

PAPER • OPEN ACCESS

A study of deep neural networks for Newtonian noise subtraction at Terziet in Limburg—the Euregio Meuse-Rhine candidate site for Einstein Telescope

To cite this article: Vincent van Beveren *et al* 2023 *Class. Quantum Grav.* **40** 205008

View the [article online](#) for updates and enhancements.

You may also like

- [Newtonian-noise cancellation in large-scale interferometric GW detectors using seismic tiltmeters](#)
Jan Harms and Krishna Venkateswara
- [Newtonian-noise reassessment for the Virgo gravitational-wave observatory including local recess structures](#)
Ayatri Singha, Stefan Hild and Jan Harms
- [Seismic array measurements at Virgo's west end building for the configuration of a Newtonian-noise cancellation system](#)
Maria C Tringali, Tomasz Bulik, Jan Harms *et al.*

A study of deep neural networks for Newtonian noise subtraction at Terziet in Limburg—the Euregio Meuse-Rhine candidate site for Einstein Telescope

Vincent van Beveren^{1,*} , Maria Bader¹, Jo van den Brand^{1,2} , Henk Jan Bulten¹, Xander Campman³ , Soumen Koley^{1,4}  and Frank Linde¹

¹ Nikhef, Science Park, 1098 XG Amsterdam, The Netherlands

² Maastricht University, 6211 LK Maastricht, The Netherlands

³ Shell Global Solutions International B.V., Den Haag, The Netherlands

⁴ Gran Sasso Science Institute (GSSI), I-67100 L'Aquila, Italy

E-mail: v.van.beveren@nikhef.nl

Received 20 April 2023; revised 25 July 2023

Accepted for publication 24 August 2023

Published 18 September 2023



CrossMark

Abstract

The Euregio Meuse-Rhine border region of Belgium, Germany and the Netherlands has been identified as a candidate site for hosting Einstein Telescope. Newtonian coupling of ground vibrations to the core optics of the detectors may limit the sensitivity of Einstein Telescope at frequencies below about 10 Hz. The contribution of Newtonian noise is site specific and depends on the ambient seismic field which in turn depends on the site's geology and the distribution of surface and underground seismic-noise sources. We have investigated the application of machine learning in combination with the deployment of seismic sensor networks to predict seismic displacement noise at specific locations on the surface and underground. Moreover we have modeled a deep neural network (DNN) that allows to subtract Newtonian noise from the strain data measured by Einstein Telescope. The seismic-field model is based on a complete solution of the elastodynamic wave equations for a horizontally-layered soil structure. The geology features soft-soil layers on hard-rock and was shown to be effective in attenuating Newtonian noise from

* Author to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

surface waves below the required sensitivity. In addition our model includes a random background of body waves with all possible angles of incidence. We show that a DNN is effective in predicting Newtonian noise. Application of our DNN allows Newtonian noise subtraction by a factor up to 4.7 at 1 Hz and 2.5 at 5 Hz.

Keywords: gravitational wave detection, Einstein Telescope, Newtonian noise, seismic noise, machine learning, noise cancellation, neural networks

1. Introduction

Einstein Telescope is a planned third-generation subterranean gravitational-wave detector, to be based in Europe [1]. Einstein Telescope will improve upon the sensitivity of current detectors by novel sensor technology, greater seismic isolation and an evolved observatory design. Seismic perturbations in the subsurface around the gravitational-wave detector may limit the sensitivity of the observatory at low frequency. This is due to mass density fluctuations that induce acceleration-noise in the optical components of the detector. This effect, called Newtonian noise or gravity-gradient noise, is expected to dominate [2] the measured detector strain below frequencies of about 10 Hz.

At present there are two candidate sites for hosting Einstein Telescope: the Euregio Meuse-Rhine (EMR) in the border region of the Netherlands, Belgium and Germany, and Sardinia in Italy. An initial study has been carried out to quantify Newtonian noise for the EMR candidate site [3] and results show that for frequencies below 8 Hz Newtonian noise may be about a factor two higher than the design sensitivity of Einstein Telescope (the so called ET-D).

Here we discuss the use of a seismic sensor network to model and subtract Newtonian noise from the detected signal. Noise subtraction can be accomplished when the current state of the seismic field can be inferred from the data of the seismic sensor network, by numerically integrating the change in gravitational attraction on the optical components, induced by this state.

A pragmatic approach to this problem is to apply Wiener filtering or machine learning to find the relation between the seismometer input and detector strain output. Note that this approach eliminates the necessity to calculate the seismic state of the field, which is generally a complex superposition of waves and transients in a non-homogeneous medium. We demonstrate a new approach that uses DNNs to predict seismic- and Newtonian noise activity. Machine learning for estimating the Newtonian noise contribution to a gravitational-wave detector signal is a synthesis problem, in which the machine-learning algorithm attempts to reproduce a Newtonian noise induced signal given a number of seismic input sensors. Once the noise is estimated it can be subtracted from the observed signal, in effect reducing the noise and thus improving the quality of the detector output signal. In contrast to Wiener filters, a neural network can give non-linear solutions based on the information of several sensors. This is a crucial advantage, since the density fluctuations in the soil surrounding a test mass depend on the *gradients* of the displacement field, which change sign when a wave travels in the opposite direction. We will discuss Wiener filtering and compare our neural network results with a Wiener filter in section 7.

Einstein Telescope will be sensitive for frequencies from about 3 Hz to 10 kHz, a range that largely coincides with the audio band. In this band machine learning was applied to increase signal to noise in active noise canceling technology, which uses one or more external

microphones to capture ambient noise and to apply an algorithm to cancel this signal by mixing an estimated anti-signal into the audio stream [4]. In gravitational-wave research studies were done to reduce the technical noise of a detector strain signal by applying machine learning. Vajente *et al* [5] apply an algorithm based on Volterra series to estimate stationary and non-stationary noise sources. The machine learning part relates mostly to using a gradient descent algorithm commonly used in machine learning to optimize the Volterra kernel. Ormiston *et al* [6] explored the application of a DNN called ‘DeepClean’ to improve upon the Wiener filter based noise reduction schemes in current detectors. However, no attempt was made to apply these techniques to Newtonian noise. Badaracco *et al* [7] find the optimal input seismometer placement by using a particle swarm optimizer, facilitating a Wiener filter based Newtonian noise cancellation pipeline.

