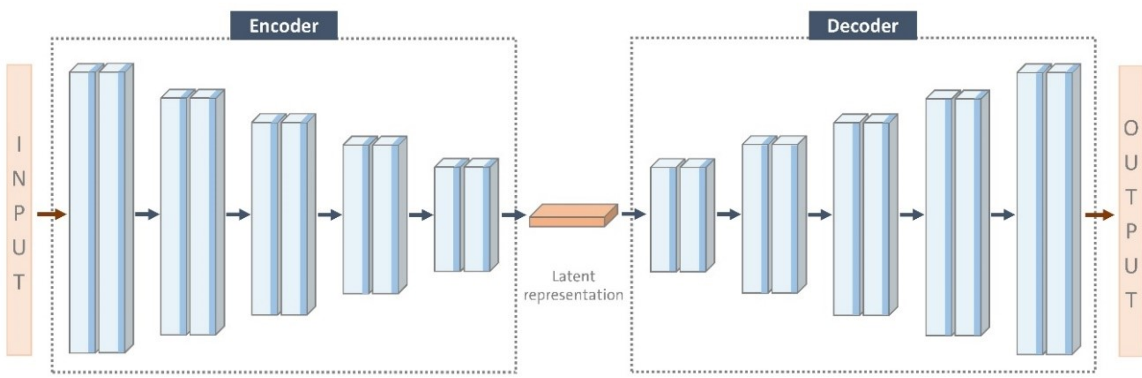


Figure 1.4 Schematic illustration of an encoder-decoder architecture: the input passes through the encoder, consisting of multiple layers that extract and compress features into a latent representation. The decoder reconstructs the output from the latent representation.

To separate earthquake signals from ambient seismic noise, Yin et al. (2022) train a multitask encoder-decoder network. The network takes noisy 3-component seismograms as input and learns to separate earthquake signals from ambient noise in the data. This allows to improve signal-to-noise ratios and to better use ambient noise signals for monitoring Earth's structure. Likewise, Zhang et al. (2020) use encoder-decoder networks for signal separation and develop a fully convolutional encoder-decoder network for separating microseismic signals from various types of noise. Their method learns features in the time-frequency domain to denoise microseismic signals, which improves the signal-to-noise ratio as well. Building on denoising, Saad and Chen (2020) use a deep-denoising autoencoder to attenuate random noise in seismic data. Their algorithm encodes noisy seismic data through several levels of abstraction to extract significant features, before decoding these features to reconstruct "the seismic signal without noise". From using neural networks to isolate specific signals within seismic data, studies have also explored their potential for forecasting future seismic events. Moustra et al. (2011) try to predict earthquakes in Greece using time series magnitude data. Their model uses historical earthquake magnitude data as input to predict the magnitude of seismic events for the following day as output. However, with a success rate of only 58.02%, the model's accuracy drastically declines for major seismic events. This example highlights the challenges and potential of using neural networks for time series forecasting, which can be applied to many different prediction tasks.

1.3. Research objectives

This work aims to address the limitations of traditional seismic arrays and improve seismic data collection and analysis through an advanced machine learning approach. The primary objectives are to improve the continuity and reliability of seismic measurements, mitigate the impact of array failures and offline periods, and increase spatial coverage of seismic observations by addressing gaps in data collection. Furthermore, this research aims to optimize resource management in seismic monitoring by exploring a cost-effective alternative to traditional array deployments. To achieve this, this work investigates the application of encoder-decoder networks to learn transfer functions between seismic stations. Approximating the transfer functions and thus how seismic waves translate between seismic station pairs, this approach has the potential to virtually replace physical stations by data prediction. Beyond reducing the need for physical station deployments, this also expands the potential uses of encoder-decoder networks in seismology. Building on this potential, this work addresses three main questions:

Q1 Can machine learning techniques be adapted to learn the transfer function between seismic stations for predicting seismic wave fields?

Q2 Can we use a single seismic station to predict the data of an entire seismic array?

Q3 What impact do different parameters have on the performance of machine learning techniques for predicting seismic wave fields?

1.4. Approach and outline

This dissertation presents three integrated studies that investigate the application of machine learning techniques, specifically encoder-decoder networks, to address challenges in seismic data collection and analysis. The first two studies introduce a novel machine learning approach and demonstrate its practical application, building upon each other. The third chapter complements the findings by providing a detailed analysis of parameter influences on the machine learning models, enabling further optimization of the approach.

Study I introduces a novel adaption of encoder-decoder networks for seismological applications, approximating the transfer function between seismic stations using 1-D time-series measurements. This approach uses data from one seismic station as input to predict data at another seismic station as the target, thereby incorporating both phase and amplitude information. While using differing input and output data in encoder-decoder networks is common, especially in sequence-to-sequence tasks, this application is new in the context of seismology.

Study II builds directly on the findings of **Study I**, applying the developed approach and test it to create Virtual Seismic Arrays. This novel idea utilizes data from one reference station within an array to predict recordings that would be measured by the entire array. Unlike traditional seismic arrays, which consist of multiple physical sensors, this would enable the acquisition of seismic data from previously instrumented areas even after the physical sensors have been removed. The performance of these Virtual Seismic Arrays is evaluated through beamforming.

Chapter III presents a parameter study on the encoder-decoder network. This examines how various parameters affect the performance of the machine learning approach in predicting seismic wave fields. Through the analysis of how specific parameters affect the results, this study provides an outlook for optimizing the encoder-decoder networks performance in future studies.

2. Study I: Encoder-decoder networks for seismic transfer

Published in *Geophysical Journal International*

Klinge J., Schippkus S., Walda J., Hadziioannou C., Gajewski D. (2025). **Predictive modelling of seismic wave fields: learning the transfer function using encoder–decoder networks**, *Geophysical Journal International*, 240(3), 1611-1621. https://doi.org/10.1093/gji/ggaf004

Author contributions:

- **JK**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing original draft, Writing review and editing.
- **SS**: Conceptualization, Data curation, Supervision, Writing review and editing.
- **JW**: Conceptualization, Software.
- **CH**: Conceptualization, Funding acquisition, Supervision, Writing review and editing.
- **DG**: Conceptualization, Funding acquisition, Supervision, Writing review and editing.

All authors have read and agreed to the published version of the manuscript

Geophysical Journal International

Geophys. J. Int. (2025) **240,** 1611–1621 https://doi.org/10.1093/gji/ggaf004
Advance Access publication 2025 January 06
GJI Seismology

Predictive modelling of seismic wave fields: learning the transfer function using encoder–decoder networks

Jana Klinge, Sven Schippkus, Jan Walda,* Céline Hadziioannou and Dirk Gajewski

Institute of Geophysics, Centre for Earth System Research and Sustainability (CEN), Universität Hamburg, 20146 *Hamburg, Germany. E-mail: jana.klinge@uni-hamburg.de*

Accepted 2024 December 24. Received 2024 December 2; in original form 2024 July 17

SUMMARY

Wouldn't it be beneficial if we could predict the time-series at a seismic station even if the station no longer exists? In geophysical data analysis, this capability would enhance our ability to study and monitor seismic events and seismic noise, particularly in regions with incomplete station coverage or where stations are temporarily offline. This study introduces a novel adaption of encoder–decoder networks from the subfield of deep learning, modified to predict the development of seismic wave fields between two seismic stations. Using 1-D time-series measurements, our algorithm aims to learn and predict signal transformations between the two stations by approximating the transfer function. Initially, we evaluate this proof of concept in a simplified controlled setting using synthetic data, before we incorporate field data gathered at a seismic exploration site in an area containing several roads, wind turbines, oil pump jacks and railway traffic. Across diverse scenarios, the model demonstrates proficiency in learning the transfer function among various seismic station configurations. Particularly, it achieves high accuracy in predicting a majority of seismic wave phases across different data sets. Diverging significantly from encoder–decoder networks that estimate time-series forecasts by analysing historical trends, our approach places greater emphasis on the wave propagation between nearby locations. Thereby, the analysis incorporates both phase and amplitude information and provides a new approach to approximate the transfer function relying on machine learning techniques. The gained knowledge enables to reconstruct data from missing, offline or defunct stations in the context of temporary seismic arrays or exclude non-relevant data for denoising.

Key words: Machine learning; Time-series analysis; Seismic interferometry; Seismic noise; Wave propagation.

1 INTRODUCTION

Signal recording and processing hold significant importance across a range of scientific disciplines, including the field of geophysics. Capturing and analysing various types of signals, such as seismic waves, electromagnetic waves and gravity anomalies, enables the understanding of the Earth's subsurface and its geological characteristics. As waves propagate through the Earth, their interaction with geological structures, such as sediment layers or fault lines, affects the recorded signals and leads to changes in the wave's propagation characteristics. Deploying seismic stations enables the measurement of signals and the derivation of insights regarding the subsurface characteristics and nature of the area.

In seismic analysis, understanding these measurements involves the identification of different wave types, along with analysing frequency spectra, amplitude variations, phase shifts and other wave properties (M. Bath 1973; Rost & Thomas 2002; Barnes 2007). While many of these signal components deliver valuable information and are essential for seismic investigations, there are also parts known as seismic noise that introduce more complexity to the data interpretation process. Natural sources such as wind or ocean waves, atmospheric disturbances or geological activities, as well as artificial sources including human activities and industrial operations, emit noise signals in various frequency bands and contribute to seismic measurements. In order to interpret measurements and mitigate the influence of undesired signals on the results, it is important to understand how seismic waves interact with geological structures (Kawakami & Oyunchimeg 2003). The transfer function captures this relationship by describing how initial seismic signals change as they travel through the medium, leading to the signals measured

*Present address: Emetriq GmbH, 20459 Hamburg, Germany.

by seismic sensors. For instance, the relation between the ground motion and the recorded seismogram is named instrument response (Walden & White 1998; Havskov & Alguacil 2016; Lindsey *et al.* 2020), while the Green's function describes the response at a given location to a delta source applied at another location (Snieder 2004; Sabra *et al.* 2005; Sergeant *et al.* 2020). However, estimating the transfer function in seismology can be complex due to the interaction of varying subsurface structures, variability in seismic wave propagation, noise, instrumentation limitations and source field dependence. Together, these factors contribute to intricate coupled systems of differential equations. In this context, the Green's function represents the theoretical impulse response when the system's differential operator is known. If the operator is not available, we can estimate the transfer function under the assumptions of linearity and time invariance.

Machine learning has emerged as a widespread methodology in geophysical data analysis, providing an advanced alternative to conventional seismic analysis methods for uncovering relationships within seismic data. Multiple fields including seismic exploration (Helmy *et al.* 2010; Li *et al.* 2019; Tariq *et al.* 2021) and seismology (Li *et al.* 2018; Xie *et al.* 2020; Mousavi & Beroza 2023) employ machine learning methods to characterize seismic data and detect and classify relevant characteristics and patterns within the data. One fundamental architecture in the subfield of deep learning (LeCun *et al.* 2015) are encoder–decoder networks, which provide the opportunity to learn and extract dependencies between data across input and output domains. In seismic and seismological applications, encoder–decoder networks play a crucial role for tasks like denoising (M. Saad & Chen 2020; Knispel *et al.* 2022; Yin *et al.* 2022) or interpretation (Wu *et al.* 2019; Zhang *et al.* 2021).

In this paper, we introduce an adaptation of encoder–decoder networks to learn the relationship between seismic wave fields recorded at two different locations. By using 1-D time-series from a fixed seismic station as input and the measurements from a nearby seismic station as target, we aim for the network to learn the alterations that the signal undergoes between the two stations. This involves two key concepts; first, the station-to-station transfer function, which describes how arbitrary seismic signals transform between the two stations due to geological and environmental factors, and secondly, the frequency-dependent Green's function, which is the medium's response to a delta source. We aim to demonstrate that a modification of the encoder–decoder architecture is capable of learning data characteristics that closely resemble the principle of the station-to-station transfer function within the setup of two seismic stations. While the foundation of the concept originates from the established practice of detecting and learning patterns and structures of and within time-series data (Malhotra *et al.* 2016; Badrinarayanan *et al.* 2017; Du *et al.* 2020; Beveren *et al.* 2023), our approach focuses more on the parts that influence the propagation of waves between nearby locations. Thereby, our analysis incorporates both phase and amplitude information, while also extending beyond the Green's function to capture additional aspects such as the source time function and radiation pattern. This provides a new approach to approximate the principles outlined by Curtis *et al.* (2012) without requiring a time-reversal mirror, instead using a single sensor and deep learning techniques.

We will guide through this study by introducing the encoder–decoder network setup and the most important metrics for this specific use case (Section 2) first. Following this, Section 3 outlines the characteristics of the measurement region and provides an overview of the selected seismic stations and data. Section 4 will involve evaluating the findings across different scenarios and data sets before discussing (Section 5) and drawing conclusions on the potentials and limitations of the presented method (Section 6).

2 NEURAL NETWORK SETUP

The methodology employed in this study follows the overall aim of testing the feasibility of a network that is able to learn the transfer properties between two seismic stations. We make use of an encoder–decoder architecture in a supervised fashion and train it by using input data from a fixed reference station A and target data from a second station B (Fig. 1, top). The form of the network traces the traditional U-Net shape (Ronneberger *et al.* 2015; Zhu & Beroza 2019; Li *et al.* 2022; Zhong *et al.* 2022) while an equal amount of convolutional and deconvolutional blocks defines its structure. Each block consists of a convolutional layer, a batch normalization layer and an activation layer. Furthermore, we use a dropout layer after each block to prevent overfitting by randomly setting a fraction of input units to zero during training. To make sure that every input connects to every output, we extend the architecture by a dense layer in the latent space bottleneck. To enable the direct transfer of information from the encoder to the decoder, we introduce skip connections between the respective convolutional and deconvolutional blocks. The depth of the network is five, while we use hyperbolic tangens as final activation layer in each of the individual use cases introduced in Section 3. As an outcome of the learning process from the input to the target data, the network delivers a prediction that ideally resembles the shape of the target data. Fig. 1 illustrates the schematic network architecture subdivided into the use of input and target data, the encoder part, the latent space and the decoder part.

To assess the model performance, we select different metrics to evaluate the similarity between the predicted ($\hat{y}$) and the observed value (y). This includes the mean squared error (MSE), which measures the average squared difference between predicted and observed values, and the mean absolute error (MAE), which quantifies the average absolute difference. In order to optimize the model during the training process of the algorithm, the error between the model prediction and the actual target data is estimated using the Huber loss function implemented by Keras (Chollet & others 2015). The Huber loss l (eq. 1) combines MSE and MAE with parameter ∂ defining the threshold for the transition from quadratic to linear components of the loss. This helps the Huber loss function to be robust to outliers in the data.

$$l\left(y,\hat{y}\right)=\begin{cases}\frac{1}{2}(y-\hat{y})^2 \text{ for } |y-\hat{y}|\leq\ \partial\\ \partial\left(|y-\hat{y}|-\frac{1}{2}\partial\right) \text{ for } |y-\hat{y}|\ >\partial\end{cases}. \tag{1}$$

To review the model's training progress after completion, we analyse the shape of the Huber loss curve. In addition, we visually inspect the MSE and MAE loss curves as supplementary Keras metrics (Chollet & others 2015), noting that they are not integrated into the optimization during training. To assess the performance of the model and the goodness of its prediction after training, we select two metrics to independently assess both amplitude and phase fit, and subsequently consider them in equal measure for an overall indication of the model quality.

In order to assess the degree of similarity between the target and prediction time-series, we use the normalized cross-correlation function as an evaluation metric. When calculated without applying a time-shift, this is equivalent to the Pearson correlation coefficient.

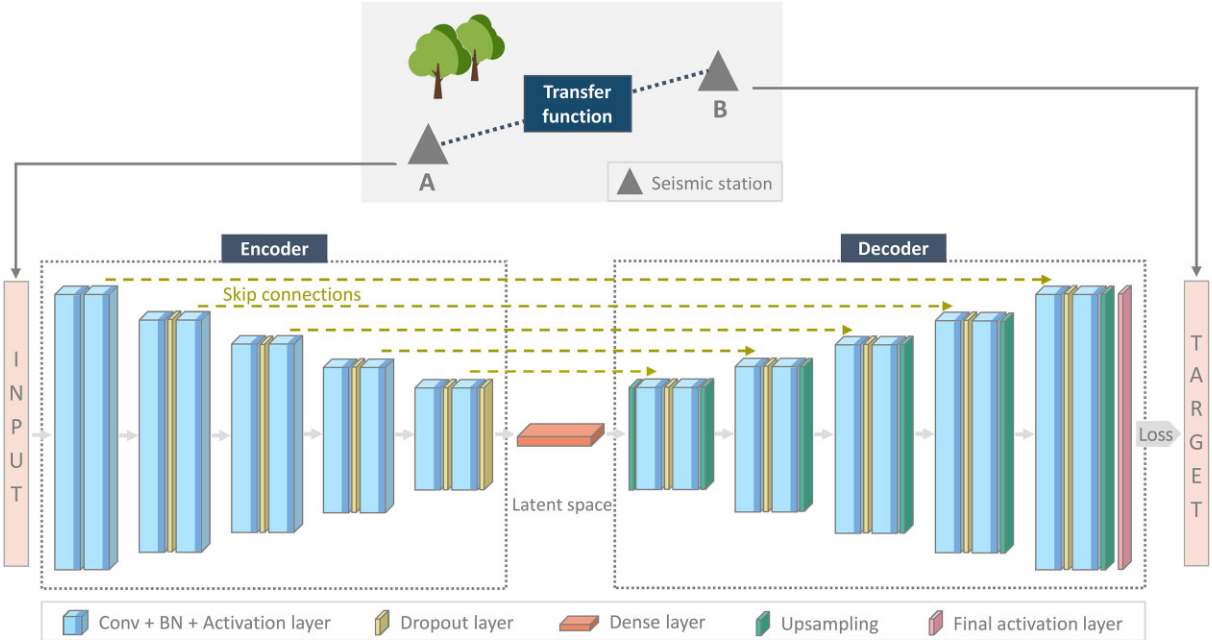


Figure 1. Simplified visualization of the network architecture consisting of an encoder and decoder part. Data from seismic station A serve as input, while data from another seismic station B provide the target data. Skip connections (dashed lines) link corresponding convolutional and deconvolutional blocks. Within the encoder, each block consists of a Convolutional layer (Conv), Batch Normalization (BN) and an Activation layer. A dropout layer follows almost every block. Within the decoder, each block with dropout layer complements by an Upsampling layer.

The similarity between the two time-series is assessed by calculating the correlation from their displacements, followed by normalizing the results using the overall standard deviation of the signals. While phase shift in seismology denotes the time displacement of a waveform, we employ this metric to emphasize the temporal alignment between the two signals. Assuming a good model and thus an accurate prediction, we expect both signals to be identical and align well without any offset. Under this assumption, we compute the cross-correlation without shifting samples at zero time and determine the cross-correlation coefficient at that point. By doing so, a value of 1 indicates a strong positive similarity, −1 indicates an anticorrelation and 0 reflects no relationship between the two time-series. Assessing the cross-correlation on the entire time-series as well as in smaller segments of about 10.24 s helps in determining the quality of the results in detail.

Classifying the amplitude differences between the predicted and the actual target values, the root mean squared error (RMSE) quantifies the accuracy of a model while being sensitive to the magnitude of errors. Thereby, the RMSE indicates how far the predicted value deviates from the target value, providing an indication of prediction accuracy. By employing RMSE as the second evaluation metric aids in comparing the amplitudes of the actual target data with those predicted by the model. RMSE defines as shown in eq. (2), where n represents the total number of data points and i refers to the i-th observation. With increasing errors, the RMSE score tends to rise linearly, indicating that a smaller value corresponds to a closer alignment between the model's predictions and the actual data. Thereby, RMSE shares units with the actual target values.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}. \tag{2}$$

While the Huber loss is estimated as part of the model training process to enhance the models understanding of the data iteratively, cross-correlation coefficients and RMSEs are calculated post-training. Estimating both provides a comprehensive approach to quantify the model's predictive capability of how well it captures the phases and amplitudes of the target data. Thereby, we do not combine the metrics numerically, but rather use them to comprehend the quality of the results.

3 DATA AND PROCESSING

In order to demonstrate the viability of the proposed method in capturing the relationship between two seismic stations, we will employ 1-D time-series measurements, starting with the exploration of synthetic data. With this, we aim to validate the viability of the general approach in a controlled setting, before we proceed to analyse field data gathered during a seismic exploration campaign.

3.1. Synthetic data

To generate synthetic data, we simulate two seismic stations with a constant interstation distance of 200 m located on top of a homogeneous, acoustic half-space with a medium velocity of 1500 $\frac{m}{s}$. This implies that, with a sampling rate of 100 Hz, the corresponding wavelength is approximately 150 m, placing the second station well within one wavelength. Initially, we create an ideal scenario with a single source, consistently arriving from the same direction, located near the stations but changing its position for each example (Fig. 2(c), S1). The resulting time-series resembles a cross-correlation function with a peak value of one at given traveltime and zero elsewhere, incorporating time-shifts corresponding to the traveltime between the stations. By generating morlet wavelets with

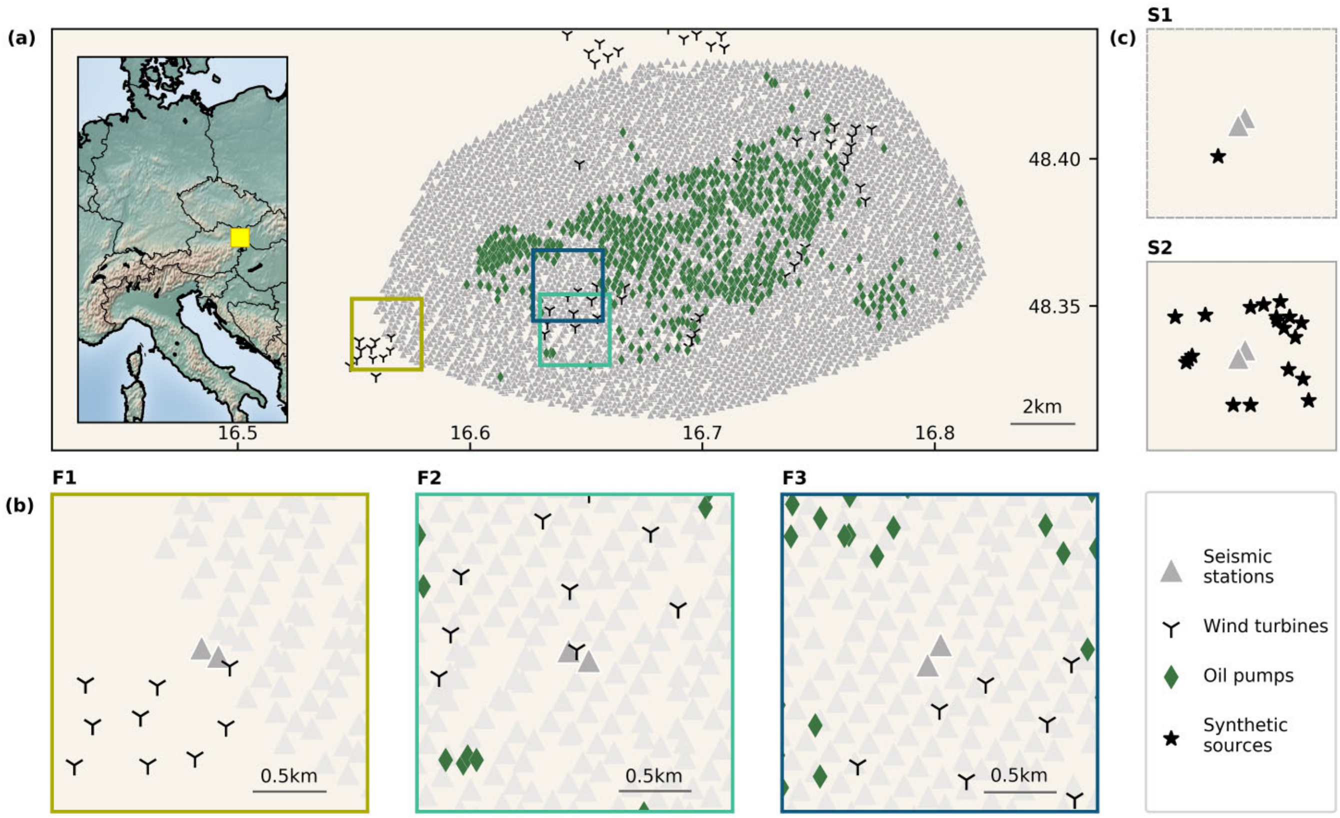


Figure 2. Geometry of the experiment and synthetic setups. (a) Map of the study area northeast of Vienna, where seismic stations are positioned with an interstation distance of approximately 200 m. The chosen pairs of analysed stations are indicated by coloured boxes. (b) Detailed view of the three chosen station pairs F1, F2 and F3 from the array deployment. The plot's border colour corresponds to its location on the map. (c) Configuration of synthetically generated station pairs S1 and S2 with surrounding sources.

a fundamental frequency of approximately 9.9 Hz (with parameters $\mathbf{w}_0 = 5,\ \ s = 1,\ \ M = 101$ samples) and convolving them with the previously described time-series, we are able to generate surface waves that are not undergoing any reflections, refractions or conversions. As a result of activating the source only once, we receive a simplified time-series that contains a single wavelet at a specific traveltime. In order to resemble real-world setups with multiple ambient noise sources more accurately, we increase the complexity of the setting by incorporating 20 randomly distributed sources into the scenario and dispersing them with varying spacing around the station pair (Fig. 2(c), S2). While keeping the position of the sources stationary, we introduce some sources activating multiple times with varying offsets for each of the active sources. This results in a time-series of overlapping signals from various distances and directions. To stabilize the procedure, we further add random noise to the data of both scenarios. Furthermore, we establish identical initial conditions for the model training through the generation of an equivalent amount of synthetic data compared to that present in the field data measurements.

3.2. Field data

To evaluate the applicability of the proposed method across various data scenarios, we employ not only synthetic data but also incorporate field data gathered at a seismic exploration campaign conducted by the OMV E&P GmbH in the Vienna basin, Austria. The array setup consisted of in total 4907 seismic stations, each tooled with either 12 or 24 geophones (vertical components), and spaced with an interstation distance of approximately 200 m (Fig. 2(a)). The measurement period comprises a total duration of about four weeks during March and April 2019 using a sampling rate of 100 Hz. Major and minor roads surround and intersect the region, and a railway line runs along its southern boundary. In addition to these sources of seismic signals, wind turbines and oil pump jacks appear throughout the region (Fig. 2(a)). The wind parks Prottes-Ollersdorf and Grossengersdorf are situated northeast and southwest within the array, while the wind park Deutsch-Wagram is located on its southwestern boundary. Oil pump jacks position in various setups, ranging from individual placements to small clusters and larger groupings within the array. Ocean noise reaches the stations predominantly from the northwest direction. Schippkus *et al.* (2022) provide another detailed description of the array used in this study. The authors explore the impact of an isolated noise source within the framework of seismic interferometry using the same data set. Furthermore, a detailed description of the study area offering background information on the present industry and additional potential sources of noise is given by Schippkus *et al.* (2020). While using a different array than the one in this study, the authors provide detailed insights into the source characteristics of the region by examining spectrograms and power spectral densities.

For the model training, we select three station pairs within the southwestern quarter of the array. The choice of station pairs thereby depends on the respective area conditions in terms of wind turbine and oil pump jack distribution and the distances of these sources to the stations. We successively enhance complexity between the scenarios by increasing the number of surrounding noise sources and consider their spatial proximity to the stations. We evaluate the close vicinity of a wind park in absence of other sources, as well as configurations with and without a wind turbine positioned directly between the stations. Fig. 2(b) shows the three scenarios and their surrounding noise sources. The first station pair F1 situates at the western edge of the array in an area surrounded by fields. The wind park Grossengersdorf is located in the southwestern vicinity of the stations, having its closest wind turbine east of the stations in about ∼110 m to the target station and ∼250 m to the reference station. Situated more towards the centre of the array, the second station pair F2 encircles by wind turbines and oil pump jacks, appearing either individually or in smaller groups. With a distance of ∼55 m to the reference station and about ∼125 m to the target station, a single wind turbine locates between the stations. For the third station pair F3, the number of surrounding sources further increases, particularly witnessing a greater number of oil pump jacks in close proximity to the stations. In contrast to the other station combinations, there is no wind turbine directly next to the stations in this case. The closest wind turbine is located at a distance of about ∼440 m, while the nearest oil pump jack is ∼950 m away. To ensure an appropriate and consistent amount of training and testing data, we limit the measurement of each station to a period of two days.

3.3. Data processing and model training

Prior to starting the training of models for each of the data sets, it is essential to perform pre-processing on the data, as it directly impacts the network's ability to learn accurately. In addition to filtering the data below 10 Hz using a Butterworth low-pass filter, data preparation for both synthetic and field data includes the alignment of all amplitudes to the same range through data scaling. For the synthetic data, we implement normalization to consistently scale the data within the range of $[-1, 1]$. Given the data generation process, we expect only minimal variations within the data, thus eliminating the need for independent centring and scaling using standard scaling methods. With regard to the variety of sources influencing the characteristics of the field data, we anticipate greater variations in range and distribution within this data set. Therefore, we combine both standard scaling and normalization to account for these variations. Initially, standard scaling is applied to centre the data around zero and standardize its deviation to one, followed by normalization to adjust the data to fit within the range of $[-1, 1]$. Before scaling, we allocate 80 percent of the data to the training set and 20 percent to the testing set. Additionally, 20 percent of the training data is automatically determined as the validation set during model training. To ensure successful model training, it is important to provide a sufficient amount of training and testing examples. To do so, we divide the overall time-series of two days into chunks of 10.24 s each, while each chunk corresponds to 1024 samples based on a sampling rate of 100 Hz. Like this, we receive 13.500 chunks for training and 3.375 chunks for testing. In the following, we will refer to these chunks as traces.

We train our encoder–decoder model using pre-processed input traces from reference station A and provide the traces from station B as the target we aim to predict. This way, we obtain a unique model for each station pair that outputs predictions based on the individual data set provided. Subsequently, we compute relevant metrics between the target and prediction to assess the model performance. Following architectural investigations, we empirically determine the optimal network depth by analysing accuracy, convergence and model performance on a sample of the data before starting model trainings. In order to capture the complexity of the data and avoid overfitting, we use a network depth of five layers. Fig. 1 shows the schematic layout of the network having five convolutional and deconvolutional blocks. We train our models with a learning rate of 10^{-4} for 1500 epochs each, as further training beyond this point does not significantly improve performance.

4 RESULTS

Following the training phase, we evaluate the models by calculating the RMSE and cross-correlation coefficient (CC) between the target data and the corresponding model prediction. We assess both metrics on the overall target time-series of two days, and on smaller segments of it. To facilitate the analysis of results and improve the visual representation, we analyse our results within output segments that are half the size (512 samples) of the training and testing traces. When the model captures all relevant transfer features from the data, its predictions will accurately correspond with the unseen target data. To scrutinize the results in terms of positive and negative amplitude deviations, we visualize each sample of the entire target time-series against the model prediction by density plots (Fig. 3(d)). In order to comprehend the correlation dynamics across the whole data set, we further estimate correlation coefficients for each window of 512 samples without any shift and visualize their distribution through a histogram (Fig. 3(e)). Going into further detail, we will analyse a representative example trace (Fig. 3(b)) for each scenario along with its corresponding prediction, correlation coefficient and input data (Fig. 3(a)) from the model training. To understand how the cross-correlation coefficient evolves throughout the data, we link each section of the trace to its corresponding correlation coefficient, as shown in Fig. 3(c). For this, we compute these correlation coefficients using moving intervals of 20-sample windows with a 10-sample offset and visualize the results.

4.1 Synthetic data

Fig. 3 illustrates the results for the two scenarios of synthetic data. For the single source case S1 (Fig. 3(a)–(e)), the model prediction closely aligns with the actual target data (Fig. 3(b)), showing the algorithm's general capability to predict the transfer properties in a very simplified setup. Metrics support this observation, validating the accuracy of predictions and the presence of minimal errors by a small RMSE value of 0.04. While the majority of value pairs cluster around the ideal case of correct amplitude predictions as indicated by the dotted purple line in Fig. 3(d), some segments show significant deviations, highlighting instances where the prediction does not align with the target values. The overall cross-correlation coefficient of 0.90, as shown in Fig. 3(e), reflects the predominance of good fits, though it moderates by the occurrence of some less accurate predictions. The histogram shows, that a majority of traces display a correlation coefficient close to one, while another distinct cluster is observed around zero. We attribute the latter cluster, observed around zero, to the random noise introduced in the data, which adds variability but does not necessarily indicate a systematic relationship. Consequently, the predictions do not align with the target, leading to CCs near zero. This observation is confirmed by the analysis of CCs in smaller windows (Fig. 3(c)), which indicate strong correlations when predicting the wavelet at given traveltime,

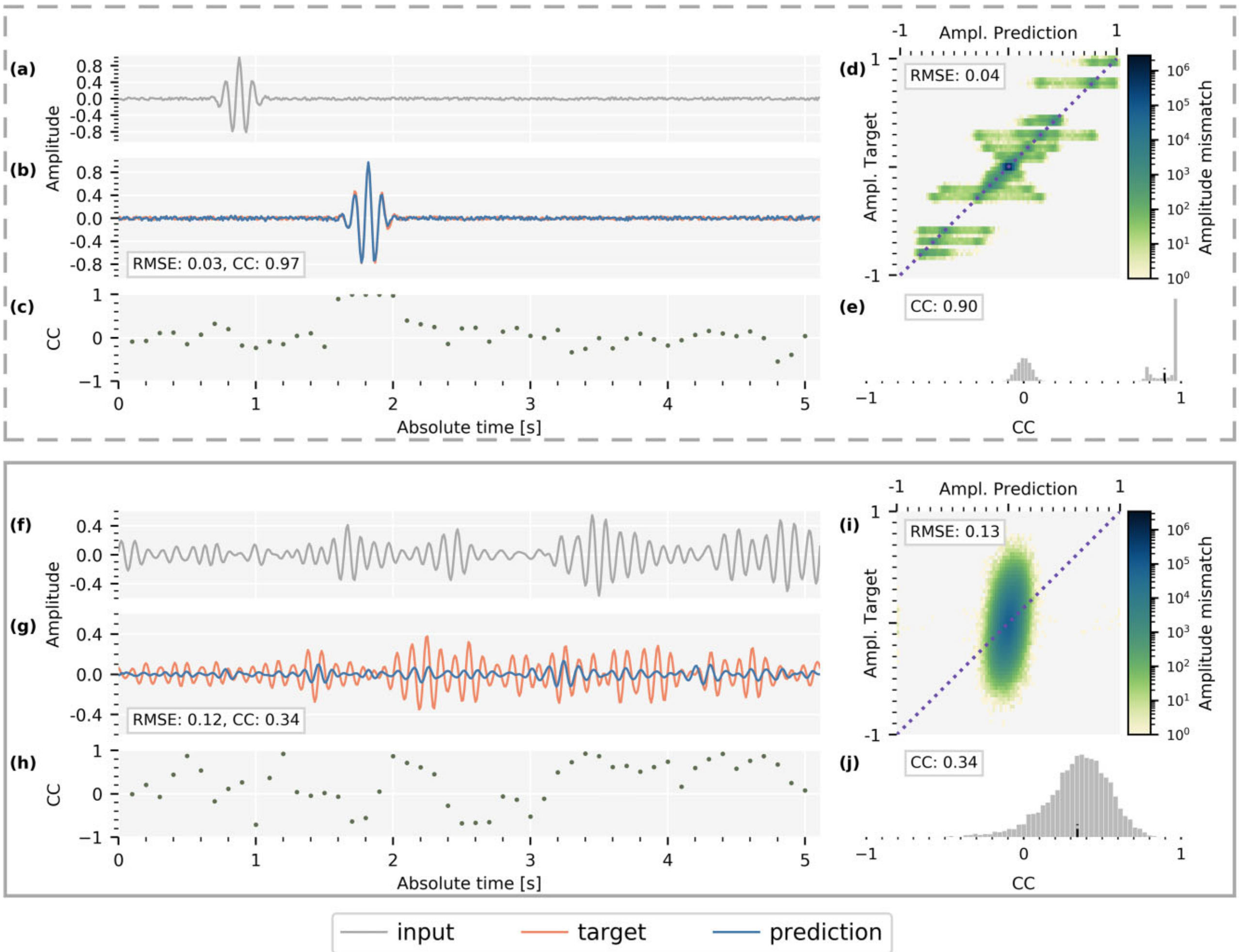


Figure 3. Results of the model analysis for two synthetic data scenarios (S1, top—dashed box; S2, bottom—solid box) using RMSE and cross-correlation coefficient (CC) across the entire target time-series and within traces. Panels (a) and (f) represent the input data, panels (b) and (g) denote the target data and network prediction, respectively. The density plots (d) and (i) show the network prediction of the target against the actual target data, while the dotted line visualizes the ideal best-fitting line for the regression. Single points ((c), (h)) depict cross-correlation coefficients for 20-sample sections beneath the corresponding example trace. Histograms ((e), (j)) show correlation coefficients for windows of 512 samples each. The black marker highlights the overall correlation coefficient of the entire time-series given in the text box.

whereas the correlations of random noise components are significantly lower.

The presence of 19 sources surrounding the stations (scenario S2, Fig. 3(f)–(j)) introduces increased variability to the data, evident in the time-series as overlapping signals with varying amplitudes (Fig. 3(g)). While certain segments of the target trace align with the model predictions (Fig. 3(g)), other parts reveal disparities in either amplitude or the general shape of the wavelet. The overall correlation coefficient of 0.34 (Fig. 3(j)), along with the RMSE of 0.13 (Fig. 3(i)) highlights larger differences in the similarity of the time-series compared to the single source case S1. However, the given overall correlation coefficient of 0.34 indicates a predominantly positive correlation, implying that the model is able to approach a modest similarity between its predictions and the target data. The analysis of the correlation coefficients for individual traces (Fig. 3(j)) reveals characteristics of Gaussian-like distribution, with the majority of values concentrated between 0.1 and 0.4, and some traces reaching an upper limit near 0.8 and a lower limit around −0.4. Analysing the kind of differences between the target and predictions (Fig. 3(i)) demonstrates a slightly tilted elliptical shape of amplitude mismatch around the centre indicating that amplitudes are more commonly underestimated than overestimated.

The outcomes from both scenarios reveal promising indications of the feasibility of this model architecture. While they exhibit significant differences in performance, both models generate predictions that indicate patterns, for example, predicting phases accurately rather than appearing random. Especially scenario S1 thereby demonstrates the potential of the overall algorithm to learn the transfer between two nearby stations, despite not accurately representing real-world conditions. Even in the second scenario, the prediction maintains a reasonable level of accuracy. This establishes a solid foundation for the transition to field data, mirroring a comparable scenario where two seismic stations are encircled by seismic sources.

4.2 Field data

Fig. 4 depicts the outcomes for the three field data scenarios. Beginning with the first station pair located at the array's edge near a wind park (F1, Fig. 4(a)–(e)), the model prediction closely aligns with the actual target data in various segments (Fig. 4(b)). While we observe positive and negative deviations in amplitudes between target and prediction in several parts of the trace, the prediction of phases exhibits accurate matches with the target time-series. Fig. 4(c) confirms this observation, as the correlation coefficients for the majority of trace subparts cluster near one, highlighting the model's accuracy in predicting phase information. With an overall correlation coefficient of 0.75, the concentration of individual correlation coefficients (Fig. 4(e)) is mostly within the positive range of 0 to 1, having its peak strength at a high correlation value of around 0.8. A cluster of values around 0.75 characterizes the central tendency of the data set and emphasizes further the models ability to make predictions of similarity to the target data. Besides, there is another notable peak around −0.58 and −0.78, likely attributable to data gaps present in this data set leading to inaccurate predictions in the negative range. Evident from the elliptical shape, the density plot (Fig. 4(d)) reveals positive and negative mismatches of amplitudes along the dotted purple best-fitting line. Thus, both positive and negative amplitude predictions display tendencies of overfitting and underfitting, reflecting some variability in the model's capacity to accurately estimate amplitude values. The RMSE reflects this with an average deviation of around 0.13 units of amplitude between predicted and actual values. In comparison to previous examples, this scenario indicates a relatively broader distribution of amplitude values stretching to the lower and upper limits of the data range.

While wind turbines are already in close proximity to the stations in the first case F1, the distance further halves for the second scenario F2 (Fig. 4(f)–(j)), where a wind turbine is located directly between both stations. Although the overall correlation coefficient of 0.77 (Fig. 4(j)) is nearly identical to one of the previous example, there are visual differences regarding the data itself and the model outcomes leading to variations in the results. Examining the example trace (Fig. 4(g)), the predicted phases largely correspond with those of the target data again. The predominance of correlation coefficients close to one supports this observation (Fig. 4(h)), although minor or negative coefficients occur occasionally. However, the amplitude predictions again exhibit greater variances compared to the targets. While the RMSE for the selected trace is 0.06, the global RMSE measures at 0.08 (Fig. 4(i)), indicating more accurate amplitude predictions for this station pair compared to case F1. This is also evident when looking at the distribution of values around the purple dotted best-fitting line of the plot. Similar to the initial example, the analysis reveals a tendency to both overfit and underfit, affecting the accuracy of predictions for both positive and negative amplitude values. However, the data set predominantly consists of smaller values, leading to reduced variability and a narrower range of data dispersion. Following this, the individual correlation coefficients of traces (Fig. 4(j)) not only approximate a nearly Gaussian distribution again but also display increased steepness, indicating a tighter clustering of values around the mean.

The third station pair, F3, unique among the combinations as it lacks a wind turbine in direct proximity to the stations, leads to an overall correlation coefficient of 0.58, as shown in Fig. 4(o). Although the overall correlation coefficient represents a decrease relative to those found in earlier field data examples, the amplitude deviations, characterized by an RMSE of 0.10, lie within an intermediate range compared to the observations from the prior two cases. In this instance, as shown in Fig. 4(n), the comparison of target and predicted amplitudes reveals an elliptical shape again. However, the ellipse appears more circular in comparison to previous cases, suggesting that it represents an intermediate scenario in terms of the spread and steepness. Additionally, there is the same tendency of over- and underestimation as in previous scenarios. The analysis of the model's performance on the example trace (Fig. 4(l)) shows disparities between the target and predicted values in certain intervals, whereas other sections align well. Correlation coefficients, derived from 20-sample segments of the example trace (Fig. 4(m)), confirm this impression: values near one mirror precise phase predictions or moderate amplitude fits, while values at or below zero point to negative predictions. The analysis of correlation coefficients for trace windows (Fig. 4(o)) reveals that most bins lie within the positive range of 0.2 to 0.8, while we identify one large peak above 0.9. This indicates the presence of a generally positive linear relationship between input features and the models output predictions, affirming the model's effectiveness in detecting data patterns to a certain degree.

All three field data examples exhibit moderate to strong linear correlation, providing predictions that resemble the actual target to a high degree. In this regard, both the visual assessment and the evaluation metrics surpass the performance of the second synthetic example S2, which mirrors comparable environmental conditions of having sources distributed around the stations. However, the field data examples do not achieve the level of accuracy seen in the perfect synthetic case S1, suggesting that the performance of field data models ranges somewhere between these two extremes. Physically, we also expect the source regime to be a hybrid between single source and evenly distributed sources, with a tendency towards a few significant sources.

5 DISCUSSION

The model architecture described in this study shows the capability to predict the transfer properties, in our case the 1-D time-series, between two seismic stations in different source-station-setups. Employing diverse scenarios of both synthetic data (Fig. 3) as a controlled environment, and field data (Fig. 4), representing real-world conditions, delivers a comprehensive proof of concept across different data sets. Overall, the models demonstrate strong predictive performance, particularly in predicting the phase of the wavefield more reliably than its amplitude, as demonstrated by both synthetic (Fig. 3) and field data examples (Fig. 4). While the models manage the novel scenario of differing input and target data effectively, further optimization by fine-tuning various factors, such as hyperparameters, could affect the algorithms performance even further (Weerts *et al.* 2020; Yang & Shami 2020; Bakhashwain & Sagheer 2021).

Particularly evident in the scenario S1 (Fig. 3(a)–(e)), the model training benefits significantly by considering only a single source from one direction and random noise, representing an idealized scenario. This simplification yields favourable results, underscoring the network's general ability to learn a given relationship in a controlled

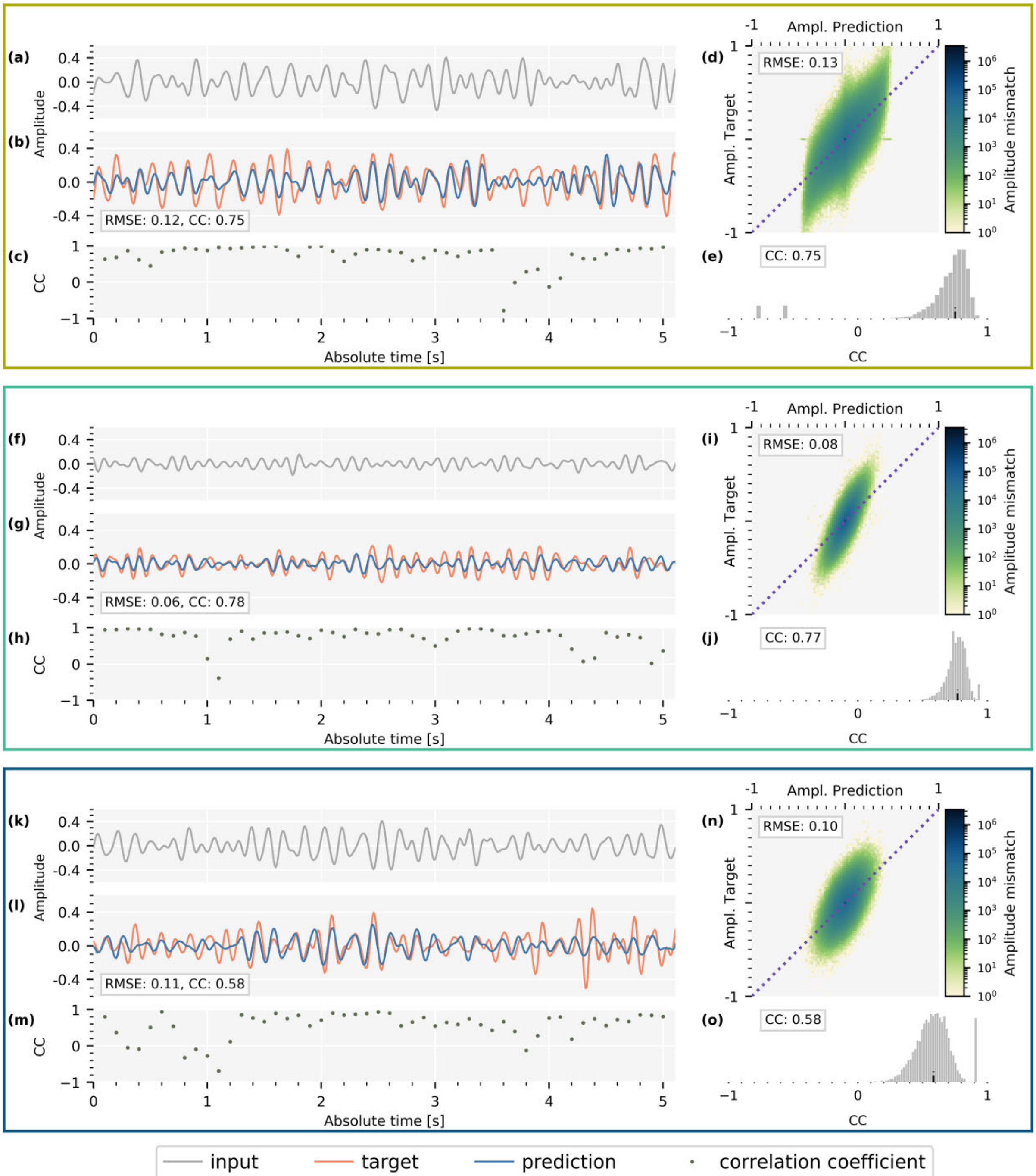


Figure 4. Results on the comparative analysis (cf. Fig. 3) for three field data scenarios (F1 top, F2 middle, F3 bottom). Plots provide insights into the examination of correlation dynamics, magnitude deviations and distribution patterns using RMSE and cross-correlation coefficient (CC) for evaluation. The results for this data set follow the same evaluation criteria and presentation as in Fig. 3. The frame colour of the box indicates the corresponding scenario.

setting. Moving to scenario S2, the approach handles a higher level of complexity introduced by simultaneous inputs from multiple directions. Despite these challenges, the algorithm maintains a decent level of performance, as evident by the mean correlation coefficient (CC) of 0.34. When comparing the performance of this second synthetic example S2 (Fig. 3(f)–(j)) with that of the field data models, all results from the field data exceed the performance observed in S2 with CCs of 0.58, 0.75 and 0.77. Although scenario S2 may seem initially favourable due to the uniform energy propagation from the synthetic sources, the observed performance improvement in the field data is likely driven by the unique characteristics of different sources around the array, such as wind turbines, oil pumps or roads. Despite similarities in source distributions between the two setups, our field data do not exhibit the extreme conditions of S2, demonstrating the robustness and practical applicability of our approach in more natural and realistic scenarios.

While the CC threshold values we obtain might be considered relatively low or almost comparable in the context of some ML studies (Wu *et al.* 2021; Verma *et al.* 2024), in seismology, Schaff *et al.*

(2002) used a normalized CC of 0.7 as a criterion for reliable relative arrival time measurements. Wegler & Sens-Schönfelder (2007) considered only cross-correlation coefficients above 0.5 for their d*v*/*v* analysis, and Castellanos *et al.* (2020) only use traveltime measurements with CCs larger than 0.5 as reliable. Although the research settings of their studies differ from ours, the range of values of CC thresholds in 'real data' studies underscores its broad applicability as a measure of data reliability. This alignment emphasizes that our use of noise field data is yielding results that are comparably robust and reliable.

In addition to the equal distribution of energy for each source in the synthetic data scenarios, the inconsistent and repeated activation of these sources may fail to generate learnable characteristics in the data set. The absence of pattern-like attributes introduces challenges for the algorithms learning process, as they represent essential relationships within the data that models are trained to learn and utilize. While this could potentially create challenges with our synthetic data, the situation shifts with the nature of sources present in the field data, which exhibit consistent and repetitive signals. Given the distribution of surrounding sources in the field data examples (Fig. 2b), we account for the presence of wind turbines at various distances in each scenario. Neuffer *et al.* (2019) demonstrate that wind turbines show directional characteristics with wind-dependent specific patterns. Anticipating these sources to introduce distinct patterns by the propagation of similar signals, we expect them to provide valuable input to the model training and enhance its predictive accuracy. Given that our results improve when wind turbines are in close proximity to the stations, the presence of such noise sources appears to resemble the characteristics of a single source and thus positively influences the model's performance.

While it is evident that consistently emitting sources such as wind turbines positively impact our results, it is not immediately clear why we observe stronger accuracy across various data sets in the prediction of phases, while our models preferentially underestimate amplitudes (Figs. 3 and 4). Given that neither the area of investigation nor the characteristics of the sources and stations indicate any physical phenomena that could account for these deviations, it appears that there are no evident physical processes to explain this behaviour. Consequently, we will focus our investigation on the data as well as the architecture and parameters of the models as potential cause. The fundamental nature of encoder–decoder networks, particularly autoencoders and the ones used for sequence-to-sequence learning, aims to capture and reconstruct patterns in the data. Identifying both the sequence of events over time and underlying patterns within the data allows these networks to extract key features and reduce dimensionality to focus on key relationships. However, encoder–decoder networks learn to prioritize certain data characteristics based on their architecture, which determines how layers and connections handle specific features in the data. Weights and biases, for example, adjust during the training process and influence which data attributes are prioritized. Furthermore, input data properties such as dimensionality or level of noise, can affect the process of prioritization and learning of various patterns. Upon visually inspecting our data, it becomes apparent that the spacing between phases of our time-series appears to be relatively consistent. This can be attributed to two key factors: the application of filtering and the dominance of a relatively narrow frequency band in the remaining frequencies. While this is true for the phases, amplitudes vary between high and low values and span from positive to negative, which poses a greater challenge for the model to learn properties of these more complex patterns. Furthermore, neural networks can exhibit a low-frequency bias, meaning they learn low-frequency components more easily than high-frequency details (Rahaman *et al.* 2019). Since an exact fit of amplitudes requires learning both high and low-frequency patterns, this bias can lead to difficulties in capturing the full amplitude spectra (Song *et al.* 2021; Rasht-Behesht *et al.* 2022; Ren *et al.* 2022) and may result in the underestimation of amplitude variations. Besides the architecture of encoder–decoder networks and the quality of training data, the choice of parameters like learning rate, batch size, activation or loss function can affect the model performance. As an example, while we initially used the tanh activation function, we also evaluated the sigmoid and linear activation functions. Both yielded comparable results but did not improve the underestimation of amplitudes observed. This highlights the need for additional research for a comprehensive understanding.

Although it is not obvious to us why amplitudes are preferentially underestimated, many seismological applications rely entirely on the phases of seismograms. Our models reliably predict the phase of seismic noise. For instance, phases from seismic waves are essential for determining arrival times of different waves, which help to locate earthquake epicentres and understand Earth's internal properties. In addition, phase-based investigations, such as ambient noise tomography and seismic interferometry, predominantly rely on phase information to extract subsurface details. As highlighted by Bensen *et al.* (2007), ambient noise data processing involves steps like cross-correlation and temporal stacking, which are inherently phase-dependent, and the accurate measurement of dispersion curves, which utilize phase and group speeds. Seismic interferometry, for example, involves the cross-correlation of seismic recordings at different stations, allowing researchers to reconstruct the Green's function between two points using phase information. This highlights that while our amplitude predictions may be less precise, the critical phase information remains robust and useful for various seismological analyses.