Newtonian noise background subtraction features the following relevant characteristics: firstly we can assume a relatively invariant ambient environment and a flexible selection of sensor channels and locations. Secondly we can presume that there are no real-time constraints, and that ample processing resources are available. Finally a Newtonian noise subtraction solution can perform *in situ* machine-learning model (re-)training, allowing flexibility and an adaptive approach when the ambient seismic noise environment would change over time.

In the following we discuss how Wiener filters and machine learning can be used to predict seismic displacement noise and Newtonian noise. We first demonstrate the prediction of seismic noise for a surface sensor array for which we have an extensive seismic velocity data set, and where we expect both approaches to provide solid predictions. Next we show that such a surface array is also capable to predict underground seismic noise at a depth of 250 m. Finally, we evaluate the capability of both a Wiener filter and a DNN to predict Newtonian noise for the geology at Terziet in Limburg, the Netherlands. This is a significantly more challenging case as Newtonian noise arises from an integral over all relevant mass-element displacements in a complicated geology extending over several kilometers. As real data from a distributed seismic sensor network are not available, this study is based on synthetic data.

2. Predicting seismic-sensor signals with DNNs

To estimate Newtonian noise, a neural network needs to resolve how seismic waves in the Earth’s subsurface relate to the strain signal measured by a gravitational-wave detector. Before embarking on this challenging task, we first carry out a feasibility study to see if it is possible to create a machine-learning model that can predict the signal of a single seismic sensor, the so called witness sensor, based on the response of all other seismic sensors that are located in the surrounding area. This question was addressed by Liu *et al* [8] who used convolutional neural networks in the context of geophysics analysis for deep-seismic-prior-based reconstruction of seismic data. They analyzed seismic traces from sensors evenly spaced along a straight line (leading to a sub-surface mapping in 2D-geometry). In our analysis, the sensors are placed randomly, at irregular distances both in the north-south and east-west direction, and the phase relations between channels are derived by the neural network itself during training.

We employ a parameterizable DNN and the implemented setup is depicted in figure 1. Once the network has been trained on an adequate amount of training data, the network is used to predict the signal of the witness sensor when using yet unseen input data. The goal is to demonstrate that the seismic activity can be retrieved (also in the time domain) at locations that are not directly monitored by a sensor. It is not assumed that the network can predict

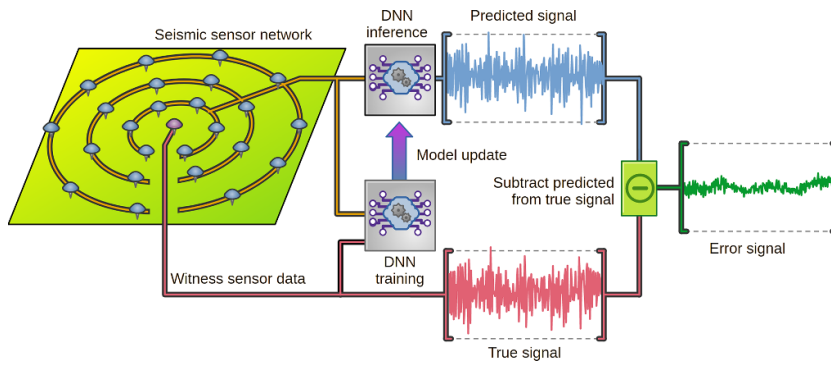


Figure 1. Example seismic network for predicting the seismic displacement signal of a witness sensor (denoted by the red sensor in the center of the network).

seismic activity at any arbitrary location without training; note that this is not a necessary requirement to subtract Newtonian noise as we will demonstrate in section 5.

2.1. Preprocessing

A parameterizable DNN termed ‘Seismonet’ was developed by using Keras [9] and Google Tensorflow [10] to solve the problem of predicting seismic activity for a witness sensor, based on data from other sensors in a seismic network.

While seismic sensors, such as geophones, provide one or more channels of displacement data in the time-domain, Seismonet operates in frequency domain, and signal transformation and signal conditioning are required. The seismic sensor data of all channels are digitized at a sample rate specific for the sensor. In the first preprocessing step, all sample data are decimated to a single sample-rate, leading to a fixed Nyquist frequency, which value is dependent on the application. Furthermore instrument response is removed, providing N time-series signals in velocity (m s^{-1}) $z_n(t)$ where n represents an index for a specific channel, t the time index, and N the total number of channels. The witness (or target) signal z_0 is the signal we wish to predict based on the data $z_{1\dots(N-1)}$ of $N - 1$ so-called reference channels.

Hereafter each time-series is scaled with a channel specific normalization factor as

$$z'_n = \frac{z_n}{2\text{Med}|z_n|}, \quad (1)$$

where Med is the median operation. Once chosen, the normalization factor should remain constant, but it can be determined from a representative subset of all channel data. Normalization such that most values are within the range of $-1 \dots 1$ is beneficial to DNN performance and common in machine learning preprocessing pipelines.