In general, choosing an encoder–decoder architecture suits the requirements of the given problem, as it is able to capture complex relationships and generalizes well to unseen data. Traditionally, this approach is used to predict future values of time-series based on historical trends, using past data as input to forecast subsequent values within the same series. Using it with input data from one seismic station and target data from another seismic station thereby diverges from this conventional application as well as from classical autoencoders. While autoencoders aim to learn a compressed representation of the input data, the proposed architecture extends this approach to learn and predict the relationship between data from distinct stations. In other words, our model learns the propagation of complex wave fields between the stations. This allows to model spatial and temporal dependencies between seismic data relying on the phase and amplitude information of the signals. While our architecture encounters challenges in explicitly modelling spatial correlations, its ability to predict the translation of seismic signals from one station to another shows that the temporal factors provide sufficient information for high-quality predictions. This highlights the U-Net's robustness, even without explicitly incorporating spatial correlations into the modelling. This also raises new possibilities by assuming the integration of only minimal spatial dependencies.

One might draw parallels between this approach and Wiener prediction filters (Chen *et al.* 2006; Chandra *et al.* 2014), which also aim to capture dependencies within signal data. However, it is important to note that Wiener prediction filters primarily deal with the autocorrelation of signals, focusing on their power spectrum without considering phase information. Wiener filtering assumes non-deterministic signals, which contradicts seismic signals known for their deterministic nature, such as reflections from layered structures. In contrast, our method comprehensively accounts for the dynamic, nonlinear interactions and phase information essential for accurately modelling wave propagation in seismology.

To advance our approach from the proof-of-concept stage described herein to concrete applications, several aspects will likely need to be addressed. These could encompass technical and structural enhancements that include the improvement of data quality, fine-tuning of hyperparameters or the accuracy of amplitude predictions. Additionally, we might consider adjustments to the synthetic data generation process to better resemble conditions encountered in field environments with multiple ambient noise sources. Future studies might also delve into how the geographical and spatial configuration of sources and receivers impacts the results. Given the differences in model performance in predicting phases and amplitudes, experimenting with different model architectures and parameters could further be advantageous. By implementing these modifications to the model setup and understanding influences on the results in detail, we can further refine the overall performance and robustness of our approach.

6 CONCLUSION

In this study, we have successfully presented and tested an adaption of encoder–decoder networks to predict the transfer function of seismic wavefields, between two seismic stations. By introducing 1-D time-series data from a fixed seismic station as the input to the network and data from a nearby station as the target, our approach effectively learns the transfer function between the locations. Initially tested with synthetic data, the approach was validated further with field data from a seismic exploration campaign. Employing a range of scenarios with varying surrounding conditions—from a controlled environment with synthetic data to field data including several sources of ambient noise—we demonstrate a broad proof of concept.

Our findings confirm that our approach effectively predicts the wavefield recorded as time-series at a seismic station using input from a neighbouring seismic station, resulting in machine learning models with varying degrees of accuracy. Notably, our models not only achieve high precision in predicting the phases of seismic waves but also perform adequately in estimating amplitudes, demonstrating significant potential for the field of geophysical research. This makes our approach particularly valuable for applications requiring precise seismic isolation or compensation, such as active vibration isolation in photolithography, semiconductor manufacturing and 3-D microfabrication (Kerber *et al.* 2007; Kim *et al.* 2009). It is also highly relevant for projects like the Einstein Telescope (Punturo *et al.* 2010; Harms *et al.* 2022), where extremely sensitive gravitational wave detections need to be free from seismic disturbances. Additionally, our approach also opens up the potential for the novel concept of virtual seismic arrays.

FUNDING

This work is financially supported by the Federal Ministry of Education and Research (BMBF) project '3G-GWD' with references 05A20GU5 and 05A23GU5 and partially funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 955515 (SPIN ITN—https://spin-itn.eu).

ACKNOWLEDGMENTS

The authors thank OMV E&P (OMV E&P GmbH 2019) for providing access to the seismic data and granting permission to publish these findings. Special thanks goes to Conny Hammer and Alexander Bauer for their insightful comments on technical questions. We thank the reviewers, Leonard Seydoux and one anonymous, and the editor Bertrand Rouet-Leduc and assistant editor Louise Alexander for their insightful comments that helped improve the manuscript.

Author contributions: JK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing review and editing. SS: Conceptualization, Data curation, Supervision, Writing review and editing. JW: Conceptualization, Software. CH: Conceptualization, Funding acquisition, Supervision, Writing review and editing. DG: Conceptualization, Funding acquisition, Supervision, Writing review and editing. All authors have read and agreed to the published version of the manuscript

DATA AVAILABILITY

The data underlying the findings of this study were provided by OMV Exploration and Production GmbH. Restrictions apply to the availability of these data, which were used under license for this study. The colour scheme utilized for the scatterplots in this study was sourced from Crameri (2023). The colour scheme for the time-series data and the frames used to emphasize the selected station pairs was taken from Tol, Paul (2021).

REFERENCES

Badrinarayanan, V., Kendall, A. & Cipolla, R., 2017. SegNet: a deep convolutional encoder–decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.,* **39,** 2481–2495.

Bakhashwain, N. & Sagheer, A., 2021. Online tuning of hyperparameters in deep LSTM for time series applications, *Int. J. Intell. Eng. Syst* **14,** 1, 212.

Barnes, A.E., 2007. A tutorial on complex seismic trace analysis, *Geophysics,* **72,** W33–W43.

Bath, M., 1973. *Introduction to Seismology,* Halsted Press book. Wiley.

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.,* **169,** 1239–1260.

Beveren, V. van, Bader, M., Brand, J. van den, Bulten, H.J., Campman, X., Koley, S. & Linde, F., 2023. A study of deep neural networks for Newtonian noise subtraction at Terziet in Limburg—The Euregio Meuse-Rhine candidate site for Einstein Telescope, *Class. Quantum Gravity,* **40,** 205008.

Castellanos, J.C., Clayton, R.W. & Juarez, A., 2020. Using a time-based subarray method to extract and invert noise-derived body waves at Long Beach, California, *J. geophys. Res. Solid Earth,* **125,** e2019JB018855.

Chandra, G., Yadav, S., Krishna, B.A. & Kamaraju, M., 2014. Performance of wiener filter and adaptive filter for noise cancellation in real-time environment, *Int. J. Comput. Appl.,* **97.**

Chen, J., Benesty, J., Huang, Y. & Doclo, S., 2006. New insights into the noise reduction wiener filter, *IEEE Trans. Audio Speech Lang. Process.,* **14,** 1218–1234.

Chollet, F., others, 2015. Keras [WWW Document]. URLhttps://github.com/fchollet/keras

Crameri, F., 2023. Scientific colour maps. Available at: https://www.fabiocrameri.ch/colourmaps/

Curtis, A., Behr, Y., Entwistle, E., Galetti, E., Townend, J. & Bannister, S., 2012. The benefit of hindsight in observational science: retrospective seismological observations, *Earth planet. Sci. Lett.,* **345–348,** 212–220.

Du, S., Li, T., Yang, Y. & Horng, S.-J., 2020. Multivariate time series forecasting via attention-based encoder–decoder framework, *Neurocomputing,* **388,** 269–279.
Harms, J., Naticchioni, L., Calloni, E., De Rosa, R., Ricci, F. & D'Urso, D., 2022. A lower limit for Newtonian-noise models of the Einstein Telescope, *Eur. Phys. J. Plus,* **137,** 687.
Havskov, J. & Alguacil, G., 2016. *Correction for Instrument Response in Instrumentation in Earthquake Seismology,* pp. 197–230, Springer International Publishing. doi:
Helmy, T., Anifowose, F. & Faisal, K., 2010. Hybrid computational models for the characterization of oil and gas reservoirs, *Expert Syst. Appl.,* **37,** 5353–5363.
Kawakami, H. & Oyunchimeg, M., 2003. Normalized input–output minimization analysis of wave propagation in buildings, *Eng. Struct.,* **25,** 1429–1442.
Kerber, F., Hurlebaus, S., Beadle, B.M. & Stöbener, U., 2007. Control concepts for an active vibration isolation system, *Mech. Syst. Signal Process.,* **21,** 3042–3059.
Kim, S.J., Dean, R., Flowers, G. & Chen, C., 2009. Active vibration control and isolation for micromachined devices, *J. Mech. Des.,* **131**.
Knispel, S., Walda, J., Zehn, R., Bauer, A. & Gajewski, D., 2022. A self-attention enhanced encoder–decoder network for seismic data denoising, in *Second International Meeting for Applied Geoscience & Energy, Society of Exploration Geophysicists and American Association of Petroleum,* Society of Exploration Geophysicists and American Association of Petroleum, pp. 2922–2926.
LeCun, Y., Bengio, Y. & Hinton, G., 2015. Deep learning, *Nature,* **521,** 436–444.
Li, D., Peng, S., Lu, Y., Guo, Y. & Cui, X., 2019. Seismic structure interpretation based on machine learning: a case study in coal mining, *Interpretation,* **7,** 1–44.
Li, W., Chakraborty, M., Fenner, D., Faber, J., Zhou, K., Rümpker, G., Stöcker, H. & Srivastava, N., 2022. EPick: attention-based multi-scale UNet for earthquake detection and seismic phase picking, *Front. Earth Sci.,* **10**.
Li, Z., Meier, M.-A., Hauksson, E., Zhan, Z. & Andrews, J., 2018. Machine learning seismic wave discrimination: application to Earthquake early warning, *Geophys. Res. Lett.,* **45,** 4773–4779.
Lindsey, N.J., Rademacher, H. & Ajo-Franklin, J.B., 2020. On the broadband instrument response of Fiber-optic DAS arrays, *J. Geophys. Res. Solid Earth,* **125,** e2019JB018145.
Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P. & Shroff, G., 2016. LSTM-based encoder-decoder for multi-sensor anomaly detection, CoRR abs/1607.00148.
Mousavi, S.M. & Beroza, G.C., 2023. Machine learning in earthquake seismology, *Annu. Rev. Earth Planet. Sci.,* **51,** 105–129.
Neuffer, T., Kremers, S. & Fritschen, R., 2019. Characterization of seismic signals induced by the operation of wind turbines in North Rhine-Westphalia (NRW), *Germany. J. Seismol.,* **23,** 1161–1177.
OMV E&P GmbH, 2019. OMV.
Punturo, M. *et al.*, 2010. The Einstein Telescope: a third-generation gravitational wave observatory, *Class. Quantum Gravity,* **27,** 194002.
Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y. & Courville, A., 2019. On the spectral bias of neural networks, in *Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research,* eds Chaudhuri, K. & Salakhutdinov, R., PMLR, Long Beach, CA, pp. 5301–5310.
Rasht-Behesht, M., Huber, C., Shukla, K. & Karniadakis, G.E., 2022. Physics-informed neural networks (PINNs) for wave propagation and full waveform inversions, *J. geophys. Res. Solid Earth,* **127,** e2021JB023120.
Ren, P., Rao, C., Chen, S., Wang, J.-X., Sun, H. & Liu, Y., 2024. SeismicNet: physics-informed neural networks for seismic wave modeling in semi-infinite domain, Computer Physics Communications, **295,** 109010.
Ronneberger, O., Fischer, P. & Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference,* October 5-9, 2015, Proceedings, Part III 18. Springer, Munich, Germany, pp. 234–241.
Rost, S. & Thomas, C., 2002. Array seismology: methods and applications, *Rev. Geophys.,* **40,** 2–1.
Saad, O.M. & Chen, Y., 2020. Deep denoising autoencoder for seismic random noise attenuation, *Geophysics,* **85,** V367–V376.
Sabra, K.G., Gerstoft, P., Roux, P., Kuperman, W.A. & Fehler, M.C., 2005. Extracting time-domain Green's function estimates from ambient seismic noise, *Geophys. Res. Lett.,* **32**.
Schaff, D.P., Bokelmann, G.H.R., Beroza, G.C., Waldhauser, F. & Ellsworth, W.L., 2002. High-resolution image of Calaveras Fault seismicity, *J. geophys. Res. Solid Earth,* **107,** ESE 5–1-ESE 5-16.
Schippkus, S., Garden, M. & Bokelmann, G., 2020. Characteristics of the ambient seismic field on a large-N seismic array in the Vienna Basin, *Seismol. Res. Lett.,* **91,** 2803–2816.
Schippkus, S., Snieder, R. & Hadziioannou, C., 2022. Seismic interferometry in the presence of an isolated noise source, *Seismica,* **1**.
Sergeant, A., Chmiel, M., Lindner, F., Walter, F., Roux, P., Chaput, J., Gimbert, F. & Mordret, A., 2020. On the Green's function emergence from interferometry of seismic wave fields generated in high-melt glaciers: implications for passive imaging and monitoring, *The Cryosphere,* **14,** 1139–1171.
Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: a derivation based on stationary phase, *Phys. Rev. E,* **69,** 046610.
Song, C., Alkhalifah, T. & Waheed, U.B., 2021. Solving the frequency-domain acoustic VTI wave equation using physics-informed neural networks, *Geophys. J. Int.,* **225,** 846–859.
Tariq, Z., Aljawad, M., Hassan, A., Murtaza, M., Mohammed, E., El Husseiny, A., Alarifi, S. & Abdulraheem, A., 2021. A systematic review of data science and machine learning applications to the oil and gas industry, *J. Pet. Explor. Prod. Technol.,* **11,** 4339–4374.
Tol, P., 2021. "Colour schemes," Technical note SRON/EPS/TN/09-002 3.2. SRON. Available at: https://personal.sron.nl/∼pault/data/colourschemes.pdf
Verma, N. *et al.*, 2024. Comparison of neural networks techniques to predict subsurface parameters based on seismic inversion: a machine learning approach, *Earth Sci. Inform.,* **17,** 1031–1052.
Walden, A.T. & White, R.E., 1998. Seismic wavelet estimation: a frequency domain solution to a geophysical noisy input-output problem, *IEEE Trans. Geosci. Remote Sens.,* **36,** 287–297.
Weerts, H.J.P., Mueller, A.C. & Vanschoren, J., 2020. Importance of tuning hyperparameters of machine learning algorithms, arXiv preprint arXiv:2007.07588.
Wegler, U. & Sens-Schönfelder, C., 2007. Fault zone monitoring with passive image interferometry, *Geophys. J. Int.,* **168,** 1029–1033.
Wu, H., Zhang, B., Lin, T., Cao, D. & Lou, Y., 2019. Semiautomated seismic horizon interpretation using the encoder-decoder convolutional neural network, *Geophysics,* **84,** B403–B417.
Wu, Y., Wang, W., Zhu, G. & Wang, P., 2021. Application of seismic multi-attribute machine learning to determine coal strata thickness, *J. Geophys. Eng.,* **18,** 834–844.
Xie, Y., Ebad Sichani, M., Padgett, J. & Desroches, R., 2020. The promise of implementing machine learning in earthquake engineering: a state-of-the-art review, *Earthq. Spectra,* **36,** 875529302091941.
Yang, L. & Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: theory and practice, *Neurocomputing,* **415,** 295–316.
Yin, J., Denolle, M.A. & He, B., 2022. A multitask encoder–decoder to separate earthquake and ambient noise signal in seismograms, *Geophys. J. Int.,* **231,** 1806–1822.
Zhang, H., Chen, T., Liu, Y., Zhang, Y. & Liu, J., 2021. Automatic seismic facies interpretation using supervised deep learning, *Geophysics,* **86,** IM15–IM33.
Zhong, T., Cheng, M., Dong, X., Li, Y. & Wu, N., 2022. Seismic random noise suppression by using deep residual U-net, *J. Pet. Sci. Eng.,* **209,** 109901.
Zhu, W. & Beroza, G.C., 2019. PhaseNet: a deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.,* **216,** 261–273.

3. Study II: Introduction of Virtual Seismic Arrays

Submitted to *Geophysical Research Letters*

Klinge J., Schippkus S., Hadziioannou C. (2025). **Unlocking the potential of single stations to replace seismic arrays**, *Geophysical Research Letters*

Author contributions:

- **JK**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing original draft, Writing review and editing.
- **SS**: Conceptualization, Resources, Methodology, Project Administration, Supervision, Writing review and editing.
- **CH**: Conceptualization, Funding acquisition, Supervision, Writing review and editing.

Unlocking the potential of single stations to replace seismic arrays

Jana Klinge[1], Sven Schippkus[1], and Céline Hadziioannou[1]

[1]Institute of Geophysics, Centre for Earth System Research and Sustainability (CEN), Universität Hamburg, Hamburg, Germany

Key Points:

- We introduce Virtual Seismic Arrays to predict array recordings from a single station for data collection after physical array removal.
- To capture transfer characteristics between stations, we train deep learning models on GRF array recordings of secondary microseisms.
- The Virtual Seismic Array's performance is assessed through beamforming, showing strong agreement between predicted and original results.

Corresponding author: Jana Klinge, jana.klinge@uni-hamburg.de

Abstract

We introduce Virtual Seismic Arrays, which predict full array recordings from a single reference station, eliminating the need for continuous deployment of all stations. This innovation can reduce costs and logistical challenges while maintaining multi-station functionality. We implement a Virtual Seismic Array using a deep learning encoder-decoder approach to predict transfer properties between stations. Training on recordings from the Gräfenberg array in the secondary microseism frequency band allows us to retrieve models capturing transfer characteristics between stations. These models form the Virtual Seismic Array. To evaluate performance, we beamform original and predicted waveforms to detect dominant secondary microseism sources. We assess three scenarios: one aligning with the training dataset, another with two regimes in training but testing on one, and a third where training data does not align with the testing regime. Our results show strong agreement between predicted and original beamforming results, demonstrating the potential of Virtual Seismic Arrays.

Plain Language Summary

In geophysics, analyzing various signals, including those from earthquakes and environmental noise, is essential for understanding Earth's behavior and seismic activity. To collect seismic data, groups of sensors known as seismic arrays are used. This approach helps analyzing seismic data and improves our ability to locate where seismic waves come from. In this study, we introduce Virtual Seismic Arrays, which use data from one reference station within the array to predict the recordings captured by the entire array. This reduces the need to have all stations actively collecting data all the time. We create this Virtual Seismic Array using a deep learning method called encoder-decoder networks, which learns how seismic signals propagate between stations. We train our algorithm using data from the Gräfenberg array, focusing on a frequency range of 0.1 to 0.25 Hz associated with ocean-generated seismic waves. This way, we create models that learn and predict how the reference station is related to the other stations in the array. To evaluate how well our Virtual Seismic Array works, we compare original and predicted data across different cases. Our results show a strong similarity between original and predicted data, demonstrating the potential of Virtual Seismic Arrays for future applications.

1 Introduction

Seismic arrays are an essential approach to collect and analyze seismic data, improving the understanding of geophysical processes like seismic source localization and determination of large- and fine-scale structures of the Earth's interior (Gibbons & Ringdal, 2006; Rost & Thomas, 2009; Schweitzer et al., 2012). By increasing the capability to detect seismic events, seismic arrays significantly improve seismic monitoring and allow for insights into wave propagation phenomena. An essential processing technique enabling this is beamforming, which allows directional signal detection by combining multiple sensor inputs, thereby improving signal-to-noise ratio and allowing seismic arrays to operate as wave number filters (Capon et al., 1967; Rost & Thomas, 2002; Wang et al., 2020). Array beamforming is also widely applied in other fields, such as ultrasound and astronomy, where it improves image resolution and diagnostic precision (J. Y. Lu et al., 1994; Holfort et al., 2009; Luijten et al., 2020), as well as the sensitivity of observations (van der Veen et al., 2004; Warnick et al., 2016).

This research aims to advance seismic observation techniques by introducing the concept of Virtual Seismic Arrays, which could help vastly decrease costs and improve monitoring in environments with limited resources or insufficient seismic infrastructure. A Virtual Array is able to acquire seismic data from previously instrumented areas even after the physical sensors have been removed. The concept includes the prediction of array recordings based on data from a single reference station, which was originally part of the array, thereby eliminating the need for continuous deployment of all array stations. While multiple stations often provide advantages in event location and signal characterization (Gibbons & Ringdal, 2006; Rost & Thomas, 2009), they can be challenging to deploy, require regular maintenance, and are associated to high operating costs. Our new approach uses deep learning to achieve capabilities similar to those of seismic arrays with just a single station. By learning signal propagation characteristics between a reference station and all stations within a seismic array, our method maintains the ability to monitor seismic activity effectively, while significantly reducing the physical infrastructure required. We note that the term "virtual seismic array" has been used before in a different context (Alhukail, 2012), where it refers to an approach for enhancing the response of an existing seismic array. In contrast, our approach aims to allow the continued operation of seismic arrays after most stations have been removed from the field.

Encoder-decoder networks, a type of neural network, have been widely applied in seismology to learn complex patterns from seismic data enabling earthquake event classification (W. Li et al., 2022), fault detection (S. Li et al., 2019) or seismic inversion (Gelboim et al., 2023). Klinge et al. (2025) employed encoder-decoder networks that represent the transfer function between two seismic stations by learning the underlying signal transformations. Within a supervised framework, the network is trained with time series data from a fixed seismic reference station to successfully predict measurements of different neighboring stations (Klinge et al., 2025).

We investigate the applicability of encoder-decoder networks to realize Virtual Seismic Arrays. As a proof-of-concept, we train a Gräfenberg Virtual Seismic Array in the secondary microseism frequency band (0.1 to 0.3 Hz), where wind-driven ocean waves interact with the solid Earth, resulting in continuous seismic noise detectable on land (Longuet-Higgins & Jeffreys, 1950; Hasselmann, 1963; Ardhuin et al., 2019). By using beamforming techniques, we analyze the noise signals recorded in the given frequency band to identify and differentiate the dominant wave type regimes present in the seismic noise field. In the following, we describe the data used, the network architecture and compare beamforming results on original and predicted recordings.

2 Data and Methods

2.1 Seismic data and beamforming

We train the neural networks with data from the Gräfenberg seismic array (GRF), which consists of 13 seismic broadband stations (Harjes et al., 1977). The array is located in the Franconian Jura in central Bavaria, Germany, extending approximately 100 kilometers north-south and 40 kilometers east-west (Fig. 1a). We select two time frames for analysis, each consisting of two days of data from all array stations, with one frame corresponding to summer (July 2013) and the other to winter (November 2013), chosen to avoid earthquakes. To prepare the data for the neural network training, we remove the instrument response, detrend, and demean. Seismograms are filtered in the secondary microseism frequency band using a Butterworth bandpass filter from 0.1 Hz to 0.25 Hz. Finally, we resample the data to 20 Hz.

Beamforming enables the extraction of propagation characteristics of seismic waves by analyzing the waveforms recorded across the array (Rost & Thomas, 2002; Ruigrok et

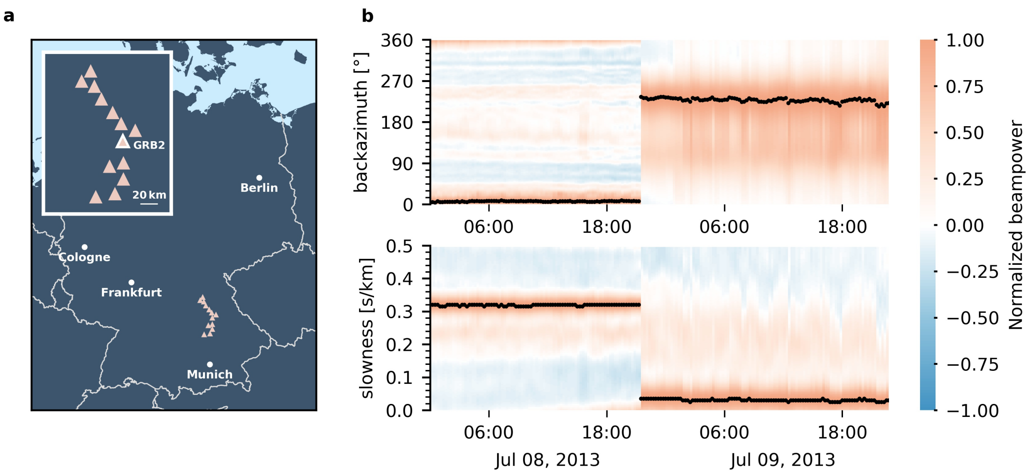


Figure 1. Gräfenberg array beamforming. **a** Map of Germany showing the location of the Gräfenberg (GRF) array, with seismic stations indicated by orange triangles. The reference station GRB2 is indicated with a white frame. An inset in the top-left corner provides a zoomed-in view of the station arrangement within the array. **b** Beamforming results for the selected two-day time period during summer. Colors indicate the normalized beampower in each slice along the best-fitting backazimuth and slowness dimensions. The best-fitting backazimuth and slowness are indicated with a black dot.

al., 2017). We use cross-correlation beamforming, which applies the delay-and-sum approach (Rost & Thomas, 2002) to correlation functions in order to estimate the dominant direction of arrival (backazimuth) and slowness. This method assumes plane waves propagating across the array and is closely connected to Bartlett beamforming (Baggeroer et al., 1988). Both, backazimuth and slowness, not only provide important insights into the seismic waves being analyzed, but, in this study, are the quantities we use to validate the quality of the Virtual Seismic Array.

We apply beamforming to the original GRF recordings using 1-hour windows with 75% overlap. It is important to note that station GRB2 serves as the reference station and is therefore excluded in the beamforming. Figure 1b shows backazimuth and slowness for the selected two-day summer period. Background colors are slices through the slowness domain normalized by beampower, highlighting the best-fitting backazimuth and slowness with a black dot. During the first 22 hours, we observe waves from the north, with a single dominant backazimuth of 7°, measured clockwise from North, and slowness of 0.32 s/km (Fig. 1b). We call this the surface wave-dominated regime due to the presence of Rayleigh waves in this frequency range (Juretzek & Hadziioannou, 2016). We further observe a stark transition to waves arriving from the southwest, with backazimuth and slowness values of 235° and 0.03 s/km, respectively, indicating the transition to a body wave-dominated regime (Landès et al., 2010; Pedersen & Colombi, 2018; Y. Lu et al., 2022; Zhang et al., 2023). These are the regimes we refer back to later in the text.

2.2 Waveform prediction

To obtain the models that predict seismic waveforms and together constitute the Virtual Seismic Array, we build on the approach introduced by Klinge et al. (2025). By learning the transfer functions between a reference station and all other stations within a given seismic array, we obtain unique models that capture the transfer characteristics for each station pair individually. This allows modeling data at each station even if they are no longer in operation, provided that the reference station is still installed and running. We select GRB2 as the reference station because it is located near the center of the array (Fig. 2a).

Klinge et al. (2025) used an encoder-decoder network to learn the transfer properties between two seismic stations. We use the same network architecture with minor changes to account for the different sampling frequency and frequency band. The approach involves feeding input data from a seismic reference station to the network, with the aim of learning the transfer to target data from a neighboring seismic station. As a result, the network generates predictions that ideally approximate the waveforms of the target data. We demonstrate that this methodology is applicable not only to the original study's data but also to the GRF array. While the original study involved an array with interstation distances of hundreds of meters (about seven wavelengths) and varying sources, like oil pumps, at frequencies below 10 Hz, we now apply the methodology to the GRF array with tens of kilometers between stations and in the frequency range of 0.1–0.25 Hz. Figure 2b illustrates example results from training the network with GRF array data, comparing target data and predictions for each station alongside the corresponding correlation coefficient (CC) for quality assessment. While amplitude predictions show variability, resulting in over- or underestimation, phase information is consistently well predicted, which is essential for effective beamforming applications. Although our average CC values are lower than those reported by Klinge et al. (2025), the maximum CC value we achieve is comparable to the results, highlights that the algorithm has the potential to capture key aspects of wave propagation in the GRF data as well.

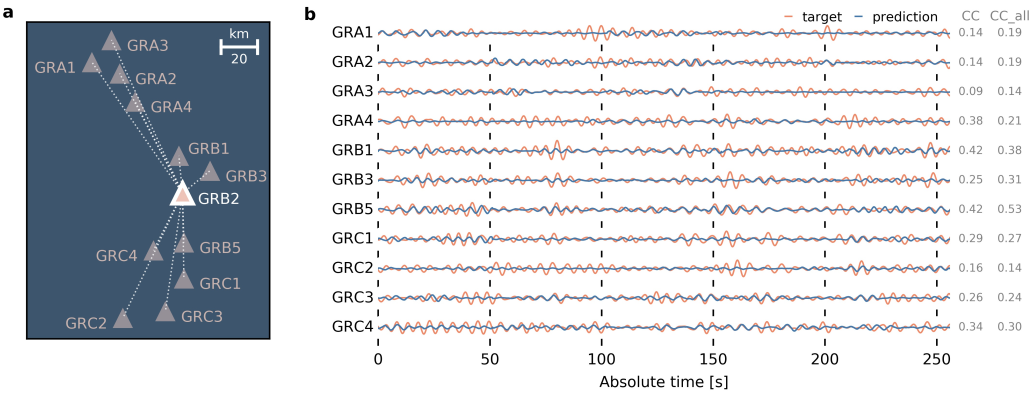


Figure 2. The Virtual Seismic Array. **a** The arrangement of stations within the Virtual GRF Array, highlighting the reference station GRB2 with a bold white outline, while the virtual stations are displayed in faded orange. Dashed lines illustrate the connections from the reference station to each of the array stations, forming station pairs for model training. **b** A selection of example data, featuring the target time series, i.e., the original recording, for each station (orange line) and the corresponding model predictions (blue line). The normalized correlation coefficient is provided on the right side for the example trace depicted (CC) as well as for the entire target data (CC_all), indicating the degree of similarity between the two time series.

Based on this methodology we perform the network training for every station combination with the reference station GRB2. Before training, the data are scaled with a combination of standard scaling and normalization. We allocate 80% of the data to the training set and 20% to the testing set. The testing set consists of data that the model has not seen during training (e.g. target data in Fig. 2b), allowing us to evaluate how well the algorithm generalizes to new data. The performance of the models on the testing data of each station represents the Virtual Seismic Array, enabling the prediction of seismic data across the array even if stations encounter failures or downtimes.

3 Performance of the Virtual Seismic Array

We evaluate the performance of the Virtual Seismic Array across three different scenarios of increasing complexity. First, we analyze a single dominant noise regime, characterized by surface waves only for both the training and testing data. Next, we evaluate its adaptability to a changing regime by training the models with data that transition from surface wave-dominance to body wave-dominance (Fig. 1b). Finally, we assess the performance for an unseen regime, where surface waves dominate in training but body waves dominate in testing. In the following, ”real array” refers to the original data, where all stations are still active. Our predictions constitute the Virtual Array.

3.1 A single dominant regime

For a single dominant regime (Fig. 3) we find that backazimuth and slowness detected by the real array in the test data (Fig. 3c,d) closely match those in the training set (Fig. 3a,b). The Virtual Array detects the same dominant backazimuth and slowness (Fig. 3e,f). Both the real and Virtual Array find surface waves incoming from North. The predicted slowness closely matches the original measurements, showing precise and focused detections.

The strong correspondence between the performance of the real array and Virtual Array demonstrates the proposed method effectively learns and predicts relevant seismic features in the data, despite low correlation coefficients. This highlights the algorithm's ability to capture the transfer characteristics between each station pair during training and to apply this knowledge to unseen data, showing its robustness with less than two days of training data. Our findings indicate the algorithm performs particularly well when the dominant noise regime is stable and aligns with the training dataset. Given that the model was trained on and applied to a single type of wave regime, we anticipated good performance. The question remains whether the Virtual Array can achieve similar predictive accuracy in more complex scenarios.

3.2 A changing regime

We investigate a more complex scenario, where the dominant regime changes (Fig. 4). Here, the training set contains a transition from surface to body wave-dominated regime, indicated by backazimuths and slownesses. We show the training set twice (Fig. 4a,b & g,h) to visually emphasize the application to two different testing datasets. First, we apply the models to body wave-dominated test data from the summer period post training set (Fig. 4c-f). Second, we evaluate the performance of the models on surface wave-dominated test data from the winter period (Fig. 4i-l). This cross-application helps to

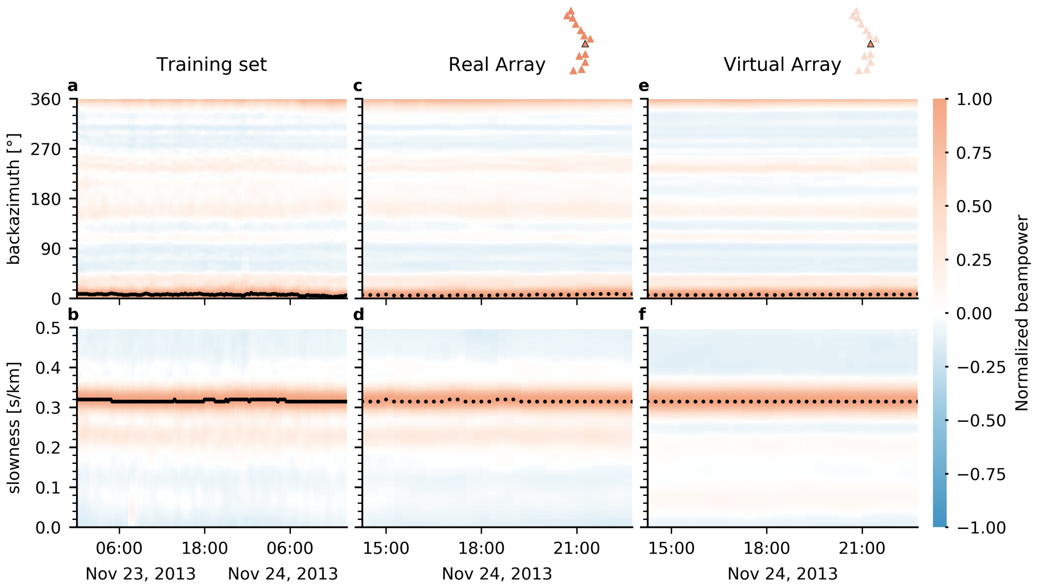


Figure 3. Beamforming results for the single dominant regime, with backazimuth in the top row and slowness in the bottom row. Panels **a** and **b** show results for the training set, which includes the first 80% of the two-day winter data dominated by surface waves. Panels **c** and **d** highlight results from the real array as ground truth, representing the last 20% of the two-day winter data, while panels **e** and **f** illustrate findings for the Virtual Array. Each plot includes background slices through the slowness domain, where blue indicate negative correlation and red indicate positive correlation. Black dots mark the maximum beampower, representing best-fitting waves.

investigate the algorithm's ability to generalize and precisely capture seismic wave behavior under varying conditions.

In the first case, beamforming the Virtual Array reveals body waves arriving from southwest (Fig. 4e,f), similar to the direction observed by the real array (Fig. 4c,d). While the Virtual Array shows a very stable distribution of beampower values with time (Fig. 4e), the real array results deviate slightly from the average (Fig. 4c). In this example, the real array (Fig. 4c,d) detects less well-focused beampowers compared to those in the single dominant regime (Fig. 3), likely due to the presence of more complex wavefields and lower resolution at low slownesses. In contrast, the Virtual Array (Fig. 4e,f) finds remarkably sharp detections.

In the second case, beamforming the Virtual Array (Fig. 4k,l) detects predominantly surface waves coming from the north, which aligns well with detections by the real array (Fig. 4i,j). The observed values are in line with the surface wave-dominated regime seen in the first 22 hours of the training set (Fig. 4g,h), although body waves dominate the second part of the training set. For both the real array and Virtual Array, the beampowers are sharply focused around the maxima.

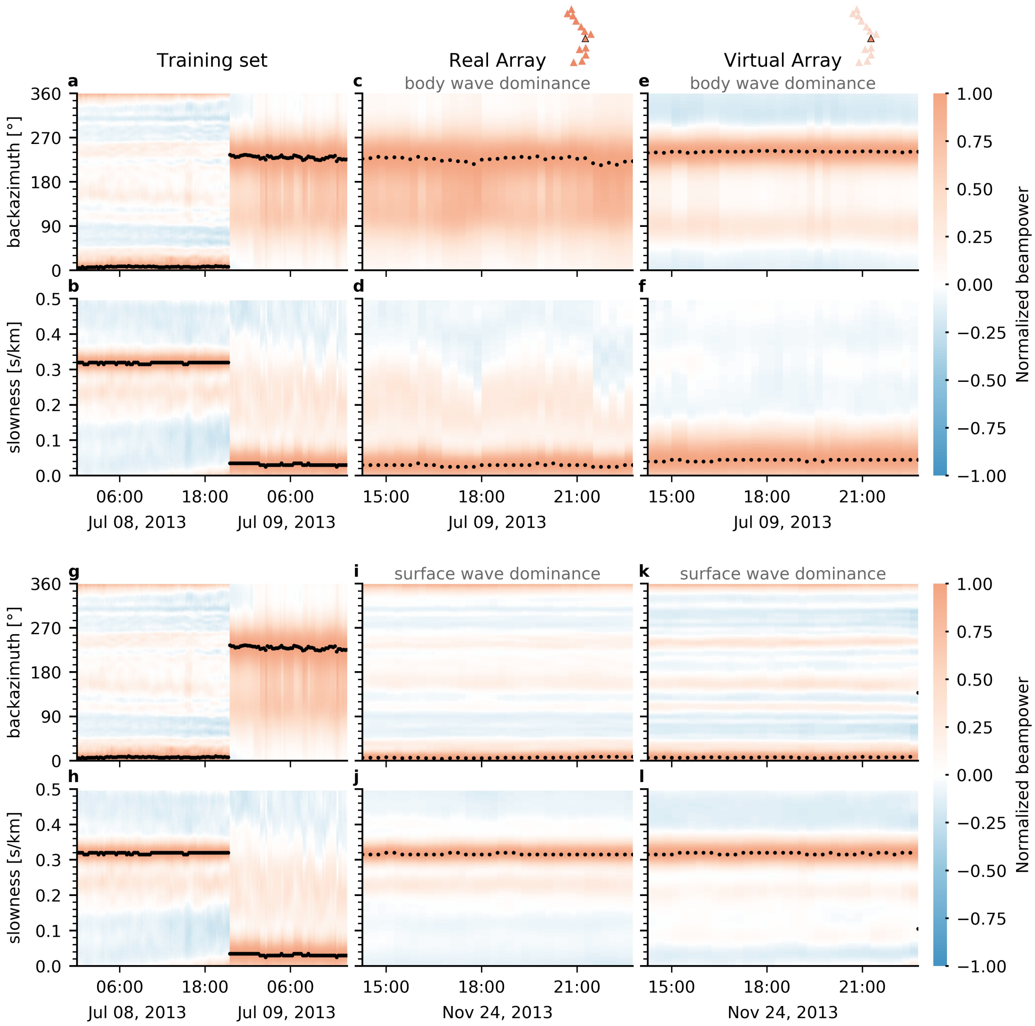


Figure 4. Beamforming results for the changing regime. The training set consists of the first 80% of the two-day summer data dominated by surface and body waves (**a**, **b**, **g**, and **h**). Panels **c** and **d**: Results from the real array, displaying the last 20% of the summer data, which is body wave-dominated, while panels **i** and **j** feature data from the real array during winter with a surface wave-dominance. Outcomes for the respective Virtual Arrays are illustrated in panels **e**, **f** and **k**, **l** . For further description of the plot see caption of Figure 3.

These two scenarios show that the algorithm is able to predict the wavefield from a single-station recording as long as the wavefield regime, i.e., dominant wave type and direction, has been part of the training. This highlights the algorithm's ability to generalize to more complex training sets compared to the single dominant regime while effectively differentiating between two distinct regimes. The sharper beampower focus for body waves detected by the Virtual Array compared to the real array supports this further (Fig. 4a-f). Although the algorithm was trained on both body and surface wave-dominated regimes, in the first case, the real array is dominated by body waves only, while in the second case, it is dominated by surface waves only. The models predict a simpler dominant wavefield compared to the original training set, resulting in sharper predictions and enhanced predictive accuracy.

Note that each time window is predicted independently from its neighboring time windows so that the prediction for a beamforming window is not affected by previous or later predictions. The algorithm furthermore successfully predicts the correct wavefield across seasons, with models trained on data from summer being applied to winter data. This demonstrates its ability to generalize across seasonal variations of noise.

3.3 An unseen regime

We demonstrate the main limitations of this approach by evaluating the algorithm's performance when encountering an unseen wavefield regime. The training set consists mainly of surface waves arriving from the north (Fig. 5a,b). Meanwhile, the original recordings of the real array are dominated by body waves arriving from the southwest (Fig. 5c,d). The Virtual Array is unable to predict the wavefield seen by the real array (Fig. 5e,f). Instead, it predicts surface waves arriving from north, the only wavefield regime it was trained on. This misalignment indicates a lack of generalization, suggesting that the models cannot adapt to the conditions present in the real array.

The algorithm fails to make accurate predictions when faced with a wavefield regime that was not included in the training set, as is the case for the unseen regime. The training set predominantly includes surface waves, which do not match the characteristics present in the real array, which are predominantly body waves. These limited generalization capabilities highlight that the algorithms effectiveness in making accurate predictions relies on the characteristics it encounters during the training process. However, when sim-

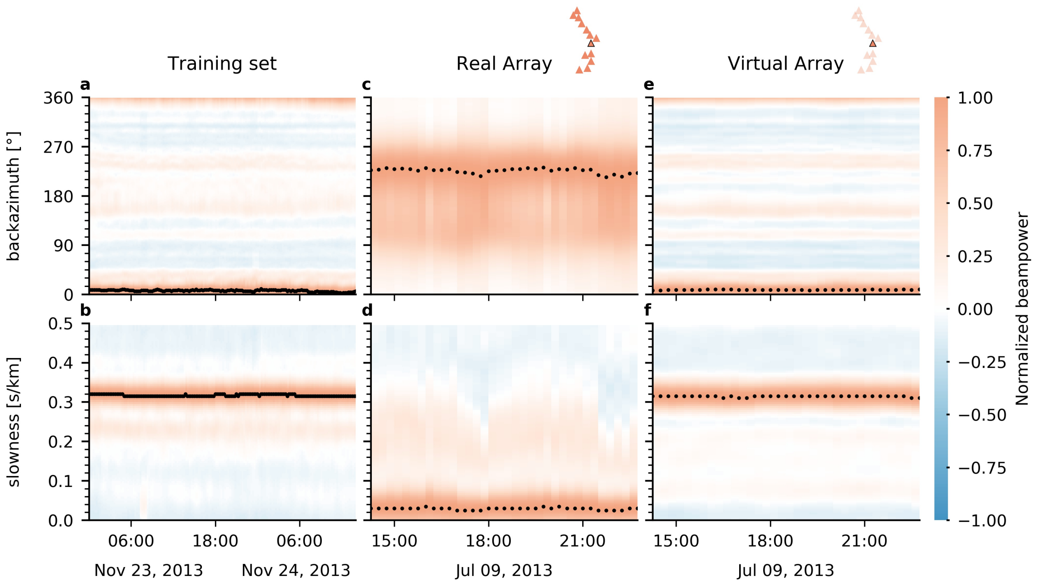


Figure 5. Beamforming results for the unseen regime. Panels **a** and **b** show the training set consisting of the first 80% of the two-day winter data dominated by surface waves. In panels **c** and **d**, results from the real array are displayed, showing body wave-dominance during the summer period. The outcomes of the Virtual Array are illustrated in panels **e** and **f**. For further description of the plot see caption of Figure 3.

ilar characteristics are present in the training set, as demonstrated in the single and the changing regime, the beamforming results for the Virtual Array are of high accuracy (Figs. 3,4). Furthermore, it is likely that the models can only predict cases that have been encountered frequently during the training process. For earthquakes, which are intentionally excluded in the selected data, accurate predictions would therefore require training on datasets that include many examples. This limitation of encoder-decoder models for earthquake recordings has been reported before (Mousavi et al., 2020; Yin et al., 2022; Zlydenko et al., 2023). For our study, this underscores the importance of having a diverse training set that covers various regimes to enhance the algorithm's ability to generalize and make accurate predictions on different data characteristics. Therefore, we chose ocean microseism noise for this first demonstration of Virtual Seismic Arrays, as it is particularly well-suited due to its stability over days and weeks (Ardhuin et al., 2019).

4 Applications and Outlook

Our findings demonstrate that Virtual Seismic Arrays can work and deliver promising wavefield predictions. As a result, several potential applications emerge, especially in remote areas where seismic array deployment can be challenging. By training models capable of predicting waveforms even without physical sensors, we can reduce the need for continuously deploying all array stations. For example, data from previous short-term deployments could be used as the training set, enabling ongoing predictions in the area of interest and maintaining data coverage without physical deployment of the full array. New deployments can be planned that involve using a minimal number of stations initially and repositioning those stations step by step to achieve full regional coverage over time while reducing the amount of resources needed. On an operational level, this approach also allows to compensate for temporary outages of individual stations. However, to fully realize Virtual Seismic Arrays in production, further steps are necessary. These include evaluating the performance in more complex datasets, such as those with frequent transitions between different wave type regimes and including transient sources such as earthquakes. It is important to integrate our findings with other approaches to enhance the adaptability of this method in different contexts. Additionally, understanding the most suitable conditions for implementing Virtual Seismic Arrays should be accompanied by detailed parameter studies. It is particularly important to understand, why certain station combinations are more effective than others and to assess how data quality and specific hyperparameters influence the model training, which we aim to pursue in the future.

5 Conclusion

In this study, we evaluate the applicability of encoder-decoder networks to implement Virtual Seismic Arrays, which predict data for an entire array using a single reference station. As a proof-of-concept, we train a Gräfenberg Virtual Seismic Array in the secondary microseism frequency band. By leveraging data from a single reference station to predict array recordings for all other stations within the array, we train models that successfully capture the wavefield propagation across the stations. Beamforming the resulting predicted waveforms reveals good agreement between the real and Virtual Array when the dominant wave regime encountered is included in the training set. This highlights the effectiveness of our approach in capturing underlying wave dynamics and the potential for future applications of Virtual Seismic Arrays. We propose to expand the application of this framework to diverse regions and seismic conditions, unlocking its potential to significantly enhance approaches for measuring and analyzing seismic data, particularly in challenging, remote areas. Temporary array deployments, for instance, can enable long-term "virtual operation" that allows seismic monitoring to continue when the physical array is unavailable. We test the most extreme version of a Virtual Array, where all but one station are removed and show that this approach can effectively compensate temporarily unavailable stations. This cost-effective advancement offers a promising approach for improving data availability and has the potential to substantially improve the reliability and efficiency of our global seismic monitoring capabilities.

Data Availability Statement

We use publicly available seismograms provided by the German Regional Seismic Network (GR) operators (Federal Institute for Geosciences and Natural Resources, 1976), accessed via the ORFEUS European Integrated Data Center (EIDA). We use accessible colors (Crameri, 2023; Tol, 2025).

Acknowledgements

The authors thank the BGR for seismic data access. This work is financially supported by the Federal Ministry of Education and Research (BMBF) project "3G-GWD" with references 05A20GU5 and 05A23GU5 and partially funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 955515 (SPIN ITN - https://spin-itn.eu).

References

Alhukail, I. (2012). *The Concept of Virtual Arrays in Seismic Data Acquisition* (Unpublished doctoral dissertation). Texas A&M University.

Ardhuin, F., Gualtieri, L., & Stutzmann, E. (2019, March). Physics of Ambient Noise Generation by Ocean Waves. In (pp. 69–108). doi: 10.1017/9781108264808.005

Baggeroer, A. B., Kuperman, W. A., & Schmidt, H. (1988, February). Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem. *The Journal of the Acoustical Society of America*, *83*(2), 571–587. doi: 10.1121/1.396151

Capon, J., Greenfield, R., & Kolker, R. (1967). Multidimensional maximum-likelihood processing of a large aperture seismic array. *Proceedings of the IEEE*, *55*(2), 192–211. doi: 10.1109/PROC.1967.5439

Crameri, F. (2023). Scientific colour maps. Retrieved from `https://doi.org/10.5281/zenodo.8409685`

Federal Institute for Geosciences and Natural Resources. (1976). *German Regional Seismic Network (GRSN).* Bundesanstalt für Geowissenschaften und Rohstoffe. doi: 10.25928/MBX6-HR74

Gelboim, M., Adler, A., Sun, Y., & Araya-Polo, M. (2023). Encoder–Decoder Architecture for 3D Seismic Inversion. *Sensors*, *23*(1). doi: 10.3390/s23010061

Gibbons, S. J., & Ringdal, F. (2006, April). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, *165*(1), 149–166. doi: 10.1111/j.1365-246X.2006.02865.x

Harjes, H.-P., Seidl, D., & others. (1977). Digital recording and analysis of broadband seismic data at the Gräfenberg (GRF) array. *Journal of Geophysics*, *44*(1), 511–523.

Hasselmann, K. (1963). A statistical analysis of the generation of microseisms. *Reviews of Geophysics*, *1*(2), 177–210. doi: https://doi.org/10.1029/RG001i002p00177

Holfort, I. K., Gran, F., & Jensen, J. A. (2009). Broadband minimum variance beamforming for ultrasound imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *56*(2), 314–325. doi: 10.1109/TUFFC.2009.1040

Juretzek, C., & Hadziioannou, C. (2016). Where do ocean microseisms come from? A study of Love-to-Rayleigh wave ratios. *Journal of Geophysical Research: Solid Earth*, *121*(9), 6741–6756. doi: https://doi.org/10.1002/2016JB013017

Klinge, J., Schippkus, S., Walda, J., Hadziioannou, C., & Gajewski, D. (2025, January). Predictive modeling of seismic wave fields: learning the transfer function using encoder-decoder networks. *Geophysical Journal International*, ggaf004. doi: 10.1093/gji/ggaf004

Landès, M., Hubans, F., Shapiro, N. M., Paul, A., & Campillo, M. (2010). Origin of deep ocean microseisms by using teleseismic body waves. *Journal of*

Geophysical Research: Solid Earth, *115*(B5). doi: https://doi.org/10.1029/2009JB006918

Li, S., Yang, C., Sun, H., & Zhang, H. (2019, March). Seismic fault detection using an encoder–decoder convolutional neural network with a small training set. *Journal of Geophysics and Engineering*, *16*(1), 175–189. doi: 10.1093/jge/gxy015

Li, W., Chakraborty, M., Fenner, D., Faber, J., Zhou, K., Rümpker, G., … Srivastava, N. (2022). EPick: Attention-based multi-scale UNet for earthquake detection and seismic phase picking. *Frontiers in Earth Science*, *10*. doi: 10.3389/feart.2022.953007

Longuet-Higgins, M. S., & Jeffreys, H. (1950). A theory of the origin of microseisms. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *243*(857), 1–35. doi: 10.1098/rsta.1950.0012

Lu, J. Y., Zou, H., & Greenleaf, J. F. (1994). Biomedical ultrasound beam forming. *Ultrasound in medicine & biology*, *20*(5), 403–428. doi: 10.1016/0301-5629(94)90097-3

Lu, Y., Pedersen, H. A., Stehly, L., & Group, A. W. (2022). Mapping the seismic noise field in europe: spatio-temporal variations in wavefield composition and noise source contributions. *Geophysical Journal International*, *228*(1), 171–192.

Luijten, B., Cohen, R., de Bruijn, F. J., Schmeitz, H. A. W., Mischi, M., Eldar, Y. C., & van Sloun, R. J. G. (2020). Adaptive Ultrasound Beamforming Using Deep Learning. *IEEE Transactions on Medical Imaging*, *39*(12), 3967–3978. doi: 10.1109/TMI.2020.3008537

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020, August). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, *11*(1), 3952. doi: 10.1038/s41467-020-17591-w

Pedersen, H., & Colombi, A. (2018, August). Body waves from a single source area observed in noise correlations at arrival times of reflections from the 410 discontinuity. *Geophysical Journal International*, *214*(2), 1125–1135. doi: 10.1093/gji/ggy191

Rost, S., & Thomas, C. (2002). ARRAY SEISMOLOGY: METHODS AND APPLICATIONS. *Reviews of Geophysics*, *40*(3), 2–1–2–27. doi: https://doi.org/10.1029/2000RG000100

Rost, S., & Thomas, C. (2009, October). Improving Seismic Resolution Through Array Processing Techniques. *Surveys in Geophysics*, *30*(4), 271–299. doi: 10.1007/s10712-009-9070-6

Ruigrok, E., Gibbons, S., & Wapenaar, K. (2017, May). Cross-correlation beamforming. *Journal of Seismology*, *21*(3), 495–508. doi: 10.1007/s10950-016-9612-6

Schweitzer, J., Fyen, J., Mykkeltveit, S., Gibbons, S. J., Pirli, M., Kühn, D., & Kværna, T. (2012). Seismic arrays. In *New manual of seismological observatory practice 2 (NMSOP-2)* (pp. 1–80). Deutsches GeoForschungsZentrum GFZ.

Tol, P. (2025). Paul tol's notes, colour schemes and templates. Retrieved from `https://personal.sron.nl/~pault/data/colourschemes.pdf.`

van der Veen, A.-J., Leshem, A., & Boonstra, A.-J. (2004). Signal processing for radio astronomical arrays. In *Processing Workshop Proceedings, 2004 Sensor Array and Multichannel Signal* (pp. 1–10). doi: 10.1109/SAM.2004.1502901

Wang, K., Lu, L., Maupin, V., Ding, Z., Zheng, C., & Zhong, S. (2020). Surface Wave Tomography of Northeastern Tibetan Plateau Using Beamforming of Seismic Noise at a Dense Array. *Journal of Geophysical Research: Solid Earth*, *125*(4), e2019JB018416. doi: https://doi.org/10.1029/2019JB018416

Warnick, K. F., Maaskant, R., Ivashina, M. V., Davidson, D. B., & Jeffs, B. D. (2016). High-Sensitivity Phased Array Receivers for Radio Astronomy. *Proceedings of the IEEE*, *104*(3), 607–622. doi: 10.1109/JPROC.2015.2491886

Yin, J., Denolle, M. A., & He, B. (2022, July). A multitask encoder–decoder to separate earthquake and ambient noise signal in seismograms. *Geophysical Journal International*, *231*(3), 1806–1822. doi: 10.1093/gji/ggac290

Zhang, R., Boué, P., Campillo, M., & Ma, J. (2023, March). Quantifying P-wave secondary microseisms events: a comparison of observed and modelled backprojection. *Geophysical Journal International*, *234*(2), 933–947. doi: 10.1093/gji/ggad103

Zlydenko, O., Elidan, G., Hassidim, A., Kukliansky, D., Matias, Y., Meade, B., . . .