The total dataset has to be segmented in records of fixed duration. Each record contains s samples of data for all channels z'_n . To arrive at an optimal value for s , we consider the maximal lag between the witness sensor and any reference sensor in the field for which one can expect sizable coherent contributions. This maximal lag depends on the spatial separation between the sensors and the velocities v with which waves travel through the field. Generally for seismic waves at the surface, the group velocity is lower than the phase velocities in the wave packet, and all velocities decrease with increasing frequency. The lowest relevant velocity v_{\min} is then mainly determined by the Rayleigh waves at the Nyquist frequency. Longer travel

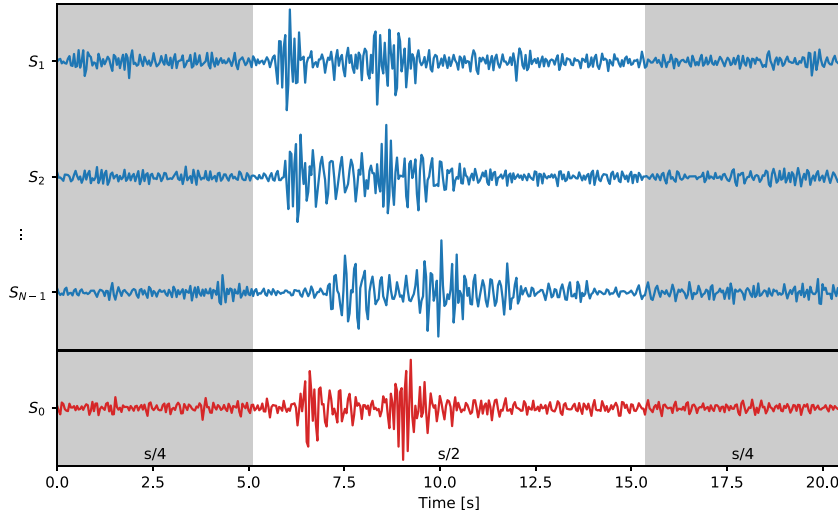


Figure 2. Example of a record S showing the witness channel in red and the reference channels in blue. Signals in the shaded regions of length $s/4$ may receive contributions from previous or next records. All signals are windowed.

times between sensors can be neglected. The maximal lag τ_{\max} in the correlation that needs to be considered can therefore be estimated as

$$\tau_{\max} = \frac{d_{\max}}{v_{\min}}, \quad (2)$$

where d_{\max} is the maximal distance between the witness sensor and the furthest edge of the sensor network. Correlations need only be considered for time lags smaller than or equal to τ_{\max} . We choose the record length s to equal

$$s = 4 \lceil f_s \tau_{\max} \rceil_2, \quad (3)$$

where $\lceil \dots \rceil_2$ denotes rounding up to the nearest power-of-two (which is numerically convenient when using fast Fourier-transform calculations).

An example of a (time domain) record S is shown in figure 2. The input of the neural network is focused on the central $s/2$ samples. As signals in the shaded regions of length $s/4$ may receive contributions from previous or next records, all signals are windowed. Partly overlapping records are obtained by using a stride of $s/2$ samples. For our application, the chosen value of s is optimal, since longer records and higher values of s would increase the parameter size of the neural network and degrade the numerical performance.

Seismonet operates in the frequency domain and thus the time domain record S needs to be transformed to its spectral representation R by using a discrete Fourier transform (\mathcal{F}) as

$$R_n(i) = \mathcal{F}(S_n(i)w_n). \quad (4)$$

Here, i and n represent the record and channel index, respectively and w_n represents a window function. The window function used depends on whether it is a witness channel ($n=0$) or reference channel ($n \in \{1 \dots N-1\}$) and is given as

$$w_n = \begin{cases} \text{padded Tukey}(s, 0.5) & \text{if } n = 0 \\ \text{Hann}(s) & \text{otherwise.} \end{cases} \quad (5)$$

For reference channels we mitigate spectral leakage and improve the spectral resolution by application of a Hann window. For the witness channel we concentrate the predictive capabilities of the DNN on the coherent time span. To achieve this, we utilize a padded Tukey window function applied to the witness channel data. The padded Tukey window has a value of 1 for the central $s/2$ portion of the window and smoothly tapers to 0 outside of this region. This approach not only reduces spectral leakage, and focuses the DNN's predictions on the coherent time span, but also avoids that the prediction is degraded as the reference channels may not have enough history and/or future samples to make accurate predictions for the witness channel.

In the following we denote the witness channel with y , such that $S_0 = S_y$ and $R_0 = R_y$, and the reference channels with x such that $S_{\{1\dots N-1\}} = S_x$ and $R_{\{1\dots N-1\}} = R_x$.

2.2. DNN for predicting seismic displacement noise

Each record $R(i)$ contains all needed context for seismic activity prediction in the coherent region of the witness sensor. The records are fed to Seismonet and figure 3 shows that Seismonet is composed of three segments. The first 'band correlation' segment consists of the input layer and a number of sequential layers that process frequency-domain data in a number of overlapping bands. Each consecutive layer reduces the number of parameters (i.e. the number of neurons and their respective weights) for that specific band, guiding the optimizer to abstract the in-band correlations. The second 'full correlation' segment consists of a number of fully connected layers, allowing global abstractions. For this segment the number of neurons in the first layer and the subsequent neuron reduction per layer can be configured, though the last layer always contains the same number of neurons as the number of output values. Both 'band correlation' and 'full correlation' segments contain dropout layers [11] in between their layers and both use the exponential linear unit ELU [12] as activation function. The final segment contains a single layer whose only purpose is to translate and scale the output of the 'full correlation' segment to the output range. This layer has no activation function.

While the global composition of Seismonet is fixed, it can be customized by hyperparameters, facilitating to adapt Seismonet to a certain bandwidth and to a specific number of channels. The following hyperparameters are available for customization: the width of each overlapping band, the overlap ratio between each band, the number of neurons in the first layer, the number of band correlation layers, the dropout factor used for band correlation layers, the ratio of reduction of each subsequent band correlation layer, the number of full correlation layers, the width and shape of the full correlation layers, and the drop-out factor between each full correlation layer.

Finding optimal values for such a large parameter space cannot easily be done manually and for this reason hyperparameter optimization (HPO) [13] is used to find a near-optimal set of values for a given training and validation data set.

The predicted output of Seismonet for record i is denoted $R_{\hat{y}}(i)$, and represents Seismonet's inference for a specific training record. Seismonet uses the mean absolute error as loss function. A more comprehensive description of Seismonet is provided in [appendix](#).