Bar-Sinai, Y. (2023, July). A neural encoder for earthquake rate forecasting. *Scientific Reports*, *13*(1), 12350. doi: 10.1038/s41598-023-38033-9

4. Optimizing network performance – A parameter study

In this section, the influence of different parameters on the performance of the encoder-decoder network introduced in **Study I** is investigated. To understand and optimize its predictive capabilities, an analysis of how specific parameters affect the results is conducted, gaining insights into the algorithms behaviour under specific conditions. While a comprehensive testing of individual parameters and their interactions is beyond the scope of this study, the analyses offer some initial insights into the current strengths and limitations of the approach. In the following, it will be investigated how the network performance depends on interstation distance and frequency, how data scaling affects the results, and how the network responds to parameters such as network depth, number of training epochs, final activation functions, and information content. Therefore, the corresponding parameters are adapted in the network and model training is performed for each configuration, which allows to evaluate how these adjustments influence the network's behaviour and thus the results.

4.1. Parameter testing

This parameter study builds on the principles introduced in **Study I** where an encoder-decoder architecture with a depth of five and 1500 training epochs is developed and used. The focus is again on the Vienna Basin, which is characterized by seismic velocities that approximate acoustic velocities of 300 m/s or below and the presence of several noise sources, such as roads, wind turbines and oil pumps (Schippkus et al., 2020). To allow for direct comparison, the primary focus is on a station pair previously investigated in **Study I**, situated towards the center of the array. In order to evaluate the influence of the interstation distance on the results, nearby stations are further considered in ascending order of their distance

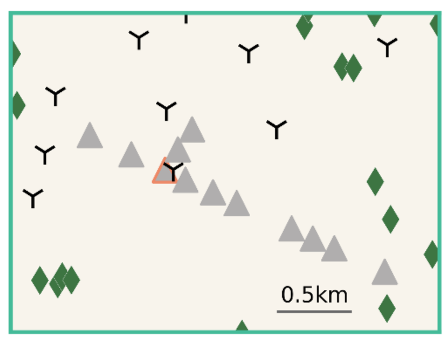


Figure 4.1. Map view of selected stations (grey triangles). The reference station is outlined in orange. The surrounding region includes seismic sources like oil pumps (green) and wind turbines (black).

from the reference station. Figure 4.1 shows the locations of the selected reference station that was earlier part of the analysis in **Study I**, its proximity to other stations, and the surrounding region including nearby sources. The seismic data used in this study are filtered below 10 Hz, similarly to **Study I**.

Interstation distance

In order to interpret seismic measurements and optimize the performance of machine learning models such as the network, it is important to understand the propagation behaviour of seismic waves. As they travel through the Earth, seismic waves interact with the medium they pass through, which changes their characteristics like frequency, amplitude, and velocity (Bormann and Wielandt, 2013; Haendel et al., 2018). Thereby, local geology and processes like attenuation, scattering, and frequency-dependency alter both the wave's properties and the data that can be extracted from it (Haendel et al., 2018). Interstation distance is therefore important to analyse, because it influences the degree to which different seismic stations record coherent seismic signals.

The effect of interstation distance on model performance is investigated by systematically increasing the distance between station pairs. Utilizing a fixed reference station, it is paired with other stations at greater distances. This approach allows observing and analysing how the quality of the model's prediction changes as the interstation distance grows. To interpret and compare the results, two evaluation metrics are used: Root Mean Square Error (RMSE) and correlation coefficient (CC). By using both metrics, the model's performance is assessed in terms of both amplitude prediction (RMSE) and phase alignment (CC). Thereby, lower RMSE values indicate better amplitude prediction, while a CC closer to one shows good alignment of the phases.

Figure 4.2 displays the relationship between model performance and interstation distance for RMSE (Figure 4.2, top) and CC (Figure 4.2, bottom). The reference station is at zero meters distance, while station locations of target stations are indicated by grey triangles on both plots. The analysis reveals that the predicted and observed waveforms have a strong positive correlation at the shortest distance of 165 meters, with a CC of approximately 0.75. With increasing interstation distance, there is a decay in CC values, with a CC dropping to 0.22 at 570 meters and further decrease to a CC of 0.05 at 600 meters. For all distances beyond 600 meters and up to 1800 meters, the CC reaches zero or even slightly goes negative, suggesting no relevant phase correlation. Due to the lack of station coverage in this range, it is anticipated that important developments in the data that could influence the understanding of the relationship between interstation distance and model performance may be missed. In contrast to the CC plot, the RMSE plot does not show a trend with distance. The RMSE values for interstation distances up to 600 meters range from 0.08 to 0.13, while the RMSE values for distances above 1000 meters primarily cluster around 0.06, with one exception at 1040 meters. In conclusion, both plots show distinct behaviour up to 600 meters, which is equivalent to about 20 wavelengths at a velocity of 300 m/s and a maximum frequency of 10 Hz, and then clearly change for distances greater than 1000 meters, approximately 33 wavelengths.

The trends found in the CC and RMSE plots show some interesting details of the model performance with increasing interstation distance in the observed frequency range up to 10 Hz. The decay in Correlation Coefficients (CCs) with distance is attributed to several factors related to wave propagation and the algorithms learning process. As seismic noise travels through the Earth, it attenuates at varying speeds because it is composed of distinct signals with a broad spectrum of frequencies. For instance, higher frequency components attenuate more quickly over distance than lower frequency components, which leads to a loss of signal coherence (Bormann and Wielandt, 2013). Because of scattering, reflection, and refraction, seismic waves that travel farther also encounter more geological heterogeneities, which further mitigates correlation (Schimmel and Gallart, 2007). Based on these wave propagation effects and the influence of local site conditions, the transfer functions from the reference station to each of the other stations become more complex for increasing distances. This makes it more difficult for the machine learning algorithm to learn the transfer function and explains why CC values decrease with distance. The effects of sampling rate and signal periods on the results will further be investigated later in this work, which may also explain why the models perform worse the farther they are from the reference station.

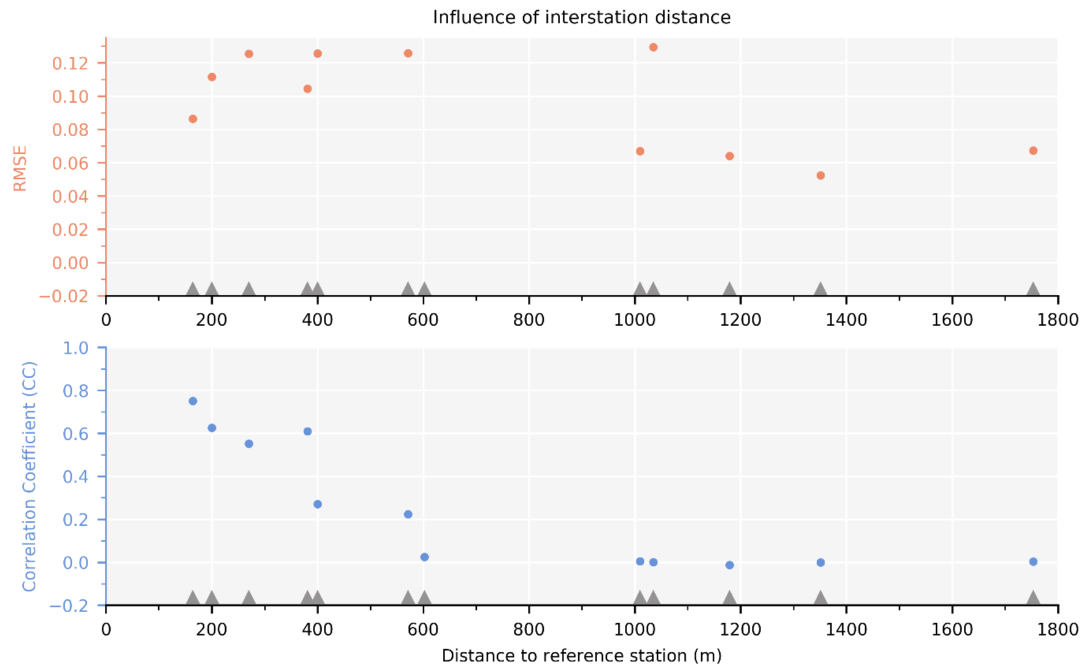


Figure 4.2. Relationship between model performance and interstation distance. The top panel shows the Root Mean Square Error (RMSE) against distance, while the bottom panel displays the correlation coefficient (CC) against distance. The reference station is at zero meters distance, and grey triangles on both plots mark target station locations.

In contrast to the CC trend, the RMSE shows a decrease of values with greater distances, indicating a better amplitude fit, whereas the CC values suggest a poorer phase fit with increasing distance. This apparent improvement is, in fact, a misleading artefact of the models output. The algorithm predicts constant zeros as distance increases because it is unable to capture wave propagation effects. This leads to smaller RMSE values because the differences between the actual signal and the zero prediction cancel each other out. While the algorithm fails to capture the transfer function between station pairs at larger distances, the findings in the lower distance range are promising, yet they indicate amplitude differences present in the data. This is consistent with the findings from **Study I** and **Study II,** showing the need for additional research to determine the cause of this and being able to modify the algorithm to address this issue. To gain a better understanding of the algorithms performance across smaller frequency ranges, its performance in discrete frequency bands will be analysed in the following.

Frequency content

This frequency band analysis, more detailed than the previous broadband analysis, allows investigating how the model performs in specific frequency bands that were previously not analysed separately. This helps to identify which components of the seismic noise field are most predictable by the models in this specific dataset and provides insights on the variations in performance across different frequencies and interstation distances. To do so, the model is trained on 10 Hz lowpass filtered data and additional filters are applied afterwards to determine the model's performance in six specific frequency ranges: 0-1 Hz, 1-2 Hz, 2-3 Hz, 3-5 Hz, 5-7 Hz, and 7-10 Hz. The CC is estimated in the respective frequency band for each target station and plotted against the distance to the reference station (Figure 4.3).

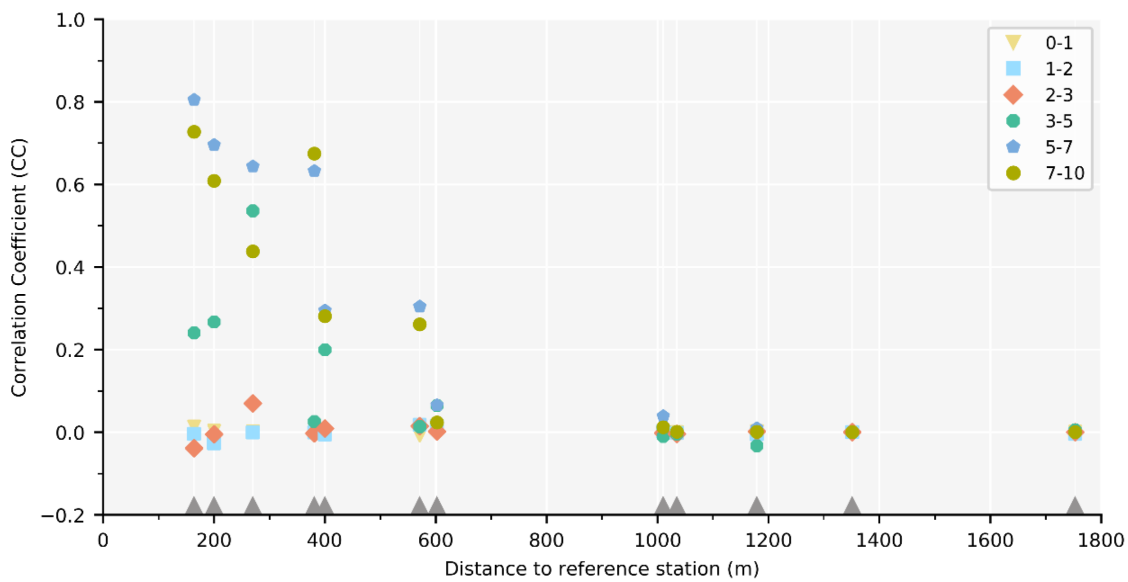


Figure 4.3. Frequency-dependent model performance across interstation distance. Shown are correlation coefficients for discrete frequency bands (0-1 Hz, 1-2 Hz, 2-3 Hz, 3-5 Hz, and 7-10 Hz), each represented by a distinct coloured marker.

These frequency-dependent CCs are shown as coloured markers across various distance ranges in Figure 4.3. The analysis shows that the CC values for frequencies between 0 and 3 Hz are approximately zero at all distances, indicating that the model performs poorly in this lower frequency ranges. On the other hand, frequencies higher than 3 Hz demonstrate superior model performance, particularly for station pairs that are up to 600 meters apart. Here, the 5-7 Hz and 7-10 Hz ranges show the strongest results, with CC values almost consistently above 0.6 up to 380 meters, then decreasing to around 0.30 between 400 meters and 570 meters, before dropping close to zero at 600 meters. This overall decrease again follows a nearly linear trend, with an exception at 380m. The 3-5 Hz range further shows positive CC values up to 400 meters, though generally lower than the higher frequency ranges. Its maximum CC reaches 0.54, but values mostly find around 0.20 or towards zero. Beyond 600 meters, CC values for frequencies above 3 Hz also converge towards zero, indicating a general decline in model performance at greater distances, similar to what is observed for lower frequency bands.

These observations provide information about the performance of the algorithm and the seismic noise field, which is influenced by various sources and processes. Particularly at shorter distances, the model shows higher predictability in the 5-10 Hz range. This might be due to the presence of local sources, such as the nearby wind turbine, that emit signals not only with distinct patterns (Neuffer et al., 2019), but also with strong amplitudes at higher frequencies. This makes these signals more prominent and thus easier for the model to learn, which consequently leads to higher accuracy in these ranges. While signals closer to the station are affected less by interference, the overlap of multiple sources with increasing distance might contribute to a more complex wave field the farther away from the reference station. This makes accurate predictions more difficult. Further considering the dominant seismic velocity in the study area (approximately 300 m/s), the performance drop at 600 meters occurs after about 10-20 wavelengths for the 5-10 Hz frequency range. This could be a result of seismic signals becoming more complex and harder to predict because of attenuation and scattering with distance. Moreover, due to the lack of station coverage between 600 and 100 meters, it is necessary to anticipate the possibility of missing important transitions in model performance here.

The frequency spectrum of the local seismic noise field is probably the reason for poor performance at lower frequencies (0-3 Hz) over all distances. Although this has not been analysed in detail, a predominance of higher frequency noise could affect the algorithm's ability to learn the characteristics of lower frequency data, potentially reducing its performance in this range.

This thesis supports by Schippkus et al. (2020), who examined PSDs for various noise sources (wind turbines, oil pumps, roads, railways) in the same region. Their study showed that the power at frequencies around 10 Hz is significantly greater than at frequencies around 1 Hz. As mentioned earlier, this explains why the algorithm performs better at higher frequencies. Furthermore, the data used in this study are measured using geophones, which may contribute to the algorithms reduced performance as their sensitivity decreases towards lower frequencies (Dean and Grant, 2024).

Seismic wave attenuation and the complexity of the wave field over longer distances are likely the reason why the models performance decreases with distance, particularly for higher frequencies. This decay highlights the predictive limitations for this specific dataset. Due to the lack of station coverage between 600 and 1800 meters, the model's performance cannot accurately assessed in this range, which makes it hard to draw conclusions on the general behaviour of the model across different frequency ranges at these distances. However, the same algorithm successfully predicts lower frequency data in the range of 0.1-0.25 Hz, as shown in **Study II** with seismometer data from the GRF array. This suggests that rather than fundamental limitations of the model itself, the observed limitations are probably caused by the dataset's characteristics, such as instrumentation, local geology, and noise sources.

Data scaling

Data pre-processing is one of the most important steps in developing successful machine learning models (Kotsiantis et al., 2006). Data scaling, which is part of this process, makes sure all features are on a similar scale and contribute equally to the model's learning. This has the potential to receive significant improvements in model performance and training effectiveness (Ahsan et al., 2021; Sharma, 2022), ranging from basic linear regression to more complex neural network tasks. This subsection explores how different scaling approaches affect the performance of the encoder-decoder network. The aim is to show how different scalings impact the results of the machine learning method and highlight the importance of data scaling in this context.

The data in the earlier studies **Study I** and **II** were scaled using a combination of standard and min-max scaling. In order to investigate better the effects of different scalings, three additional approaches are tested: no scaling, standard scaling alone, and min-max scaling alone. For this comparison, the focus is on the station pair from **Study I**, training the encoder-decoder network on its data using each of these scaling approaches. For easy comparison, the results are plotted using the same method as in **Study I**, which provides a visual representation of the data as well as the evaluation metrics RMSE and CC to assess the results. It is important to note that RMSE values are not directly comparable across different scaling methods, as they are expressed in different units or scales. However, the CC remains scale-invariant and can be compared directly. In this analysis, the colours around the boxes represent the different scaling methods applied.

No scaling

If there is no scaling applied to the data, the amplitude range of the data remains notably small between [-0.0013, 0.0013]. Observing the network output (Figure 4.4**b**, blue line), the prediction of the network appears as horizontal line. This indicates that the network was unable to learn meaningful features from the unscaled data and fails to capture the underlying structure of the seismic signal. This reflects further by the CC of zero (Figure 4.4**b**), showing no correlation between predicted and actual values. The RMSE is zero here as well, which is misleading, as it results from very small scales of the original data. The histogram plot (Figure 4.4**e**) displaying CC values across the entire test dataset confirms that this poor performance is further consistent across all traces.

MinMax scaling (mm)

The MinMax scaling approach, which normalizes the data based on the minimum and maximum values to a range of [-1, 1], yields significantly different results compared to the no-scaling method (Figure 4.4**f-j**). The prediction closely follows the target data (Figure 4.4**g**), with the blue line aligning well with the orange line in many areas. This alignment is particularly noticeable for the seismic phases, while the amplitudes are generally over- and underestimated. Subplot **h** provide a detailed view of this, showing the distribution of CC values over the example trace. Most values are around 0.5 or higher, which highlights accurate phase fitting along this trace. As seen in the scatter plot **I,** the data cloud is widely dispersed rather than closely condensed along the ideal fit line (purple dashes). This reflects the model's tendency to over- and underestimate amplitudes. However, the models overall CC of 0.64 demonstrates that it can identify important patterns in the data. This is further supported by the histogram plot **j**, which shows the CC values for the entire dataset. CC values are broadly distributed between zero and one, with a gradual increase and peak towards higher values, suggesting generally strong prediction capabilities throughout the dataset.

Standard scaling (std)

With standard scaling applied, which standardizes the data by subtracting the mean and dividing by the standard deviation, the data amplitudes range approximately from [-8, 8]. The prediction (Figure 4.4**l**, blue line) shows a somewhat static pattern, characterized by consistent upward movement followed by small plateaus and subsequent downward movement. These oscillations further tend to reach similar positive and negative amplitudes throughout the trace, which is further visible in the scatter plot **n** that appears as a vertical column. The plot further shows that amplitudes are both over- and underestimated again, as it does not distribute along the best-fit line. Because of this repetitive nature, the phase fit is quite good, reflected in the overall CC of 0.66. Compared to the MinMax scaling, the histogram plot **o** displays a distribution that is spread less across the entire range, but with CC values between zero and one again. Its peak occurs around 0.75, before and after which the distribution drops off steeply. Notably, there are very few instances of CC values close to zero, indicating that the model consistently achieves some level of correlation across the dataset.

Combination (std + mm)

In the combination of MinMax and standard scaling, the amplitudes are again between [-1, 1] due to the mm scaling included. The predictions closely align with the target data (Figure 4.4**q**), which results in the highest CC of 0.75 among all scalings tested. The histogram plot **t** also shows this, with no values near zero. Starting from around 0.50 the distribution peaks at approximately 0.90, indicating consistently strong correlations across the dataset. While some amplitude mismatch persists, it is fewer pronounced than in previous examples. The scatter plot **s** supports this with a cloud of points distributed along the best-fit line, though deviations in both positive and negative directions are still present.

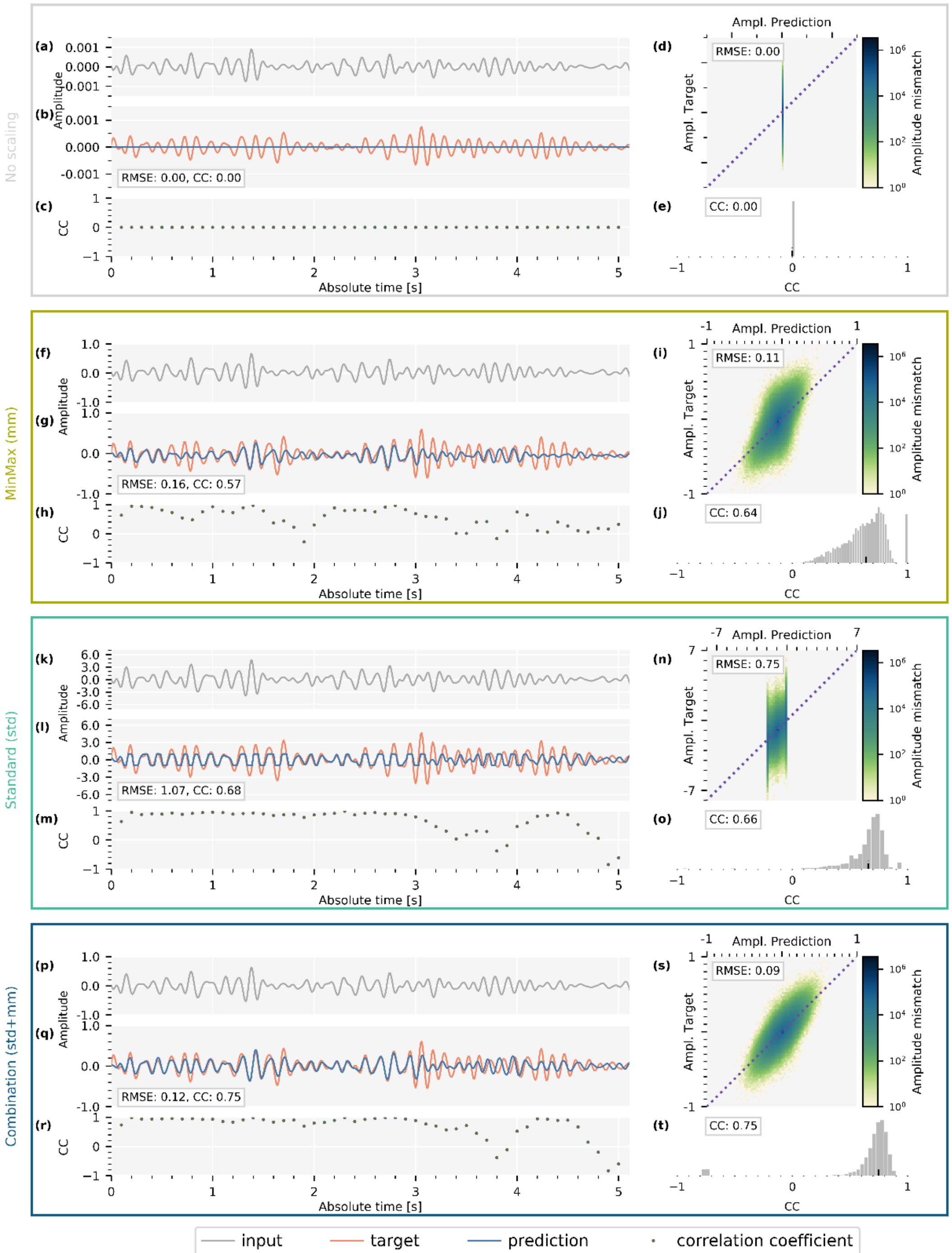


Figure 4.4. Results for different scalings using RMSE and cross-correlation coefficient (CC) across the entire target time-series. **(a, f, k, p)** Input data. **(b, g, l, q)** Target data and network prediction. **(d, i, n, s)** Density plots of prediction vs. actual target data; the dotted line is the ideal best-fitting line for the regression. **(c, h, m, r)** CC for 20-sample sections. **(e, j, o, t)** histograms of CC for 512 sample windows. The black marker highlights the overall CC.

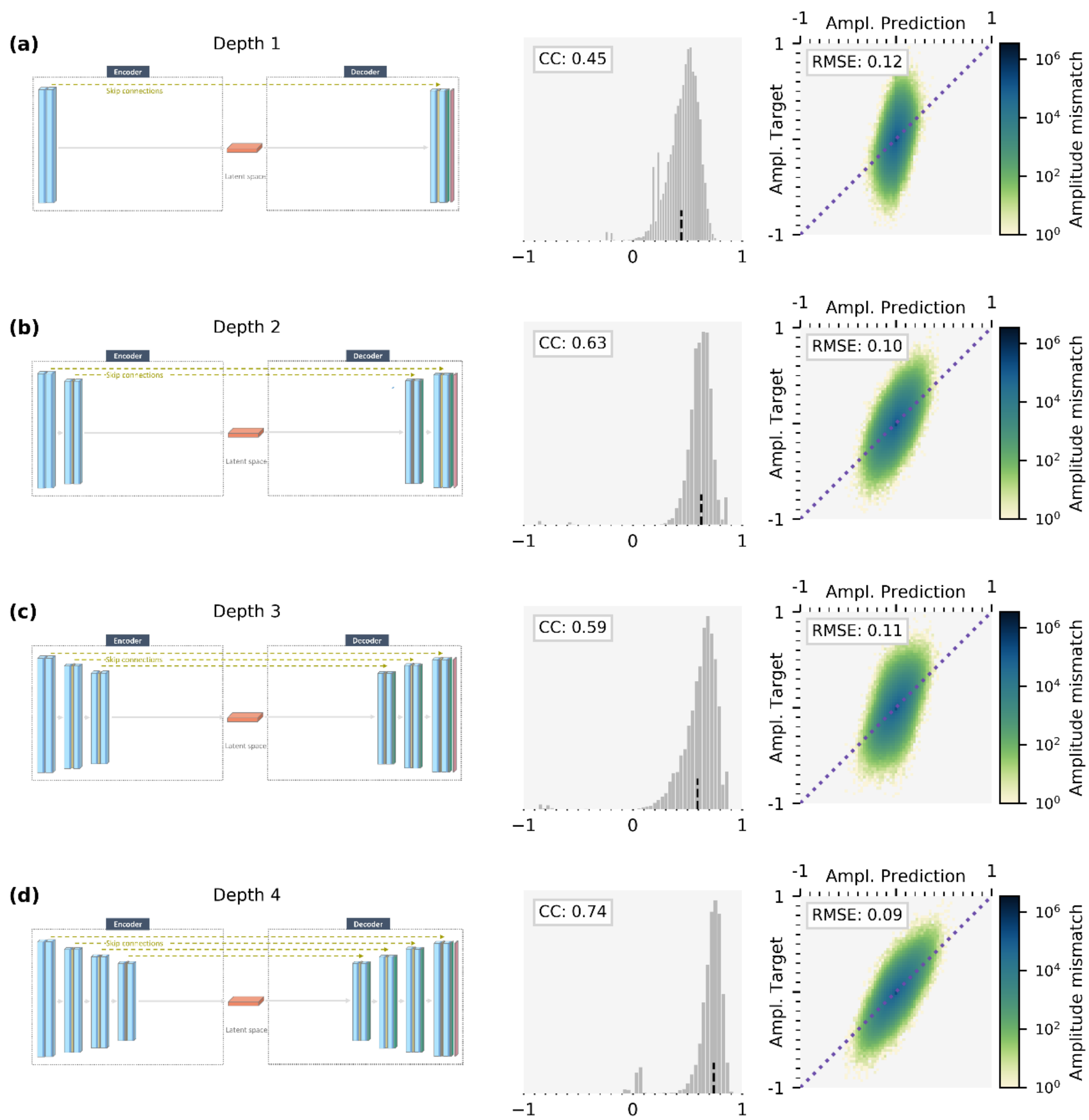
The results show, that the performance of the encoder-decoder network can significantly be impacted by the choice of scaling method. Without scaling, the model fails to identify and learn meaningful patterns, which results in poor predictions. Because features with different ranges cause large values to dominate smaller ones, numerical instability may be the cause of this (Braiek and Khomh, 2020; Sun et al., 2022), making it challenging for the model to learn appropriate weights. The optimization process may also fail or converge slowly because of backpropagation of unscaled data, which can cause very large or very small gradients.

MinMax scaling significantly improves the network performance and outputs more accurate predictions by scaling data between -1 and 1. The normalization of the data range is expected to stabilize the learning process and thus contribute positively to the results with a common scale for all data. Standard scaling normalizes data to zero mean and unit variance and achieves a similar phase prediction quality with a CC of 0.66. The predictions show a repetitive pattern of similar-sized waveforms forming small plateaus. The scaling might cause these consistent patterns, which emphasizes the relative changes in the data rather than their absolute values. Finally, the combination of standard scaling and MinMax scaling is the most effective, since it results in the highest CC of 0.75. In conclusion, using both standard and MinMax scaling together is likely the most effective approach, as it addresses the large- and fine-scale aspects as well as the data distribution. This provides the model with a more robust input representation for capturing global trends and local variations in the data.

Network depth

The number of layers in a neural network defines its depth, which affects how effectively it can identify and learn complex patterns. If the network depth is too low, the model fails to capture complex patterns in the data, which leads to underfitting and an insufficient representation of features (Sun et al., 2016; Telgarsky, 2016). If the network is too deep, the computational complexity is increased and can result in overfitting (Hastie et al., 2009; Nichani et al., 2021). Therefore, it is essential to find the right depths and balance between these options.

In this section, the impact of network depth on the algorithm's performance is investigated. Therefore, the depth of the network is systematically increased from one to seven layers (Figure 4.5, left), while each setting is trained individually. The results are evaluated using CC (Figure 4.5, middle), and Root Mean Square Error (RMSE) (Figure 4.5, right) again. Both of these evaluation metrics were introduced previously in **Study I** and Figure 4.4.



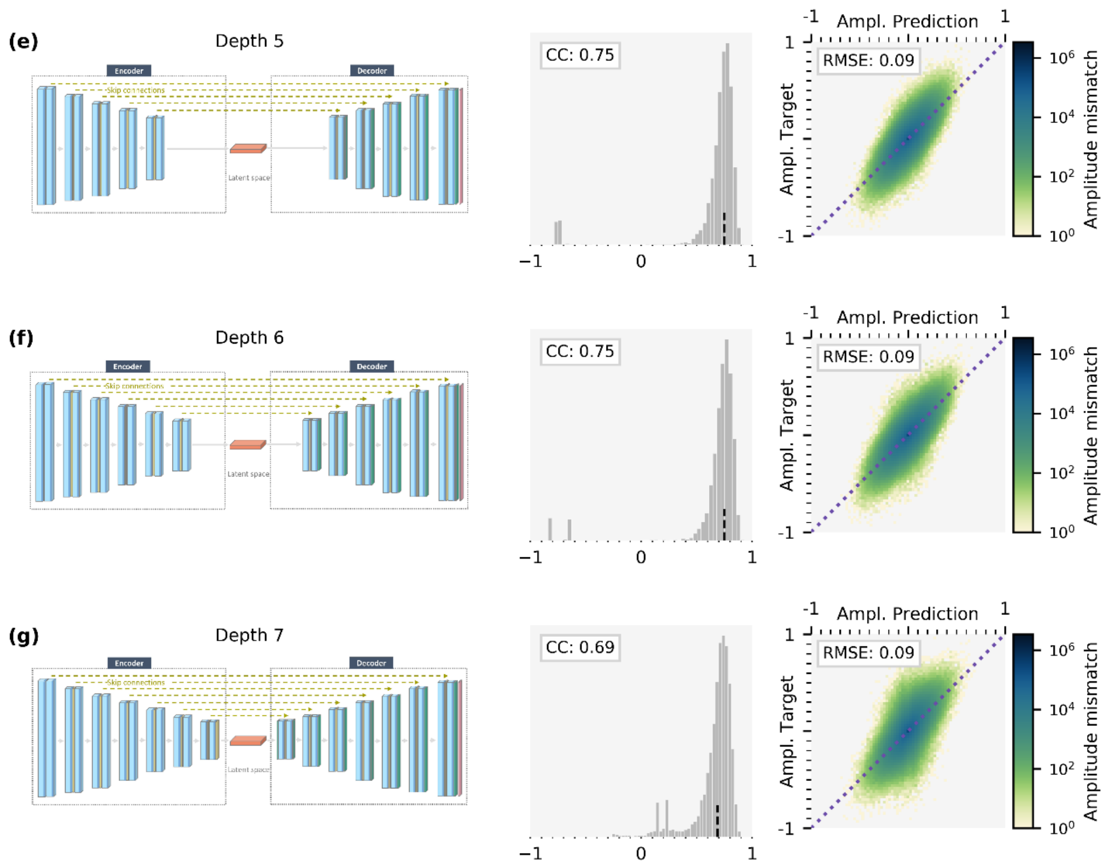


Figure 4.5. Impact of network depth on the network performance. Left: Schematic illustration of the network architecture with one to seven **(a-g)** layers. Middle: Histograms of correlation coefficients (CC) for 512 sample windows. Right: Density plots of prediction vs. actual target data as RMSE. For further description, see Figure 4.4.

The findings show that the CC generally improves with depth, going from 0.45 at depth one (Figure 4.5**a**) to 0.75 at depth six (Figure 4.5**e,f**), while the RMSE decreases corresponding and stabilizes after depth five. However, two exceptions are observed: depth two (Figure 4.5**b**), which outperforms depth three in both metrics (Figure 4.5**c**) but remains lower than higher depths (Figure 4.5**e,f**), and depth seven, which drops to a CC value of 0.69 again (Figure 4.5**f**). The CC distribution, as visualised in the bar plot (Figure 4.5, middle), shows a notable shift towards higher CC values as network depth increases. Initially for depth one, the distribution is broadly dispersed across the range from [0, 0.8], with a mean around 0.45 (Figure 4.5**a**, middle). For higher depths (four, five, six, and seven), the distribution is visually very similar and more condensed between values of 0.40 and 0.90 (Figure 4.5**d,e,f,g**, middle). The distribution of depth seven, however, becomes broader again, showing more CC values closer to zero compared to depth four to six.

As depth increases, the distribution becomes more concentrated, showing a narrower spread and a more pronounced central peak at higher CC values up to depth six. Scatter plots, showing the amplitude fit between target and prediction data, evolve from a near vertical distribution at depth one (Figure 4.5**a**, right) to a more diagonal arrangement along the best-fit line, which indicates improved amplitude predictions with increasing depth. Overall, depth four, five and six yield results that are almost identical (Figure 4.5**d**,**e**,**f**, right), with depth four showing a somewhat more condensed distribution along the best-fit line compared to depth six. For depth seven, the data scatter more widely again (Figure 4.5**g**, right), resulting in a broader distribution in both positive and negative directions, while the RMSE remains consistent with that of the previous depths.

The findings of varying network depth show, that the performance of the algorithm further depends on the choice of the number of layers in the network. For depth one, the performance of the model is already surprisingly good. This might be due to the regular intervals between seismic phases in the data, which may allow the model to recognize this even with minimal processing and a shallow network structure. Furthermore, it is surprising that a lower depth two outperforms the next higher depth three while not outperforming even higher depths. For this particular task, depth two might achieve a better balance between model capacity and generalization, whereas the network at depth three might encounter more optimization challenges during training, such as vanishing gradients. Additionally, features learnt at depth two might be more relevant or effective for this particular task than those learned at depth three. However, the significant improvement in performance from depth three to depth four suggests that the network can now more effectively identify the underlying patterns in the data. Notably, performance varies very little at depths four to five, suggesting little to no further improvements across the settings. However, depth seven shows a performance decrease again, which might be due to overfitting and indicates that this higher depth produces poorer results compared to former depths. Although we decided to use depth five for studies I and II, depth four might have been sufficient for the best results. This underscores how important it is to conduct empirical testing across different depths in order to determine the best architecture for a given task.

Training length

An epoch represents a complete cycle in which the neural network processes the entire training dataset one time. The number of epochs thereby defines how many times the network will repeat this entire process. To balance the learning capacity and prevent overfitting, the number of epochs must be set to a suitable value for a good model’s performance and generalization (Belkin et al., 2018; Too et al., 2019). To maximize learning and minimize overfitting, the ideal number of epochs is often determined by experimentation. Its number can vary widely depending on the dataset and model, typically ranging from a few dozen to hundreds or thousands (Too et al., 2019; Byrd and Lipton, 2019). It is common in practice to start with a moderate number and make adjustments according to the development of learning curves and validation metrics. Learning curves, which show training and validation metrics against the number of epochs, help in visualising the evolution of the model's performance over time. For instance, they allow the detection of when the model begins to overfit, which indicates by diverging training and validation curves, or when learning reaches plateaus (Viering and Loog, 2023; Mohr and van Rijn, 2024). The y-axis of these curves typically represents the loss function value, which quantifies the model's prediction error and should generally decrease over time for a well-learning model (Wang et al., 2022).

The model is trained with various numbers of epochs: 150, 375, 750, 1500, and 3000, arranged from top to bottom in Figure 4.6. The left column displays the loss curve for each epoch setting and shows how the model’s error changes over the training process. In the middle column, histograms of the CC for phase accuracy are shown, similar to those presented in Figure 4.5. The right column contains scatter plots of RMSE for amplitude predictions, also consistent with the ones used in Figure 4.5. This allows assessing how well the predicted amplitudes align with the true target values as the number of epochs increases and helps to identify the optimal training duration for this specific task.

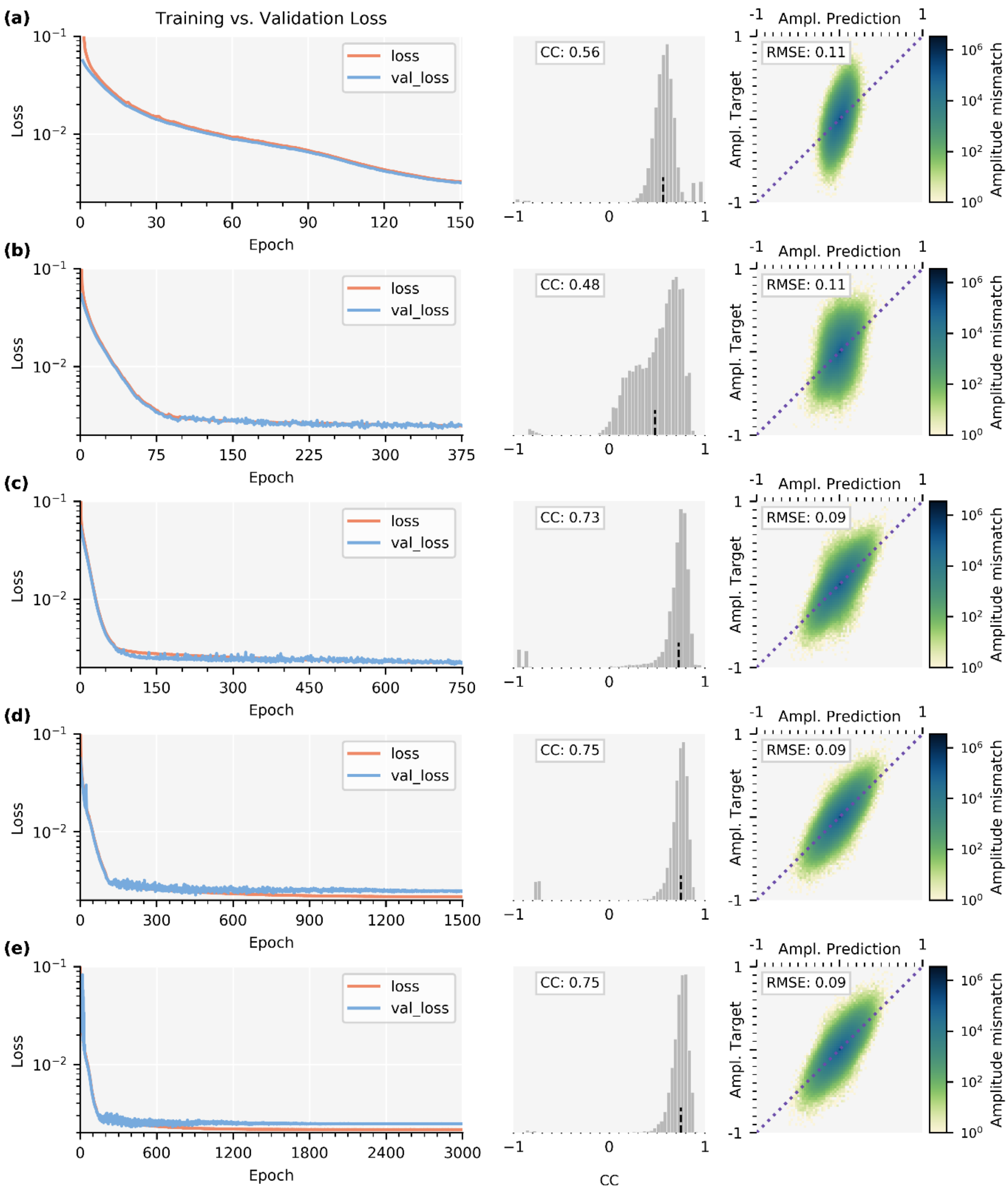


Figure 4.6. Impact of the number of training epochs on the model performance. The model is trained with 150, 375, 750, 1500, and 3000 epochs. Left column: Loss curves showing the models error evolution during training. Middle column: Histograms of CCs for phase accuracy. Right column: scatter plots of RMSE. Plots are similar to Figure 4.5.

For 150 epochs, the loss and validation loss curves consistently decrease (Figure 4.6**a**, left), closely overlaying each other. The CC histogram plot shows a condensed, nearly Gaussian distribution with a mean of 0.56 (Figure 4.6**a**, middle), while the amplitude scatter plot reveals a nearly elliptical data cloud, oriented more vertically than along the diagonal best fit line (Figure 4.6**a**, right). In contrast, over the course of 375 epochs, the loss decrease rate significantly slows down after about 150 epochs, with small oscillations starting to appear from around 100 epochs (Figure 4.6**a**, left). Nevertheless, a clear downward trend continues. The CC histogram plot for 375 epochs shows a wider spread of values, including some negative CCs (Figure 4.6**b**, middle). The amplitude scatter plot further exhibits a more dispersed pattern, resembling a rectangular shape rather than an ellipse or diagonal line (Figure 4.6**b**, right). This shows deviations in predicted amplitudes in both positive and negative directions to the target data.

For 750 epochs, the loss curve slowly continues decreasing, but is not reaching yet a plateau (Figure 4.6**c**, left). The bars of the overall CC histogram plot mostly distribute around a value of 0.73 (Figure 4.6**c**, middle), but also reveal two outliers near CCs of -1. Although there are still notable deviations in both positive and negative directions, the scatter plot displays improved alignment with the best-fit line compared to 375 epochs (Figure 4.6**c**, right). The loss curves for 1500 and 3000 epochs (Figure 4.6**d** & Figure 4.6**e**, left) reach a plateau after about 1200 epochs, with a small gap appearing between the lines for training and validation loss. Interestingly, training loss becomes smaller than validation loss in this plateau region. With overall CC values slightly increasing for higher overall epochs, the histogram plots are remarkably similar to those at 750 epochs (Figure 4.6**d** & Figure 4.6**e**, middle). Scatter plots become more elliptical and align better with the best-fit line, although deviations in both directions persist (Figure 4.6**d** & Figure 4.6**e**, right).

Analysing the results across different epoch settings shows, that in the early stages of 150 epochs, a consistent decrease in both training and validation loss is observed, indicating an ongoing learning process. The mid-range of 375-750 epochs shows a slowing rate of improvement, wider CC distributions including some negative values, and gradual alignment of amplitude predictions with the best-fit diagonal. Loss curves reach a plateau around 1200 epochs in later stages (1500-3000 epochs), suggesting that additional training yields negligible to no improvements. Notably, it can be observed that validation loss becomes smaller than the training loss in Figure 4.6**d** and **e** from 600 epochs onwards, which might be due to several reasons or a combination of them. As dropout is used to improve model generalization and prevent overfitting, it is possible that the model is generalizing too well to the validation set. However, this might result in a situation where the model performs better on the validation than on the more diverse training data. Another option is that it is easier to predict for the model when the validation data has certain characteristics, such as a mean that is comparable to the one of the training data but has less scatter. In this case, the network might become better at predicting the central tendency of the data distribution, performing well on the more focused validation set, while struggling with the training sets larger sample size. These observations suggest that lowering the dropout rate may be necessary to address the problem of validation loss decreasing relative to training loss. However, all examples demonstrate that the models learning rate is highest during the initial epochs and that training beyond 1200 epochs may not lead to significant performance improvements. According to these findings, 1500 epochs achieve a good balance between reaching stable model performance and minimizing additional computational costs linked to longer training periods. In studies I and II, 1500 epochs were chosen for training the models. As shown by the scatter plots, amplitude deviations between target and prediction data indicate further room for refinement.

Neuron activation

A neural network's activation function is a function that decides whether or not a neuron should be activated and whether an input is important to the prediction process. It transforms a neuron's input signal into an output signal, which is then passed to the subsequent layer. Without activation functions, neural networks would be limited to modeling only linear relationships, which reduces the networks power regarding more complex tasks such as image recognition or natural language processing (Sharma et al., 2017; Rasamoelina et al., 2020). The final activation function, sometimes referred to as the output layer's activation function, is applied to the final layer of a neural network and is typically distinct from the ones of the hidden layers. Its choice depends on the goal the network is designed to reach (Sharma et al., 2017). Linear activation allows any real number, while the hyperbolic tangent constrains outputs to [-1, 1], which is useful for normalized values. In contrast, sigmoid takes any input value and limits outputs from 0 to 1, which is often applied for predicting probabilities, but can also be used to decide which neuron should be activated based on the input (Sharma et al., 2017). Although there are dozens of activation functions (Ramachandran et al., 2017; Sharma et al., 2017) the focus in the following will be on the three previously mentioned ones.

To compare the algorithms performance on the three final activation layers - tanh, linear, and sigmoid - an individual model for each activation function is trained. The previously introduced metrics CC and RMSE are used to evaluate their performance, and again the same kind of histogram and density plots are used for visualisation.

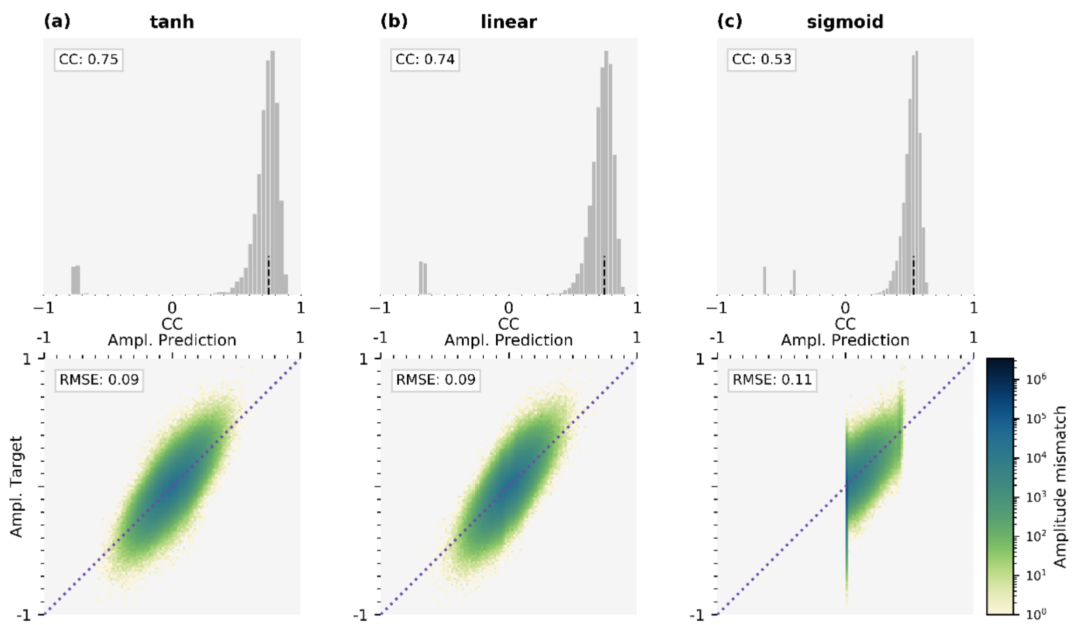


Figure 4.7. Comparison of performance of different final activation functions: **(a)** tanh, **(b)**, linear, and **(c)** sigmoid. The top row displays histogram plots showing the correlation coefficients for the entire test dataset for each activation function individually. The bottom row are density scatter plots showing the amplitude deviations as RMSEs between target data and predictions.

Figure 4.7 shows the performance of all three activation functions on the test data. The CC values for the first two examples are nearly the same, with histogram plots ranging from about 0.30 to 0.90 and a mean of 0.75 and 0.74 (Figure 4.7**a-b**, top). The scatter plots for linear and tanh activation functions are further very similar (Figure 4.7**a-b**, bottom). In contrast, the overall CC for the sigmoid activation function is lower compared to the other activation functions, with a mean CC of 0.53 (Figure 4.7**c**, top), while the scatter plot shows a sharp transition from low to high values (Figure 4.7**c**, bottom). Additionally, all three plots show outlier bars in the negative range.

The comparable performance of the tanh and linear activation functions is probably related to the data distribution and pre-processing. The target data range between [-1, 1], which allows both activation function to operate effectively. For tanh, this range is its natural output range, while linear can easily map inputs to outputs without scaling issues here. In contrast, the outputs of the sigmoid function are limited to the interval [0, 1], which leads to a mismatch with the target data range of [-1, 1], as it merges all negative values into the positive range. As a result, important information from the negative values in the target data is lost, causing a decline in performance. When input data is normalized, it benefits the tanh and linear activation functions, as it enables them to perform efficiently within their natural ranges. This same normalization, however, emphasizes the limitations of the sigmoid function for this task, as it has difficulty capturing the full range of the normalized data.

Information density

The sampling rate gives the number of samples that are recorded per second when measuring a continuous seismic signal. In general, higher sampling rates provide more data and can capture finer signal details, but also increase the storage space needed. Lower sampling rates, on the other hand, require less data storage, but may also not fully capture all frequency components of the seismic signals of interest. The sampling rate thus affects which information are in the data that can be learned by neural networks. For instance, higher sampling rates capture more detailed waveform information, which might improve the networks learning process, but also increase the size of input data to the network and thus the information that need to be processed.

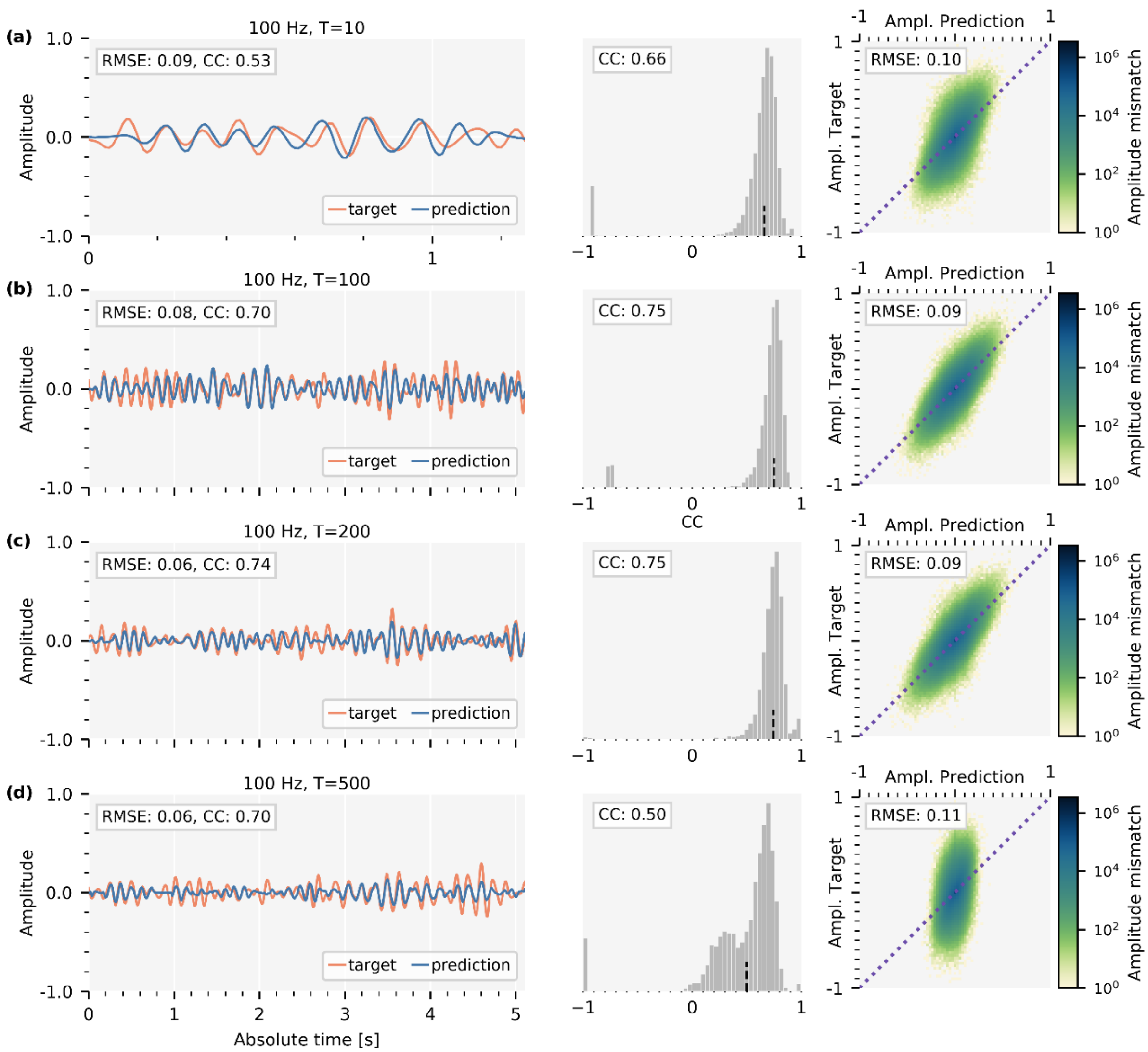


Figure 4.8. Results for data with varying numbers of periods at 100 Hz: (a) 10 periods, (b) 100 periods, (c) 200, and (d) 500 periods. Each row displays an example target trace against the model prediction (left), the distribution of correlation coefficients (CCs) over the entire range of data (middle), and the distribution of amplitudes of the target data against the prediction (right).