2.3. Post-processing

Prediction $R_{\hat{y}}(i)$ of Seismonet is a complex frequency spectrum which can be converted to the time-domain by using the inverse Fourier transform $S_{\hat{y}}(i) = \mathcal{F}^{-1}(R_{\hat{y}}(i))$. It is possible to

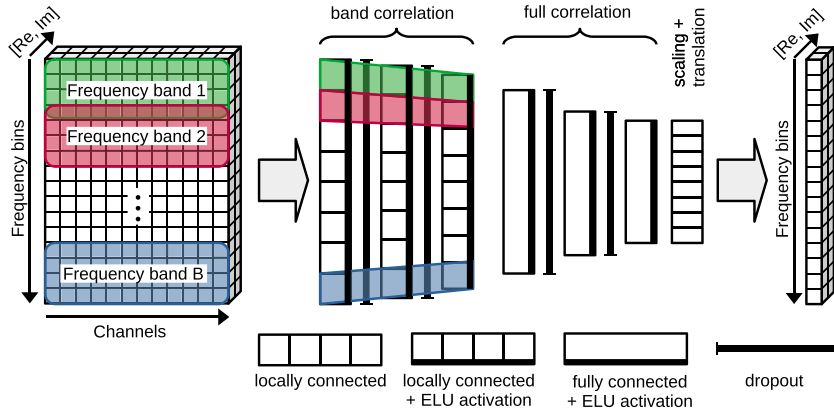


Figure 3. DNN used for seismic excitation prediction. The first three funnel-shaped layers process the seismic input data per-band. The middle two layers allow cross-band correlations. The final layer is used for scaling.

reconstruct the full prediction in the time-domain from $S_{\dot{y}}(i)$ by merging multiple consecutive predictions. Merging requires overlap between records, where each overlapping record then needs to be merged to its neighbor with a power-preserving cross-fade. For this paper we have not reconstructed a continuous time-trace, and the results only show single record predictions.

3. Estimating ambient surface displacement noise

Validation of Seismonet for the prediction of seismic surface displacement noise has been carried out by using data of a 2017 seismic survey [14] executed in the south of Limburg, near the village of Terziet, the Netherlands. The PSD at the surface and at 250 m depth are shown in figure 4. For reference Peterson’s new low-noise model and new high-noise model are shown as dashed curves [15].

3.1. Sensor network topology

The employed seismic sensor network consists of two arrays of Quantum sensor nodes [16], that are placed on the surface in expanding concentric circles. Each node provides a single channel of seismic activity data in the vertical direction. The topology of the seismic network is shown in figure 5. Each subsequent circle has about twice the distance from the array center to ensure network sensitivity to a wide band of wavelengths and to avoid spatial aliasing. The sensor at the center of Array A (depicted by the green triangle in figure 5) acts as the witness sensor for which the neural network will estimate the seismic activity.

3.2. Input data

Each node provides a single channel of sensor data with a sample rate of 250 samples per second and 24 bits of velocity noise resolution. Since we are only interested in frequencies below 10 Hz, all data are first converted to IEEE-754 32-bit float, decimated to $r = 25$ samples per second (and thus $f_{\text{Nyquist}} = 12.5$ Hz), and filtered with an 8th-order zero-phase type-I Chebychev low-pass filter, with -3 dB at 10 Hz. Data are corrected for the

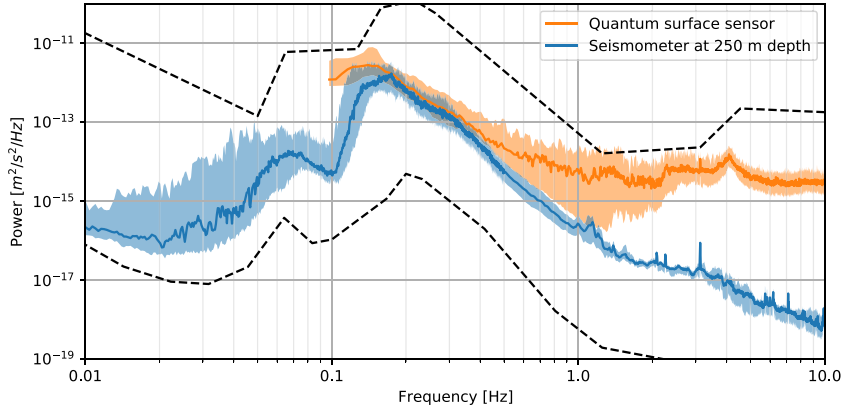


Figure 4. Power spectral density at the surface as measured by a quantum sensor and at 250 m depth by a seismometer. The colored bands represent the 10–90th percentile. Peterson’s NLNM and NHHM are shown as dashed curves.

transfer function of the 5 Hz geophone-based sensors and all channels were scaled with their channel specific normalization factor.

The distance between witness sensor and the edge of the network equals about $d_{\max} = 0.5$ km. For a lower-bound velocity detection threshold of $v_{\min} = 100 \text{ m s}^{-1}$, the required record length according to equation (3) equals $s = 512$ samples.

The data set consists of 12 days of sample data from 67 participating sensors. We employ a 75% overlap, and the final data set consists of 202432 records of which 80% are used for training.

3.3. Training

Training was executed on an Nvidia Tesla V100 with the Nadam neural network optimization algorithm [17] with a learning rate of 10^{-5} . To find the most optimal model configuration Hyperband [18] HPO with Keras Tuner [19] has been applied. The resulting model consist of over 66M trainable weights and contains 18 input bands of 778 mHz width that have an 11% overlap. Convergence was reached when the validation loss did not decrease for 90 epochs, and this condition occurred after 265 epochs.

3.4. Results

A representative validation record of the vertical surface seismic velocity information in the time-domain with actual and predicted signals is shown in figure 6(a). The seismic velocity information estimated by the DNN ($S_{\hat{y}}$) accurately matches the signal measured by the witness sensor (S_y). Note that these time-domain data are high-pass filtered with $f = 584$ mHz and low-pass filtered at 10 Hz.