By adjusting the amount of data, e.g. through upsampling or downsampling, the time resolution and frequency range present in the data can be controlled. This adaption not only changes the characteristics of the data, but also influences the information that are included in the network training. For instance, the length of the input trace can be changed to control the number of periods present in the data, which indicates how many complete waveform cycles are included in a given time series segment. To approximate the amount of periods, two key assumptions are made: the data are filtered below 10 Hz, which gives the longest possible period of 0.1 seconds (calculated as 1 / 10 Hz), and a seismic velocity of 300 m/s is assumed. With these variables, the number of periods in the data are calculated by dividing the total length of the signal by the period of one cycle (0.1 seconds). In this subsection, it will be analysed how the number of periods and the sampling rate of the input data influence the performance of the neural network approach. To investigate this, several scenarios are analysed. Firstly, the original sampling rate of 100 Hz is maintained and the number of periods in the data is increased from 10, 100, to 200 and then to 500 periods. Afterwards, the data are downsampled to 25 Hz and the same process is repeated, again by increasing the number of periods from 10 to 500.

Figure 4.8 shows the results for the original sampling rate of 100 Hz and the respective number of periods from top (**a**) to bottom (**d**) starting with the lowest number of 10 periods. Both the results for 100 and 200 periods achieve similar values with overall CCs of 0.75 and RMSEs of 0.09. The bar plot (Figure 4.8, middle) reveals that the phase prediction is further consistent across the range of test traces for both examples and with values between 0.50 and 0.90 also of moderate to high accuracy. Furthermore, the density plots (Figure 4.8, right) show that the predicted amplitudes deviate from the target trace equally in positive and negative directions for both examples. Having 100 (**b**) and 200 (**c**) periods in the data reveals similar results, both achieving high performance. In contrast, the performance for 500 periods (**d**) is notably lower, yielding an overall CC of 0.50 and RMSE of 0.11. The performance for 10 periods (**a**) falls between these two extremes, showing intermediate results with a CC of 0.66 and RMSE of 0.10. While the CCs for 10 periods (Figure 4.8**a**, middle) show a distribution a little broader but very similar to those in **b** and **c**, the CCs for 500 periods distribute over a much broader range (from approximately -0.10 to 0.90) compared to the other period examples (Figure 4.8**d**, middle). The RMSE of 0.11 indicates more significant amplitude deviations compared to former examples, which supports by the density plot's nearly vertical distribution of points.

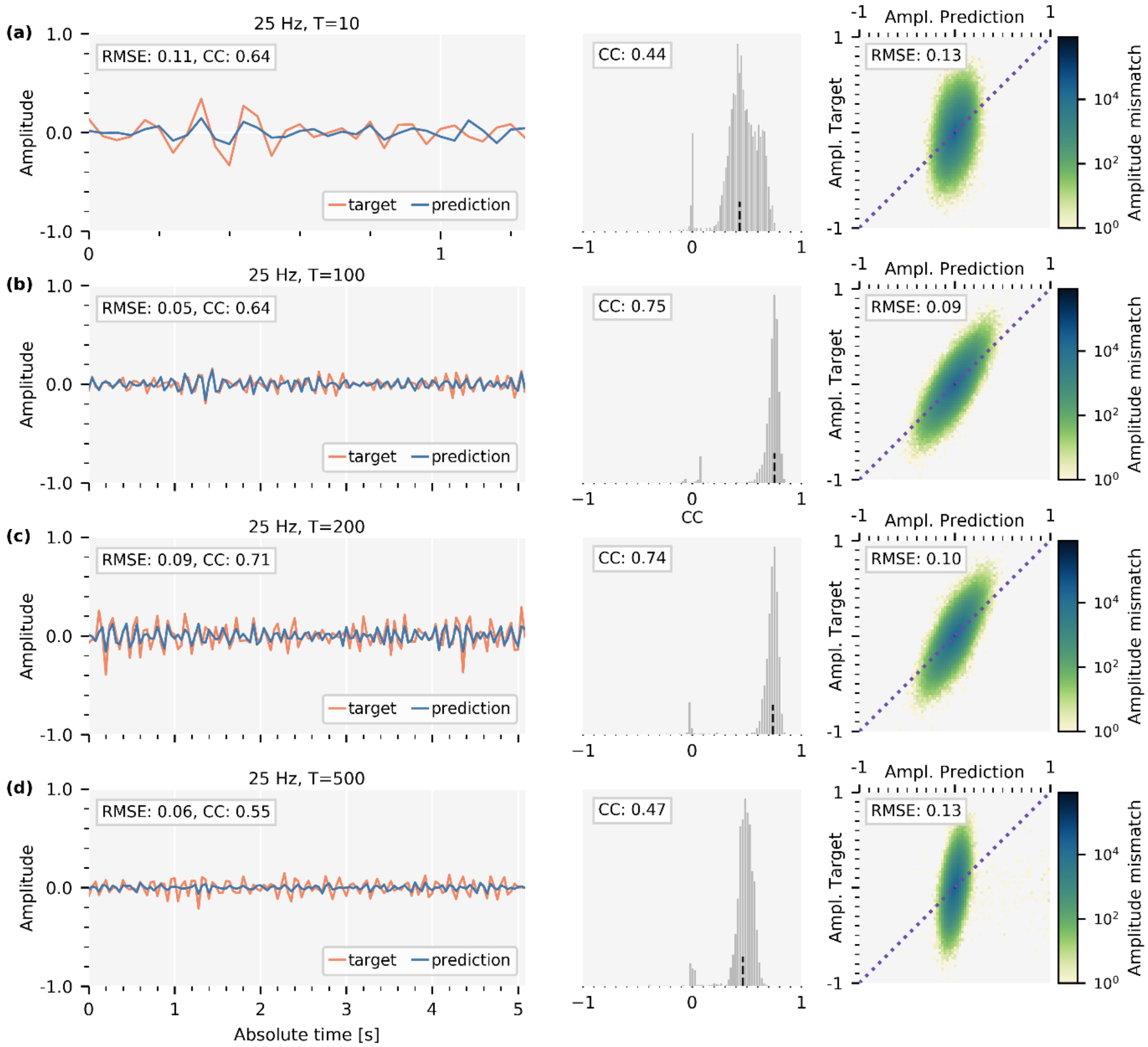


Figure 4.9. Results for data with varying numbers of periods at 25 Hz: (a) 10 periods, (b) 100 periods, (c) 200 periods, and (d) 500 periods. Description for subplots see Figure 4.8.

For a better comparison of the model performance across different sampling rates, the findings for data sampled at 25 Hz are further analysed (Figure 4.9). The results show that 100 (**b**) and 200 (**c**) periods in the data yield similar results again, with overall CCs of 0.75 and 0.74 respectively. While the overall distribution of CCs in these cases is even more focused around the mean values for 25 Hz, amplitude distributions show slightly lower performances compared to 100 Hz. Particularly in the 200-period case (Figure 4.9**c**), the amplitudes are less tightly distributed along the best-fit line, which leads to an RMSE of 0.10 compared to 0.09 for 100 Hz data. For the case of 10 periods at 25 Hz (Figure 4.9**a**), the lowest model performance is received across all examples with a CC of 0.44 and a RMSE of 0.13. Very close to this is also the results for 500 periods with a CC of 0.47 and RMSE of 0.13 (Figure 4.9**d**). This decrease in prediction accuracy is also visible in the density scatter plot, which shows a nearly vertical elliptical shape rather than an alignment with the best-fit line for 10 and 500 periods. While the overall CC histogram is clustered within a smaller range for 100 to 500 periods, the values are spread over a broader range with several peaks, including one around zero, for 10 periods.

For both 100 Hz and 25 Hz sampling rates, the consistency between 100 and 200 periods indicates that the models can learn the underlying patterns and features of the data. It seems that the signal's characteristics are represented in both 100 and 200 periods, so that an extension of the data after 100 periods do not provide new information for the model to learn. With an increased number of periods (500), a noticeable decrease in model performance is observed, with a more pronounced effect for 25 Hz compared to 100 Hz data. This may be attributed to several reasons: with an extension to 500 periods, the data contain more information and potentially include more complex patterns that the model fails to learn. This increase in data complexity can further lead to difficulties in generalization and can cause overfitting while the model tries to capture very specific details. Due to this increased amount of data, an adaption of the model architecture might also be needed, for instance, by increasing the network depth to better handle the new complexity of the data. A reduction in model performance is seen with a smaller number of periods (10), which is more pronounced at 25 Hz than 100 Hz. Data with only 10 periods contain limited information and potentially lack sufficient patterns for the model to learn effectively. This can lead to difficulties in generalization and may cause underfitting as the model struggles to capture the underlying trends. While the sampling rate does not seem to influence the results significantly, it is evident that the right amount of data for the training is important. The findings demonstrate that there can be both too little and too much information in the data, which suggests that the identification of the optimal data quantity is essential to achieve the best model performance.

4.2. Discussion and Conclusion

This study delivers first insights into parameters that affect the performance of the neural network approach and gives a place to start when trying to improve it in the future. Several parameters are investigated, including interstation distance, frequency ranges, data scaling methods, network depth, number of training epochs, final activation functions, and sampling rates. The findings highlight settings where the models perform well and identify limitations where additional parameter and data tuning is needed. Notably, the outcomes of this parameter study support that the choice of parameters taken in **Study I** and **Study II** works very well.

To find the best combination of parameters or optimize an existing setting while training a neural network, grid search can be used. It defines a grid of parameter values and trains models for all combinations, before evaluating their performance and selecting the parameter set that is performing best (Bergstra and Bengio, 2012). However, testing multiple different sets of parameters can be computationally expensive and may not provide improvements if a good set of parameters be already found. Furthermore, it efficiency depends on the definition of the parameter grid – if it does not cover optimal regions for testing, an automated grid search might not improve the results any further. Since various parameter settings were tested manually for the neural network approach, no further large improvements in the model results are expected from an automated grid search.

Based on the results, future studies need to consider that the information content of the data changes with e.g. measurement method or regional geology, which makes it highly important to understand the characteristics of the dataset before starting detailed investigations. This could include the computation of spectrograms or other time-frequency representations to better be aware of signal characteristics at different sampling rates. In the context of machine learning methods, it is further necessary to understand how different parameters influence each other, which can reveal relationships that might affect the model performance. For instance, the number of layers and the number of periods are related. If the number of layers is increased, the network can resolve increasingly finer details within the data. However, this refinement occurs only in the close neighbourhood of the features being processed, which is very helpful for image processing, but might not be ideal for capturing longer-term relationships in waveforms. Furthermore, there is a relationship between the network depth and the number of training epochs. Deeper networks often require more epochs to converge, but this also increases the risk of overfitting, where the model begins to memorize the training data instead of generalizing to unseen data. To address this, an early-stopping criterion can be implemented to end training as soon as the validation performance stops to improve. Higher sampling rates further allow capturing higher frequency components, which may affect how the data are filtered.

Moreover, it can be necessary to advance the network architecture, for instance, to improve the algorithm's ability to predict amplitudes more accurate, which would reduce the deviations between target and predicted data and enhance the overall model performance. This can include the adjustment of layer types, the optimization of activation functions or even the adoption of an entirely different neural network architecture. To continue research in these directions seems promising to improve the understanding regarding the performance, robustness, and applicability of the approach.

5. Summary and Conclusions

Machine learning offers new ways for the analysis and interpretation of complex seismic data. Building on this potential, this study presents a novel machine learning application to seismic data analysis, which focuses on the use of encoder-decoder neural networks to predict wave propagation between seismic stations. This work involves setting up an encoder-decoder model to learn and predict transfer functions between seismic stations, testing it with selected station pairs in the noise-rich Vienna Basin (Austria), and expanding the analysis to a full array of stations, assessing the results through beamforming. Additionally, the role of pre-processing and different parameter settings on the model performance was examined. These three studies allow addressing the overarching research questions regarding the application of neural networks to predict seismic wave propagation.

Q1 Can machine learning techniques be adapted to learn the transfer function between seismic stations for predicting seismic wave fields?

The findings of **Study I** demonstrate that it is possible to successfully learn the transfer function between two seismic stations using an encoder-decoder network architecture. By using one-dimensional time-series data from a reference station as input to the model training and data from a nearby station as the target, the network learns the signal transformations between the two locations and demonstrates its potential for transfer function modeling. Remarkably, the model achieved good predictions using only two days of field data consisting solely of ambient seismic noise. This highlights the robustness of the approach and shows its capacity to identify meaningful patterns in seismic noise records with a comparatively small amount of data. Interestingly, for all examples, network predictions for phases are of higher accuracy than those for amplitudes. This suggests that the network is more effective at capturing phase-related features, which is likely because the regular intervals between phases present a more predictable pattern for the model to learn in contrast to the highly variable nature of amplitudes. Overall, since resolving this issue would produce a more accurate representation of the transfer function, the difference between phase and amplitude prediction offers valuable information for additional research. This might involve modifications of the network architecture, for example by using variational autoencoders (VAEs) instead of classical encoder-decoder networks, which are probabilistic models that learn the underlying distribution of a dataset and may allow better modeling of seismic amplitude variability (Kingma et al., 2013). Furthermore, exploring the use of amplitude terms in the loss function and modifying the decoder architecture, which is important for reconstructing seismic features, might be worth considering to improve the accuracy of amplitude predictions (Wolfe and Godsill, 2003; Yin et al., 2022). This would further expand the significant strength of accurate phase prediction with improved amplitude prediction, highlighting the potential of this machine learning approach to complement conventional seismological methods and predict ground motion in a range of geological settings.

Q2 Can we use a single seismic station to predict the data of an entire seismic array?

Study I demonstrated that it is possible to learn the transfer function between two seismic stations appropriately, using the encoder-decoder approach introduced. This achievement in single station pair analysis lays the groundwork for its application in more advanced applications, such as predicting the transfer functions of an entire seismic array. The approach was applied to every station pair in a given array, and the consistency of the original and predicted records was assessed using beamforming individually. Similar to **Study I**, each station pair model exhibits differences in amplitude prediction compared to the target data. Despite these variations, the beamforming demonstrates that all models effectively learn the most important features from the data, as the plots for the target data and predictions are identical. This shows that even when predictions are not perfect, the algorithm successfully learns and predicts key wave characteristics, which highlights its robustness as well as its ability to use data from a single reference station to reconstruct seismic recordings at multiple locations. This approach creates a Virtual Seismic Array – a set of trained models, one for each array station - that enables data prediction in areas where physical sensors are no longer present. Based on these findings, Virtual Seismic Arrays are a promising new approach for seismic data acquisition and analysis, as they could reduce the need for physical sensor deployments, for example in regions that are difficult to access, and lower costs associated with the maintenance of seismic arrays. This study established a successful proof of concept for Virtual Seismic Arrays, while further investigations are needed to refine the method. The findings show that the algorithm only makes predictions based on patterns it has discovered during training. At present, the investigations focus on seismic data containing at most two regimes. This presents a great opportunity to expand the approach, since training in more complex regimes may significantly deepen understanding of the algorithm's behaviour and potentially move Virtual Seismic Arrays closer to real-world implementation. Exploring how the algorithm responds to a wider range of geological features could further lead to more robust and adaptable Virtual Seismic Arrays. This may be realized by assessing different periods, such as days, weeks or months apart from the initial training data and evaluating how well the algorithm works in regions with various geological conditions. Furthermore, extending the study to other frequency ranges – specifically, the primary microseism frequency band – would provide valuable insights into the algorithm's adaptability. Building on this, exploring the its applicability to other types of seismic signals, such as earthquakes or volcanic tremors, in addition to seismic noise, could further expand the algorithm's potential use cases.

Q3 What impact do different parameters have on the performance of machine learning techniques for predicting seismic wave fields?

To complement the promising development of an encoder-decoder approach for learning transfer functions between seismic stations (**Study I**) and applying it to a full seismic array (**Study II**), the parameter study in chapter **4** evaluates the influence of various parameters on the model performance. In this way, the robustness of the approach could be examined across various settings, including how the network responds to parameters like network depth, the amount of training epochs, final activation functions or information content in the data. According to the results of this study, network performance is impacted by data scaling, and the best results were obtained when two scaling techniques were combined. This finding confirms that the choice of combined scaling used in earlier studies **I** and **II** was reasonable. Furthermore, network depth had a complex effect; shallower networks were already capable of learning specific features effectively, while deeper networks had higher accuracy and captured even more features. Therefore, it is important to estimate the proper trade-off between them. In the earlier studies, this was achieved using a network depth of five, though a depth of four might have been sufficient and would have reduced computation time. The length of training was also found to be important, with both too little and too excessive training having a negative effect on the performance. It is interesting to note that the final activation functions tested had minimal impact on the results, while the amount of training data significantly influenced the findings, with both too little and too much information limiting the best possible learning. Understanding how all of these parameters interact is therefore crucial because these relationships can fundamentally affect the model performance. To improve the outcomes and their interpretability, it would further require appropriate techniques to better understand and visualise what the model is learning in order to be able to implement targeted next steps. As was previously discussed for **Study I**, upcoming research may also look into more advanced network architectures like Variational Autoencoders (VAEs). Overall, the parameters selected for studies **I** and **II** performed very well. While they might need to be modified for other datasets or regions, they provide a solid starting point for future applications.

By leveraging a novel machine learning approach, this study reveals the complex relationship between seismic stations, enabling to use data from one seismic station to predict signals at another. This method is unique in its application, since machine learning, specifically encoder-decoder networks, has yet not been used to model seismological transfer functions. Notably, the estimated models achieve high precision in predicting seismic phases and perform well in estimating amplitudes, still leaving room further refinement. This method is especially beneficial for applications that need precise seismic isolation or compensation, including active vibration isolation, semiconductor manufacturing, and 3-D microfabrication (Hwang et al., 2004; Kerber et al., 2007; Kim et al., 2009). In projects like the Einstein telescope (Punturo et al., 2010; Harms et al., 2022), minimizing seismic disturbances is further essential to guarantee the sensitivity of gravitational wave detections. Through precise mapping of seismic noise between two locations, the encoder-decoder approach provides a powerful tool for learning the characteristics of disturbing background sounds and thus allows for their controlled suppression. Virtual Seismic Arrays (VSAs) extend the idea of station-pair modeling to entire arrays, and offer unique potential to revolutionize the field of seismology by reducing the need for extensive sensor deployments through predictive modeling.

This work not only promises to improve the efficiency and cost-effectiveness of seismic monitoring, but it also opens the way for new insights into the dynamics of the Earth. With the ongoing evolution of this methodology, it may reshape seismological research and monitoring, leading to a greater understanding of the Earth's internal structure.

References

Ahmad, M.A., Eckert, C., Teredesai, A., 2018. Interpretable Machine Learning in Healthcare, in: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB '18. Association for Computing Machinery, New York, NY, USA, pp. 559–560. https://doi.org/10.1145/3233547.3233667

Ahmed, S., Alshater, M.M., Ammari, A.E., Hammami, H., 2022. Artificial intelligence and machine learning in finance: A bibliometric review. Res. Int. Bus. Finance 61, 101646. https://doi.org/10.1016/j.ribaf.2022.101646

Ahsan, M.M., Mahmud, M.A.P., Saha, P.K., Gupta, K.D., Siddique, Z., 2021. Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. Technologies 9. https://doi.org/10.3390/technologies9030052

An, P., Moon, W.M., Kalantzis, F., 2001. Reservoir characterization using seismic waveform and feedforword neural networks. Geophysics 66, 1450–1456. https://doi.org/10.1190/1.1487090

Ao, J., Wang, R., Zhou, L., Wang, C., Ren, S., Wu, Y., Liu, S., Ko, T., Li, Q., Zhang, Y., Wei, Z., Qian, Y., Li, J., Wei, F., 2022. SpeechT5: Unified-Modal Encoder-Decoder Pre-Training for Spoken Language Processing.

Araujo, T., Helberger, N., Kruikemeier, S., de Vreese, C.H., 2020. In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI Soc. 35, 611–623. https://doi.org/10.1007/s00146-019-00931-w

Badola, K., Gupta, M., 2021. Twitter Spam Detection Using Natural Language Processing by Encoder Decoder Model, in: 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). pp. 402–405. https://doi.org/10.1109/ICAIS50930.2021.9395862

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017b. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

Barker, J.S., Campillo, M., Sánchez-Sesma, F.J., Jongmans, D., Singh, S.K., 1996. Analysis of wave propagation in the Valley of Mexico from a dense array of seismometers. Bull. Seismol. Soc. Am. 86, 1667–1680. https://doi.org/10.1785/BSSA0860061667

Belkin, M., Ma, S., Mandal, S., 2018. To Understand Deep Learning We Need to Understand Kernel Learning, in: Dy, J., Krause, A. (Eds.), Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research. PMLR, pp. 541–549.

Bormann, P., Engdahl, B., Kind, R., 2012. Seismic wave propagation and earth models, in: New Manual of Seismological Observatory Practice 2 (NMSOP2). Deutsches GeoForschungsZentrum GFZ, pp. 1–105.

Bormann, P., Wielandt, E., 2013. Seismic signals and noise, in: New Manual of Seismological Observatory Practice 2 (NMSOP2). Deutsches GeoForschungsZentrum GFZ, pp. 1–62.

Boué, P., Denolle, M., Hirata, N., Nakagawa, S., Beroza, G.C., 2016. Beyond basin resonance: characterizing wave propagation using a dense array and the ambient seismic field. Geophys. J. Int. 206, 1261–1272. https://doi.org/10.1093/gji/ggw205

Braiek, H.B., Khomh, F., 2020. On testing machine learning programs. J. Syst. Softw. 164, 110542. https://doi.org/10.1016/j.jss.2020.110542

Byrd, J., Lipton, Z., 2019. What is the Effect of Importance Weighting in Deep Learning?, in: Chaudhuri, K., Salakhutdinov, R. (Eds.), Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research. PMLR, pp. 872–881.

Campillo, M., Brenguier, F., Hadziioannou, C., Shapiro, N., Larose, E., 2008. Monitoring changes in crustal properties with seismic noise. J. Acoust. Soc. Am. 123, 3269–3269. https://doi.org/10.1121/1.2933596

Capon, J., 1973. Signal Processing and Frequency-Wavenumber Spectrum Analysis for a Large Aperture Seismic Array* *This work was sponsored by the Advanced Research Projects Agency of the Department of Defense., in: Bolt, B.A. (Ed.), Geophysics, Methods in Computational Physics: Advances in Research and Applications. Elsevier, pp. 1–59. https://doi.org/10.1016/B978-0-12-460813-9.50007-2

Capon, J., 1969. High-resolution frequency-wavenumber spectrum analysis. Proc. IEEE 57, 1408–1418. https://doi.org/10.1109/PROC.1969.7278

Chapman, C., 2004. Fundamentals of Seismic Wave Propagation. Cambridge University Press.

Chen, J.C., Yao, K., Hudson, R.E., 2002. Source localization and beamforming. IEEE Signal Process. Mag. 19, 30–39. https://doi.org/10.1109/79.985676

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV).

Cheng, M.H., Kohler, M.D., Heaton, T.H., 2014. Prediction of Wave Propagation in Buildings Using Data from a Single Seismometer. Bull. Seismol. Soc. Am. 105, 107–119. https://doi.org/10.1785/0120140037

Chiariotti, P., Martarelli, M., Castellini, P., 2019. Acoustic beamforming for noise source localization – Reviews, methodology and applications. Mech. Syst. Signal Process. 120, 422–448. https://doi.org/10.1016/j.ymssp.2018.09.019

Chiu, C.-C., Sainath, T.N., Wu, Y., Prabhavalkar, R., Nguyen, P., Chen, Z., Kannan, A., Weiss, R.J., Rao, K., Gonina, E., Jaitly, N., Li, B., Chorowski, J., Bacchiani, M., 2018. State-of-the-Art Speech Recognition with Sequence-to-Sequence Models, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4774–4778. https://doi.org/10.1109/ICASSP.2018.8462105

Dean, T., Grant, M., 2024. A beginner's guide to seismic sensors. Preview 2024, 38–44. https://doi.org/10.1080/14432471.2024.2395647

Denolle, M.A., Dunham, E.M., Prieto, G.A., Beroza, G.C., 2013. Ground motion prediction of realistic earthquake sources using the ambient seismic field. J. Geophys. Res. Solid Earth 118, 2102–2118. https://doi.org/10.1029/2012JB009603

Dias, F.L., Assumpção, M., Peixoto, P.S., Bianchi, M.B., Collaço, B., Calhau, J., 2020. Using Seismic Noise Levels to Monitor Social Isolation: An Example From Rio de Janeiro, Brazil. Geophys. Res. Lett. 47, e2020GL088748. https://doi.org/10.1029/2020GL088748

Dixon, M.F., Halperin, I., Bilokon, P., 2020. Machine learning in finance. Springer.

Duguleană, M., Briciu, V.-A., Duduman, I.-A., Machidon, O.M., 2020. A Virtual Assistant for Natural Interactions in Museums. Sustainability 12. https://doi.org/10.3390/su12176958

Friedrich, A., Krüger, F., Klinge, K., 1998. Ocean-generated microseismic noise located with the Gräfenberg array. J. Seismol. 2, 47–64. https://doi.org/10.1023/A:1009788904007

Fujiyoshi, H., Hirakawa, T., Yamashita, T., 2019. Deep learning-based image recognition for autonomous driving. IATSS Res. 43, 244–252. https://doi.org/10.1016/j.iatssr.2019.11.008Gaci, S., 2014. The Use of Wavelet-Based Denoising Techniques to Enhance the First-Arrival Picking on Seismic Traces. IEEE Trans. Geosci. Remote Sens. 52, 4558–4563. https://doi.org/10.1109/TGRS.2013.2282422

Gibbons, S.J., Ringdal, F., 2012. Seismic Monitoring of the North Korea Nuclear Test Site Using a Multichannel Correlation Detector. IEEE Trans. Geosci. Remote Sens. 50, 1897–1909. https://doi.org/10.1109/TGRS.2011.2170429

Gibbons, S.J., Ringdal, F., 2006. The detection of low magnitude seismic events using array-based waveform correlation. Geophys. J. Int. 165, 149–166. https://doi.org/10.1111/j.1365-246X.2006.02865.x

Green, Jr., P.E., Kelly, Jr., E.J., Levin, M.J., 1966. A Comparison of Seismic Array Processing Methods. Geophys. J. Int. 11, 67–84. https://doi.org/10.1111/j.1365-246X.1966.tb03493.x

Guo, X., Zhang, Y., Lu, S., Lu, Z., 2024. Facial expression recognition: a review. Multimed. Tools Appl. 83, 23689–23735. https://doi.org/10.1007/s11042-023-15982-x

Gurney, K., 2018. An introduction to neural networks. CRC press.

Haendel, A., Ohrnberger, M., Krüger, F., 2018. Frequency-dependent quality factors from the deconvolution of ambient noise recordings in a borehole in West Bohemia/Vogtland. Geophys. J. Int. 216, 251–260. https://doi.org/10.1093/gji/ggy422

Harms, J., Naticchioni, L., Calloni, E., De Rosa, R., Ricci, F., D'Urso, D., 2022. A lower limit for Newtonian-noise models of the Einstein Telescope. Eur. Phys. J. Plus 137, 687. https://doi.org/10.1140/epjp/s13360-022-02851-z

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction. Springer.

Havskov, J., Alguacil, G., 2016. Correction for Instrument Response, in: Instrumentation in Earthquake Seismology. Springer International Publishing, Cham, pp. 197–230. https://doi.org/10.1007/978-3-319-21314-9_6

Hillers, G., Graham, N., Campillo, M., Kedar, S., Landès, M., Shapiro, N., 2012. Global oceanic microseism sources as seen by seismic arrays and predicted by wave action models. Geochem. Geophys. Geosystems 13. https://doi.org/10.1029/2011GC003875

Hu, W., Barthelmie, R.J., Letson, F., Pryor, S.C., 2019. Seismic Noise Induced by Wind Turbine Operation and Wind Gusts. Seismol. Res. Lett. 91, 427–437. https://doi.org/10.1785/0220190095

Hwang, J., Huang, Y., Hung, Y., Huang, J., 2004. Applicability of seismic protective systems to structures with vibration-sensitive equipment. J. Struct. Eng. 130, 1676–1684.

Imrie, P. and Bednar, P., 2013. Virtual Personal Assistant, in: Martinez, Marcello and Pennarolaecilia, Ferdinando (Ed.), ItAIS 2013. Proceedings of 10th Conference of the Italian Chapter of AIS. Università Commerciale Luigi Bocconi in Milan, Italy,.

Jongmans, D., Garambois, S., 2007. Geophysical investigation of landslides : a review. Bull. Société Géologique Fr. 178, 101–112. https://doi.org/10.2113/gssgfbull.178.2.101

Jozinović, D., Lomax, A., Štajduhar, I., Michelini, A., 2020. Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. Geophys. J. Int. 222, 1379–1389. https://doi.org/10.1093/gji/ggaa233

Karamzadeh, N., Heimann, S., Dahm, T., Krüger, F., 2018. Application based seismological array design by seismicity scenario modelling. Geophys. J. Int. 216, 1711–1727. https://doi.org/10.1093/gji/ggy523

Keenan, R.E., Dyer, I., 1984. Noise from Arctic Ocean earthquakes. J. Acoust. Soc. Am. 75, 819–825. https://doi.org/10.1121/1.390591

Kerber, F., Hurlebaus, S., Beadle, B.M., Stöbener, U., 2007. Control concepts for an active vibration isolation system. Mech. Syst. Signal Process. 21, 3042–3059. https://doi.org/10.1016/j.ymssp.2007.04.003

Kim, S.J., Dean, R., Flowers, G., Chen, C., 2009. Active Vibration Control and Isolation for Micromachined Devices. J. Mech. Des. 131. https://doi.org/10.1115/1.3159042

Kingma, D.P., Welling, M., others, 2013. Auto-encoding variational bayes.

Kiser, E., Ishii, M., 2012. Combining seismic arrays to image the high-frequency characteristics of large earthquakes. Geophys. J. Int. 188, 1117–1128. https://doi.org/10.1111/j.1365-246X.2011.05299.x

Kong, Q., Trugman, D.T., Ross, Z.E., Bianco, M.J., Meade, B.J., Gerstoft, P., 2018. Machine Learning in Seismology: Turning Data into Insights. Seismol. Res. Lett. 90, 3–14. https://doi.org/10.1785/0220180259

Koper, K.D., de Foy, B., Benz, H., 2009. Composition and variation of noise recorded at the Yellowknife Seismic Array, 1991–2007. J. Geophys. Res. Solid Earth 114. https://doi.org/10.1029/2009JB006307

Kotsiantis, S.B., Kanellopoulos, D., Pintelas, P.E., 2006. Data preprocessing for supervised leaning. Int. J. Comput. Sci. 1, 111–117.

Kubo, H., Naoi, M., Kano, M., 2024. Recent advances in earthquake seismology using machine learning. Earth Planets Space 76, 36. https://doi.org/10.1186/s40623-024-01982-0

LeCun, Y., Bottou, L., Orr, G.B., Müller, Klaus -Robert, 1998. Efficient BackProp, in: Orr, G.B., Müller, Klaus-Robert (Eds.), Neural Networks: Tricks of the Trade. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 9–50. https://doi.org/10.1007/3-540-49430-8_2

Lim, D., Ahn, J.-K., 2023. Horizontal seismic wave at ground surface from transfer function based on ambient noise. Front. Earth Sci. 11. https://doi.org/10.3389/feart.2023.1047667

Liu, W.Y., 2017. A review on wind turbine noise mechanism and de-noising techniques. Renew. Energy 108, 311–320. https://doi.org/10.1016/j.renene.2017.02.034

Lythgoe, K., Loasby, A., Hidayat, D., Wei, S., 2021. Seismic event detection in urban Singapore using a nodal array and frequency domain array detector: earthquakes, blasts and thunderquakes. Geophys. J. Int. 226, 1542–1557. https://doi.org/10.1093/gji/ggab135

Majer, E.L., Baria, R., Stark, M., Oates, S., Bommer, J., Smith, B., Asanuma, H., 2007. Induced seismicity associated with Enhanced Geothermal Systems. Geothermics 36, 185–222. https://doi.org/10.1016/j.geothermics.2007.03.003

McClellan, J.H., Eisner, L., Liu, E., Iqbal, N., Al-Shuhail, A.A., Kaka, S.I., 2018. Array Processing in Microseismic Monitoring: Detection, Enhancement, and Localization of Induced Seismicity. IEEE Signal Process. Mag. 35, 99–111. https://doi.org/10.1109/MSP.2017.2776798

McCowan, D.W., Lacoss, R.T., 1978. Transfer functions for the seismic research observatory seismograph system. Bull. Seismol. Soc. Am. 68, 501–512. https://doi.org/10.1785/BSSA0680020501

Meng, H., Ben-Zion, Y., 2017. Detection of small earthquakes with dense array data: example from the San Jacinto fault zone, southern California. Geophys. J. Int. 212, 442–457. https://doi.org/10.1093/gji/ggx404

Meng, L., Allen, R.M., Ampuero, J. -P., 2014. Application of Seismic Array Processing to Earthquake Early Warning. Bull. Seismol. Soc. Am. 104, 2553–2561. https://doi.org/10.1785/0120130277

Mohr, F., van Rijn, J.N., 2024. Learning curves for decision making in supervised machine learning: a survey. Mach. Learn. 113, 8371–8425. https://doi.org/10.1007/s10994-024-06619-7

Morariu, C., Morariu, O., Răileanu, S., Borangiu, T., 2020. Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems. Comput. Ind. 120, 103244. https://doi.org/10.1016/j.compind.2020.103244

Mousavi, S.M., Beroza, G.C., 2023. Machine Learning in Earthquake Seismology. Annu. Rev. Earth Planet. Sci. https://doi.org/10.1146/annurev-earth-071822-100323

Mousavi, S.M., Beroza, G.C., 2022. Deep-learning seismology. Science 377, eabm4470. https://doi.org/10.1126/science.abm4470

Moustra, M., Avraamides, M., Christodoulou, C., 2011. Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals. Expert Syst. Appl. 38, 15032–15039. https://doi.org/10.1016/j.eswa.2011.05.043

Nanni, U., Roux, P., Gimbert, F., Lecointre, A., 2022. Dynamic Imaging of Glacier Structures at High-Resolution Using Source Localization With a Dense Seismic Array. Geophys. Res. Lett. 49, e2021GL095996. https://doi.org/10.1029/2021GL095996

Neuberg, J., Luckett, R., Ripepe, M., Braun, T., 1994. Highlights from a seismic broadband array on Stromboli Volcano. Geophys. Res. Lett. 21, 749–752. https://doi.org/10.1029/94GL00377

Nichani, E., Radhakrishnan, A., Uhler, C., 2021. Increasing Depth Leads to U-Shaped Test Risk in Over-parameterized Convolutional Networks.

Pardo, E., Garfias, C., Malpica, N., 2019. Seismic Phase Picking Using Convolutional Networks. IEEE Trans. Geosci. Remote Sens. 57, 7086–7092. https://doi.org/10.1109/TGRS.2019.2911402

Perrot, V., Polichetti, M., Varray, F., Garcia, D., 2021. So you think you can DAS? A viewpoint on delay-and-sum beamforming. Ultrasonics 111, 106309. https://doi.org/10.1016/j.ultras.2020.106309

Punturo, M., Abernathy, M., Acernese, F., Allen, B., Andersson, N., Arun, K., Barone, F., Barr, B., Barsuglia, M., Beker, M., Beveridge, N., Birindelli, S., Bose, S., Bosi, L., Braccini, S., Bradaschia, C., Bulik, T., Calloni, E., Cella, G., Mottin, E.C., Chelkowski, S., Chincarini, A., Clark, J., Coccia, E., Colacino, C., Colas, J., Cumming, A., Cunningham, L., Cuoco, E., Danilishin, S., Danzmann, K., Luca, G.D., Salvo, R.D., Dent, T., Rosa, R.D., Fiore, L.D., Virgilio, A.D., Doets, M., Fafone, V., Falferi, P., Flaminio, R., Franc, J., Frasconi, F., Freise, A., Fulda, P., Gair, J., Gemme, G., Gennai, A., Giazotto, A., Glampedakis, K., Granata, M., Grote, H., Guidi, G., Hammond, G., Hannam, M., Harms, J., Heinert, D., Hendry, M., Heng, I., Hennes, E., Hild, S., Hough, J., Husa, S., Huttner, S., Jones, G., Khalili, F., Kokeyama, K., Kokkotas, K., Krishnan, B., Lorenzini, M., Lück, H., Majorana, E., Mandel, I., Mandic, V., Martin, I., Michel, C., Minenkov, Y., Morgado, N., Mosca, S., Mours, B., Müller–Ebhardt, H., Murray, P., Nawrodt, R., Nelson, J., Oshaughnessy, R., Ott, C.D., Palomba, C., Paoli, A., Parguez, G., Pasqualetti, A., Passaquieti, R., Passuello, D., Pinard, L., Poggiani, R., Popolizio, P., Prato, M., Puppo, P., Rabeling, D., Rapagnani, P., Read, J., Regimbau, T., Rehbein, H., Reid, S., Rezzolla, L., Ricci, F., Richard, F., Rocchi, A., Rowan, S., Rüdiger, A., Sassolas, B., Sathyaprakash, B., Schnabel, R., Schwarz, C., Seidel, P., Sintes, A., Somiya, K., Speirits, F., Strain, K., Strigin, S., Sutton, P., Tarabrin, S., Thüring, A., Brand, J. van den, Leewen, C. van, Veggel, M. van, Broeck, C. van den, Vecchio, A., Veitch, J., Vetrano, F., Vicere, A., Vyatchanin, S., Willke, B., Woan, G., Wolfango, P., Yamamoto, K., 2010. The Einstein Telescope: a third-generation gravitational wave observatory. Class. Quantum Gravity 27, 194002. https://doi.org/10.1088/0264-9381/27/19/194002

Rajabi, N., Rajabi, O., 2015. Real time earthquake prediction using cross-correlation analysis & transfer function model, in: 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI). pp. 238–242. https://doi.org/10.1109/KBEI.2015.7436053

Ramachandran, P., Zoph, B., Le, Q.V., 2017. Searching for Activation Functions.

Rasamoelina, A.D., Adjailia, F., Sinčák, P., 2020. A Review of Activation Function for Artificial Neural Network, in: 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI). pp. 281–286. https://doi.org/10.1109/SAMI48414.2020.9108717

Riahi, N., Gerstoft, P., 2015. The seismic traffic footprint: Tracking trains, aircraft, and cars seismically. Geophys. Res. Lett. 42, 2674–2681. https://doi.org/10.1002/2015GL063558

Ritzwoller, M.H., Lin, F.-C., Shen, W., 2011. Ambient noise tomography with a large seismic array. Comptes Rendus Geosci. 343, 558–570. https://doi.org/10.1016/j.crte.2011.03.007

Rondenay, S., 2009. Upper Mantle Imaging with Array Recordings of Converted and Scattered Teleseismic Waves. Surv. Geophys. 30, 377–405. https://doi.org/10.1007/s10712-009-9071-5

Rost, S., Thomas, C., 2009. Improving Seismic Resolution Through Array Processing Techniques. Surv. Geophys. 30, 271–299. https://doi.org/10.1007/s10712-009-9070-6

Rost, S., Thomas, C., 2002. ARRAY SEISMOLOGY: METHODS AND APPLICATIONS. Rev. Geophys. 40, 2–1. https://doi.org/10.1029/2000RG000100

Saad, O.M., Chen, Y., 2020. Deep denoising autoencoder for seismic random noise attenuation. GEOPHYSICS 85, V367–V376. https://doi.org/10.1190/geo2019-0468.1

Sabra, K.G., Gerstoft, P., Roux, P., Kuperman, W.A., Fehler, M.C., 2005. Extracting time-domain Green's function estimates from ambient seismic noise. Geophys. Res. Lett. 32. https://doi.org/10.1029/2004GL021862

Saccorotti, G., Maresca, R., Pezzo, E.D., 2001. Array analyses of seismic noise at Mt. Vesuvius Volcano, Italy. J. Volcanol. Geotherm. Res. 110, 79–100. https://doi.org/10.1016/S0377-0273(01)00204-9

Sato, H., Fehler, M.C., Maeda, T., 2012. Seismic wave propagation and scattering in the heterogeneous earth. Springer.

Saygin, E., Kennett, B.L.N., 2010. Ambient seismic noise tomography of Australian continent. Tectonophysics 481, 116–125. https://doi.org/10.1016/j.tecto.2008.11.013

Schimmel, M., Gallart, J., 2007. Frequency-dependent phase coherence for noise suppression in seismic array data. J. Geophys. Res. Solid Earth 112. https://doi.org/10.1029/2006JB004680

Schimmel, M., Paulssen, H., 1997. Noise reduction and detection of weak, coherent signals through phase-weighted stacks. Geophys. J. Int. 130, 497–505. https://doi.org/10.1111/j.1365-246X.1997.tb05664.x

Schippkus, S., Garden, M., Bokelmann, G., 2020. Characteristics of the Ambient Seismic Field on a Large-N Seismic Array in the Vienna Basin. Seismol. Res. Lett. 91, 2803–2816. https://doi.org/10.1785/0220200153

Schippkus, S., Zigone, D., Bokelmann, G., the AlpArray Working Group, 2018. Ambient-noise tomography of the wider Vienna Basin region. Geophys. J. Int. 215, 102–117. https://doi.org/10.1093/gji/ggy259

Schweitzer, J., 2021. Seismometer Arrays, in: Beer, M., Kougioumtzoglou, I.A., Patelli, E., Au, I.S.-K. (Eds.), Encyclopedia of Earthquake Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–11. https://doi.org/10.1007/978-3-642-36197-5_191-1

Schweitzer, J., Fyen, J., Mykkeltveit, S., Gibbons, S.J., Pirli, M., Kühn, D., Kværna, T., 2012. Seismic arrays, in: New Manual of Seismological Observatory Practice 2 (NMSOP-2). Deutsches GeoForschungsZentrum GFZ, pp. 1–80.

Selby, N.D., 2010. Relative Locations of the October 2006 and May 2009 DPRK Announced Nuclear Tests Using International Monitoring System Seismometer Arrays. Bull. Seismol. Soc. Am. 100, 1779–1784. https://doi.org/10.1785/0120100006

Shailaja, K., Seetharamulu, B., Jabbar, M.A., 2018. Machine Learning in Healthcare: A Review, in: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). pp. 910–914. https://doi.org/10.1109/ICECA.2018.8474918

Shapiro, N.M., Campillo, M., Stehly, L., Ritzwoller, M.H., 2005. High-Resolution Surface-Wave Tomography from Ambient Seismic Noise. Science 307, 1615–1618. https://doi.org/10.1126/science.1108339

Shapiro, N.M., Ritzwoller, M.H., Bensen, G.D., 2006. Source location of the 26 sec microseism from cross-correlations of ambient seismic noise. Geophys. Res. Lett. 33. https://doi.org/10.1029/2006GL027010

Sharma, Sagar, Sharma, Simone, Athaiya, A., 2017. Activation functions in neural networks. Data Sci 6, 310–316.

Sharma, V., 2022. A Study on Data Scaling Methods for Machine Learning. Int. J. Glob. Acad. Amp Sci. Res. 1, 31–42. https://doi.org/10.55938/ijgasr.v1i1.4

Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. Phys Rev E 69, 046610. https://doi.org/10.1103/PhysRevE.69.046610

Snieder, R., Gret, A., Douma, H., Scales, J., 2002. Coda wave interferometry, a new method for monitoring change. J. Acoust. Soc. Am. 112, 2319–2319. https://doi.org/10.1121/1.4779351

Spudich, P., Oppenheimer, D., 1986. Dense Seismograph Array Observations of Earthquake Rupture Dynamics, in: Earthquake Source Mechanics. American Geophysical Union (AGU), pp. 285–296. https://doi.org/10.1029/GM037p0285

Stammler, K., Ceranna, L., 2016. Influence of Wind Turbines on Seismic Records of the Gräfenberg Array. Seismol. Res. Lett. 87, 1075–1081. https://doi.org/10.1785/0220160049

Sun, S., Chen, W., Wang, L., Liu, X., Liu, T.-Y., 2016. On the depth of deep neural networks: A theoretical view, in: Proceedings of the AAAI Conference on Artificial Intelligence.

Sun, Y., LAO, D., Sundaramoorthi, G., Yezzi, A., 2022. Surprising Instabilities in Training Deep Networks and a Theoretical Analysis, in: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 19567–19578.

Telgarsky, M., 2016. Benefits of depth in neural networks, in: Conference on Learning Theory. PMLR, pp. 1517–1539.

Thorsten Wuest, C.I., Daniel Weimer, Thoben, K.-D., 2016. Machine learning in manufacturing: advantages, challenges, and applications. Prod. Manuf. Res. 4, 23–45. https://doi.org/10.1080/21693277.2016.1192517

Too, E.C., Yujian, L., Njuki, S., Yingchun, L., 2019. A comparative study of fine-tuning deep learning models for plant disease identification. Comput. Electron. Agric. 161, 272–279. https://doi.org/10.1016/j.compag.2018.03.032

Toshniwal, S., Kannan, A., Chiu, C.-C., Wu, Y., Sainath, T.N., Livescu, K., 2018. A Comparison of Techniques for Language Model Integration in Encoder-Decoder Speech Recognition, in: 2018 IEEE Spoken Language Technology Workshop (SLT). pp. 369–375. https://doi.org/10.1109/SLT.2018.8639038

Viering, T., Loog, M., 2023. The Shape of Learning Curves: A Review. IEEE Trans. Pattern Anal. Mach. Intell. 45, 7799–7819. https://doi.org/10.1109/TPAMI.2022.3220744

Walden, A.T., White, R.E., 1998. Seismic wavelet estimation: a frequency domain solution to a geophysical noisy input-output problem. IEEE Trans. Geosci. Remote Sens. 36, 287–297. https://doi.org/10.1109/36.655337

Walker, K.T., Hedlin, M.A.H., 2009. A Review of Wind-Noise Reduction Methodologies, in: Le Pichon, A., Blanc, E., Hauchecorne, A. (Eds.), Infrasound Monitoring for Atmospheric Studies. Springer Netherlands, Dordrecht, pp. 141–182. https://doi.org/10.1007/978-1-4020-9508-5_5

Wang, J., Schweitzer, J., Tilmann, F., White, R.S., Soosalu, H., 2011. Application of the Multichannel Wiener Filter to Regional Event Detection Using NORSAR Seismic-Array Data. Bull. Seismol. Soc. Am. 101, 2887–2896. https://doi.org/10.1785/0120110003

Wang, Q., Ma, Y., Zhao, K., Tian, Y., 2022. A Comprehensive Survey of Loss Functions in Machine Learning. Ann. Data Sci. 9, 187–212. https://doi.org/10.1007/s40745-020-00253-5

Webb, S.C., 1992. The equilibrium oceanic microseism spectrum. J. Acoust. Soc. Am. 92, 2141–2158. https://doi.org/10.1121/1.405226

Wegler, U., Sens-Schönfelder, C., 2007b. Fault zone monitoring with passive image interferometry. Geophys. J. Int. 168, 1029–1033. https://doi.org/10.1111/j.1365-246X.2006.03284.x

Wihlborg, E., Larsson, H., Hedström, K., 2016. "The Computer Says No!" – A Case Study on Automated Decision-Making in Public Authorities, in: 2016 49th Hawaii International Conference on System Sciences (HICSS). pp. 2903–2912. https://doi.org/10.1109/HICSS.2016.364

Wolfe, P.J., Godsill, S.J., 2003. A perceptually balanced loss function for short-time spectral amplitude estimation, in: 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). p. V–425. https://doi.org/10.1109/ICASSP.2003.1199997

Yan, Q., He, J., Gao, J., 2003. A method for calculating the transfer function of digital seismograph system. Acta Seismol. Sin. 16, 693–698. https://doi.org/10.1007/s11589-003-0053-2

Yang, Y., Ritzwoller, M.H., 2008. Characteristics of ambient seismic noise as a source for surface wave tomography. Geochem. Geophys. Geosystems 9. https://doi.org/10.1029/2007GC001814

Yin, J., Denolle, M.A., He, B., 2022. A multitask encoder–decoder to separate earthquake and ambient noise signal in seismograms. Geophys. J. Int. 231, 1806–1822. https://doi.org/10.1093/gji/ggac290

Zhang, H., Ma, C., Pazzi, V., Zou, Y., Casagli, N., 2020. Microseismic Signal Denoising and Separation Based on Fully Convolutional Encoder–Decoder Network. Appl. Sci. 10. https://doi.org/10.3390/app10186621

Zhong, S., Zhang, K., Bagheri, M., Burken, J.G., Gu, A., Li, B., Ma, X., Marrone, B.L., Ren, Z.J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B.M., Xiao, X., Yu, X., Zhu, J.-J., Zhang, H., 2021. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. Environ. Sci. Technol. 55, 12741–12754. https://doi.org/10.1021/acs.est.1c01339

Zhu, L., Peng, Z., McClellan, J., Li, C., Yao, D., Li, Z., Fang, L., 2019. Deep learning for seismic phase detection and picking in the aftershock zone of 2008 Mw7.9 Wenchuan Earthquake. Phys. Earth Planet. Inter. 293, 106261. https://doi.org/10.1016/j.pepi.2019.05.004

Zhu, W., Beroza, G.C., 2018. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. Geophys. J. Int. 216, 261–273. https://doi.org/10.1093/gji/ggy423

Zhu, W., Mousavi, S.M., Beroza, G.C., 2019. Seismic Signal Denoising and Decomposition Using Deep Neural Networks. IEEE Trans. Geosci. Remote Sens. 57, 9476–9488. https://doi.org/10.1109/TGRS.2019.2926772

Zimmer, V.L., Sitar, N., 2015. Detection and location of rock falls using seismic and infrasound sensors. Eng. Geol. 193, 49–60. https://doi.org/10.1016/j.enggeo.2015.04.007

Acknowledgements

I would like to take this opportunity to express my deepest appreciation to all the people who have supported me on this journey.

Above all, I want to express my special gratitude to **Céline Hadziioannou** for being such a constant source of support and guidance throughout my PhD journey. Your advice and expertise have been invaluable in helping me navigate the challenges of my research and I am especially thankful for the discussions and meetings that kept me on track and provided encouragement during difficult times. I greatly appreciate the way you lead our group, balancing academic excellence and personal support. Your efforts to create a collaborative and engaging environment have greatly enhanced my PhD experience and made our group a truly supportive and enjoyable place to work.

I am also thankful to **Dirk Gajewski** for initiating the idea of this research project, which laid the groundwork for my research. Your contributions have been indispensable, and I appreciate your continued engagement, even after retirement, to offer insights and support contributing to the success of this project.

Many thanks to **Conny Hammer** for guiding me through the machine learning part of my research. Your thoughtful responses to my technical questions and your ability to provide clear, easy-to-understand explanations were truly remarkable. I am deeply grateful for the consistent support and encouragement you offered throughout my journey into the machine learning world. Your guidance not only played a crucial role in shaping the success of my research but also helped me become more enthusiastic and confident in my work.

I would further like to thank **Sven Schippkus** for having provided me with invaluable academic and personal support. Your ability to offer additional guidance at just the right moment was incredibly helpful and allowed me to progress with confidence. I really appreciate your commitment to make time available for discussions and feedback. The thorough comments on my papers, including detailed reading, thoughtful corrections, and restructuring suggestions, were indispensable. All of this greatly enhanced the quality of my work and academic journey.

Additionally, this work would not have been possible without the support from the **seismology working group**. I appreciate our weekly meetings (and of course our Earthcakes), and I am grateful for the collaborative spirit that has grown over time, inside and outside the institute.

I could not have reached this milestone without my **family**. Your love and backing give me the courage to dream big and always follow my heart. For that, I am forever grateful. My grandmother's always crossed fingers are a sweet reminder of the warmth and encouragement that surrounds me, including her own special way of cheering me on.

Above all, thank you **Nils** for being my rock and my soulmate. Your love and support have empowered me to face every challenge with confidence. Thank you for drying my tears, encouraging me, and always making me feel like the proudest person in the room. Your belief in me has been a constant reminder of my worth, not just for my achievements but for who I am. I am truly grateful for everything you do.

Eidesstattliche Versicherung | Declaration on Oath

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated.

If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI.

Jana Klinge

Hamburg, 20.03.2025 Jana Klinge