The performance of the Seismonet model in the frequency domain is shown in figure 6(b). The figure shows that the median power of the true and predicted signals are nearly identical in the frequency domain. The bands represent the 10–90 percentile, and the data are low-pass filtered at 10 Hz. A phase mismatch between actual and predicted signal would not show up in these spectra. The quality of the prediction should be judged by considering the residual

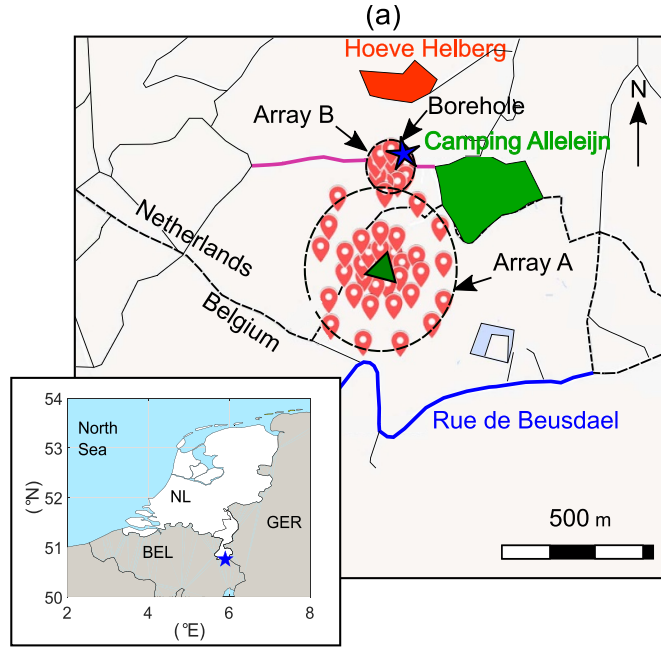


Figure 5. Layout of Array A and Array B showing seismometers laid out on rings of different radii on a map of Terziet, Limburg. The borehole at the site is marked with a star. The witness sensor at the center of Array A is at GPS coordinates 50.753 158, 5.905 013 and is shown by the green triangle. Map data © 2020 Google.

power $(|R_y - R_{\hat{y}}|^2)$ that remains after subtracting the DNN-prediction from the data. As shown in figure 6(c), the residual power is more than a factor 20 smaller than the uncorrected power in the frequency band between 3 and 10 Hz. At lower frequencies the small size of the seismic sensor array limits the spatial resolution. Together with increased instrument noise this limits the ability of the neural network to make accurate predictions.

3.5. Comparison to a Wiener filter approach

Wiener filters [20] are widely used in current gravitational wave research. In a Wiener filter one subtracts the correlated noise from a signal in an optimal manner, provided that the influence of the noise is linear. In our application we want to subtract the acceleration noise δa , induced by seismic activity, by using seismometer signals s_i . The Wiener filter employs the measured cross-correlations between all the sensors and accelerations to minimize the noise in δa ; i.e. it subtracts the correlated part of the signals between the seismometers and acceleration. For finite, discrete and real-valued data sets the circular cross-correlation between data streams s_i and s_j is defined as

$$(s_i \otimes s_j)(\tau) = \sum_{k=0}^{N-1} s_i([k - \tau]_{\text{mod } N}) s_j(k), \quad (6)$$

where the index k runs over all time samples and the subscript $\text{mod } N$ indicates modulo N . It is assumed that the finite discrete data set can be described as periodic with a period of N

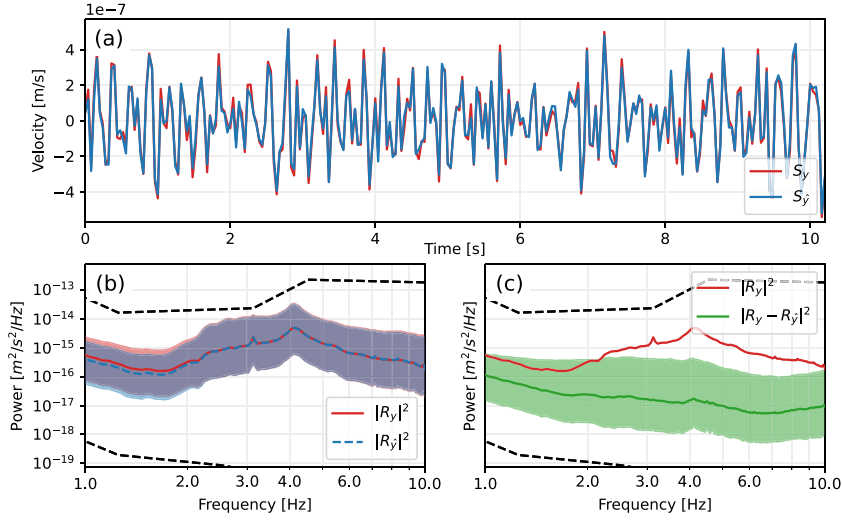


Figure 6. Panel (a) shows the time-trace of a typical single validation record of the vertical surface seismic velocity information recorded by the witness sensor (S_y). The seismic velocity information estimated by the DNN ($S_{\hat{y}}$) is in good agreement with the measured signal. Note that all data are high-pass filtered with $f = 584$ mHz and low-pass filtered at 10 Hz. Panel (b) shows the median power of the true signal (red curve) and the predicted signal (blue curve). The bands represent the 10–90 percentile, and the data are low-pass filtered at 10 Hz. For reference Peterson’s NLNM and NHHM are provided as dashed curves. Panel (c) shows the residual surface seismic velocity information when the DNN prediction is subtracted from the recorded data. For reference the median true signal is provided in red color.

samples. The correlation vector $(s_i \otimes s_j)(\tau)$ is part of a Fourier-pair; the values in the time domain can be calculated from

$$(s_i \otimes s_j)(\tau) = \mathcal{F}^{-1}(S_i(f) S_j^*(f)) \quad (7)$$

with \mathcal{F}^{-1} the inverse fast-Fourier transform operation, $S(f)$ the data stream in frequency domain ($S(f) = \mathcal{F}(s(t))$), and S^* the complex conjugate of S . This allows one to build up the normalized cross-correlation matrix $\text{Corr}_{ij}(f)$ in Fourier-domain by summing data from different time segments as

$$\text{Corr}_{ij}(f) = \frac{\sum_{m=1}^{N_{\text{segments}}} S_{i,m}^*(f) S_{j,m}(f)}{|S_i(f)| |S_j(f)|} \quad (8)$$

with $|S(f)| = \sqrt{\sum_{m=1}^{N_{\text{segments}}} S_m(f) S_m^*(f)}$.

For the Wiener filter, the prediction for the acceleration in the x -direction in frequency domain $a_x(f)$ is made from the signals in sensors $S_i(f) = \mathcal{F}(s_i(t))$ as

$$\delta a_x(f) = \sum_i C_i(f) S_i(f), \quad (9)$$

with the filter coefficients $C_i(f)$ given by

$$C_i(f) = \frac{|a_x(f)|}{|S_i(f)|} \sum_j \text{Corr}_{ij}^{-1}(f) \text{Corr}_{ja}(f). \quad (10)$$

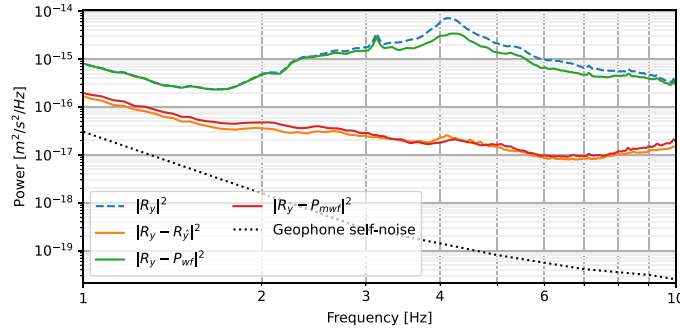


Figure 7. The PSD of actual noise (labeled R_y) is represented by the blue dashed curve. The residual noise when applying a DNN to the data is shown by the orange curve and R_j represents the prediction of Seismonet. The green and red curves represent the residual power when using the unmodified Wiener filter P_{wf} and modified Wiener filter P_{mwf} , respectively. In the latter, seismic data segments with outliers are not used for filter construction. The geophone self-noise is shown by the dotted curve [21].

Here $\text{Corr}_{ja}(f)$ represents the vector of normalized cross-correlations for frequency bin f between sensor j and acceleration a_x , $\text{Corr}_{ij}^{-1}(f)$ the inverse of the normalized cross-correlation matrix $\text{Corr}_{ij}(f)$, and $|a_x(f)|$ and $|S_i(f)|$ the average values for the amplitudes (the square root of the average power) of channels S_i and a_x .

Next we show that both Wiener filters and DNNs are highly capable in predicting the displacement field measured by a witness sensor by using the data from surrounding sensors. For Wiener filters this is the case as wave speeds are constant and the displacement field is continuous, so the optimal linear combination of a set of sensors surrounding the reconstructed sensor should have a high overlap with the witness sensor. The construction and application of the Wiener filter is done with the same data and segmentation as used for the neural network, where training data are used for construction of the filter and test data used for prediction. Sudden local excitations in one of the input channels are excluded as these may greatly reduce the Wiener filter's predictive capabilities. For this reason we drop seismic data segments from the training data if the RMS of one of the channels exceeds 3σ with respect to the median RMS of the previous seven hours of seismic data for that channel.

The PSDs of residual noise when using a Wiener filter (with and without outlier removal) and DNN are compared in figure 7. Good performance in the Wiener filter approach was obtained after removal of outliers from strong transient events (P_{mwf}), while such a procedure was not required for the DNN.

4. Estimating underground seismic displacement noise

We have shown that both a Wiener filter and Seismonet are capable of estimating surface seismic displacements. Next we study the performance for estimating underground seismic displacements.

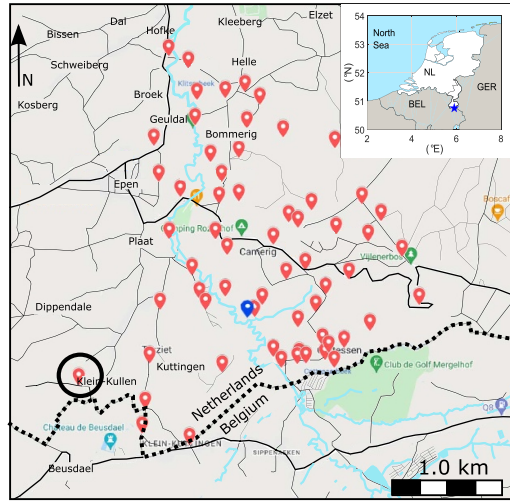


Figure 8. Seismic sensor array used to estimate vertical underground seismic displacement noise at a depth of 250 m. Red pins indicate sensor nodes, the blue pin a permanent seismometer located at Heimansgroeve at a depth of 3 m. The center of the black circle shows the location of the Terziet borehole with a 3C sensor station at the surface and one at a depth of 250 m. The vertical displacement information of the borehole seismometer is used as witness channel. Map data © 2021 Google.

4.1. Sensor network topology

A sparse seismic surface array in combination with a borehole sensor was used to predict vertical seismic activity at 250 m depth. The total number of input channels is 66 and their layout is shown in figure 8. For the sparse array two weeks of survey data were obtained from a seismic survey executed in South Limburg, the Netherlands, which lasted from 14 to 27 November 2020. Survey data from 60 Quantum sensors, located within 3.5 km from the borehole, were used. In addition data from three seismometers were added, that provide 3-channel data. The first station is located at seismic station Heimansgroeve [22], the Netherlands at a depth of only 3 m. The other seismometers are located at Terziet with a Nanometrics Trillium-240 seismometer placed at the surface and a Streckeisen STS-5A seismometer placed at the bottom of a 250 m deep borehole; the PSD of the latter is shown in figure 4. The vertical seismic displacement information of the borehole sensor acts as the witness sensor for the neural network. The horizontal channels of the borehole sensor are not used.

The network spans about $d_{\max} = 3.5$ km but is less-than-optimal considering the off-center location of the witness sensor. Moreover, due to the sparseness of the array we expect only reasonable results at frequencies below 500 mHz. For this reason all data are decimated to a sample rate of 1.25 samples per second. Assuming $v_{\min} = 200$ m s⁻¹ and applying equations (2) and (3) we obtain $s = 128$ samples per record per channel.

The data set consisted of 14 days of sample data from 67 channels. We employed an 87.5% overlap between each record resulting in a data set of 94 336 records. Data from sensors were first collected and decimated to 5 samples per second. Finally the data were low-pass filtered and decimated to 1.25 samples per second with an 8th-order zero-phase type-I Chebyshev with the -3 dB point at 0.5 Hz. Instrument response was removed and all channels were scaled with their channel-specific normalization factor. An optimal Seismonet model with 32.5M trainable

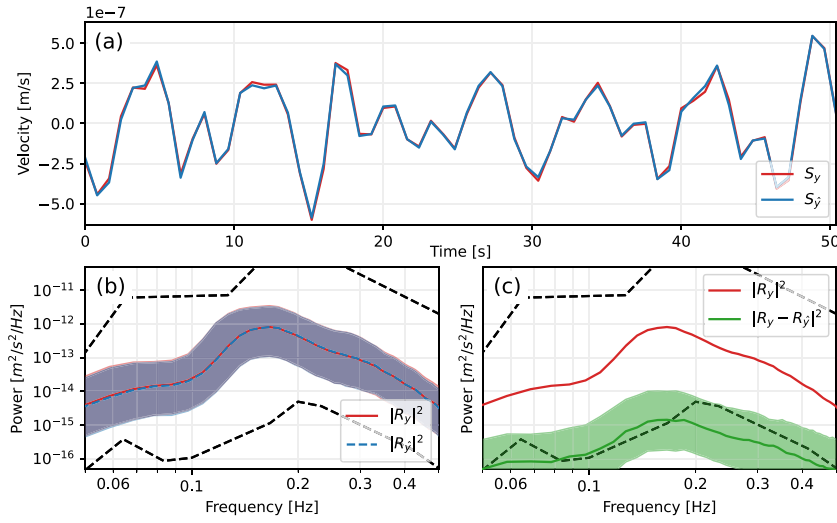


Figure 9. Panel (a) Time-trace validation record of the underground vertical seismic velocity information (S_y) recorded by the witness sensor at 250 m depth. The seismic velocity information estimated by the DNN (\hat{S}_y) is also shown and matches the measured signal reasonably well. All data are high-pass filtered with $f = 29$ mHz and low-pass filtered at 0.5 Hz. Panel (b) shows the median of the true power (red curve) and the predicted signal (blue curve) for the sensor at 250 m depth. The bands represent the 10-90 percentile and data are low-pass filtered at 0.5 Hz. For reference Peterson's NLNM and NHNM are provided as dashed curves [15]. Panel (c) shows the residual underground seismic vertical DNN information after subtracting the prediction from the recorded information. For reference the median true signal power is also provided (red curve).

weights was found with Hyperband HPO. The resulting model contains 32 input bands of 29 mHz width and with 66% overlap.

4.2. Results

The validation record presented in figure 9(a) shows the measured seismic vertical velocity from the witness sensor at a depth of 250 m and the prediction from the neural network. Again good consistency between the two wave-forms in the time domain is evident. Figure 9(b) shows both the predicted and recorded power of the vertical seismic velocity at 250 m depth in the frequency domain. The residual power is shown in figure 9(c); it is about two orders of magnitude lower than the uncorrected power in the whole frequency band. Note that the employed one-sided sparse seismic network was not optimized. More accurate and broad-band results can be expected for an optimized seismic network topology, particularly if underground sensors are included.

We compare the performance of the neural network to that of the Wiener filter in figure 10. Both the DNN and the modified Wiener filter perform equally well. In contrast to the case of the surface array survey discussed in section 3, the unmodified Wiener filter performs relatively well. This can likely be attributed to the inclusion of seismometers which have lower noise floor and are better shielded from external events such as wind and rain. Outside of the microseismic peak the DNN seems to perform better. Whether this is due to the lower signal to noise ratio or due to the presence of non-linear effects has not been explored.

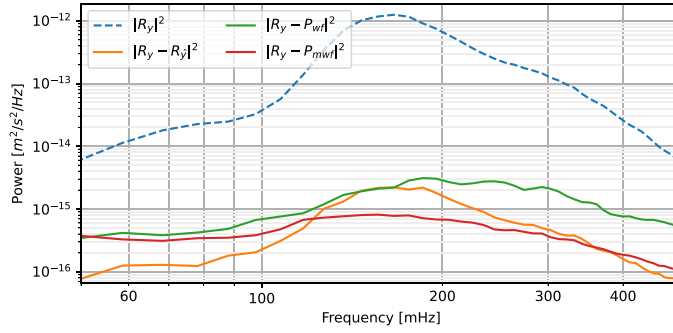


Figure 10. The PSD of actual noise (labeled $|R_y|^2$) is represented by the blue dashed curve. Seismonet's prediction for R_y is denoted $R_{\hat{y}}$ and the power of the residual noise is represented by the orange curve. The green and red curves show the residual noise of the predictions P_{wf} and P_{mwf} for the Wiener filter and the modified Wiener filter, respectively. In P_{mwf} , records with outliers have been removed from the dataset.

5. Newtonian noise subtraction

Subtraction of Newtonian noise requires knowledge of both the mass-density distribution of the local geology and the time-dependent seismic displacement field. The mass-density distribution can be obtained from passive and active seismic surveys, while information on the seismic displacement field can be obtained from a sensor network. Studies have shown [3] that in order to obtain accurate estimates of Newtonian noise, these distributions must be integrated over kilometer-scale distances from the test masses. For the geology at Terziet it was estimated that integration radii up to 2 km are required to capture more than 90% of the Newtonian noise, due to the contributions of a body-wave background.

The previous sections have shown that both a Wiener filter and neural network can be used to predict the seismic displacement field at specific locations. However it is impractical to characterize this field up to large distances and integrate this field in order to estimate Newtonian noise. Newtonian noise background subtraction requires a solution that goes beyond what was presented in figure 1 and where the actual interferometer strains are used as witness channels.

We envision that a permanent seismic sensor network will be installed surrounding each corner station as is shown in figure 11. Sensors are placed on the surface and in the tunnels of the observatory. This minimizes the construction effort and avoids the need for additional boreholes. For frequencies below about 10 Hz the strain outputs of the low-frequency interferometers of Einstein Telescope contain a sizable contribution from Newtonian noise, and this contribution *de facto* results from an integral over all relevant mass-element displacements. The difference between the predicted interferometer strain signal and the true signal will be minimized by the neural network.

5.1. Synthetic data generation

Wiener filter and Seismonet models can be trained on real data, and for Newtonian noise subtraction to be effective, the actual detector strain data must contain a significant contribution from Newtonian noise. This will be the case for the low-frequency interferometers of Einstein Telescope, where Newtonian noise is expected to dominate the strains for frequencies from 2 to about 8 Hz. Here the assumption is made that technical noise contributions (e.g. control noise) are not greatly limiting the sensitivity.

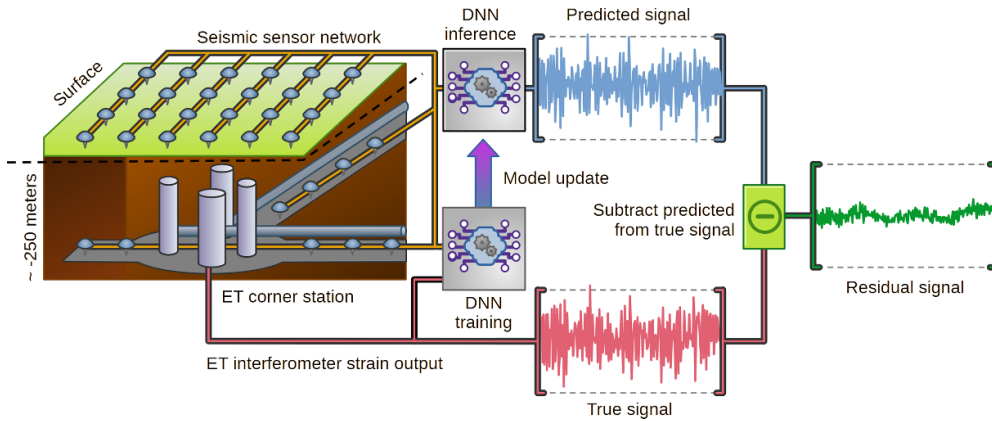


Figure 11. Overview of a machine-learning solution for Newtonian noise subtraction. A seismic-sensor network surrounds each test mass and has sensors both on the surface and underground at the Einstein Telescope (ET) site. The seismic network output and the interferometer output are used to train a machine-learning algorithm for mapping seismic network output to the interferometer output. Once the model is trained it can be used to predict the interferometer output given a seismic network output. This signal can be subtracted from the true output of the interferometer. For a Wiener filter approach the two boxes labeled ‘DNN’ would be replaced by a box labeled ‘Wiener filter cross-correlation’.

Data from an array of seismic sensors surrounding the main mirrors are needed. At present such data do not exist. Therefore, synthetic data are used that are derived from the model by Bader *et al* [3]. This model is schematically shown in figure 12 and incorporates both the geology at Terziet in Limburg and seismic noise PSDs that were measured with a surface sensor array and 3C seismometers at the surface and at 250 m depth. The numerical calculations provide the synthetic displacement field around the test mass. This allows to calculate the Newtonian accelerations on the mirror. Einstein Telescope features three low-frequency interferometers and thus three strain signals. The strain signal of each interferometer is due to the noise power of four mirrors.

In our analysis we assume the seismic noise at each corner station to be independent. Note that in reality the contributions due to Newtonian noise will be correlated in the case of three co-located interferometers, as each corner station hosts the input-mirrors of one interferometer, but the end-mirrors of the other interferometers. How such correlations can be exploited to minimize the effect of Newtonian noise remains to be investigated.

The displacement field has been calculated with the Matlab Elastodynamics Toolkit EDT [23] by using isotropically distributed excitations in a radius of 6 km around the center of the model space. Sources have random location, direction, phase and Gaussian distributed amplitude, and the strength of the sources is matched to reproduce the mean PSD at the surface. In the numerical calculations, we integrate the seismic displacement fields to a radius of 4 km around the test mass. Other observables as reported in the fore-mentioned survey from Koley *et al* [14], such as the frequency-dependent reduction of seismic amplitude with depth and the wave speeds and observed overtones in the beamforming data, are qualitatively reproduced by our model.

Apart from anthropogenic noise, we identified the presence of a body-wave background of which the strength was measured by analyzing day-night ratios for the PSDs measured at

the surface and at 250 m depth [14]. This body-wave background is currently poorly known; we only have data from a single tri-axial sensor at 250 m depth. It was modeled by randomly generated plane waves, isotropic in displacement and in traveling direction. No rescattering, reflections, damping, and/or dispersion is included, although the density of the rock layers is taken into account. The body waves feature a P-wave component traveling at 4 km s^{-1} and an S-wave component at 2.8 km s^{-1} . A more realistic model with better parameters and going beyond plane waves would require input data from a survey with more sensors at depth.

The model is used to generate 50 000 training records by adding 20 randomly distributed surface sources and 4 random body waves. Using a combination of 20 sources for each training record leads to a more realistic description, since at any given moment in time the activity in the field cannot be described by the action of a single source. The oscillations in the subsurface are recorded by a virtual mirror which is placed at position $(0, 0, -250) \text{ m}$, enclosed in a 10 m radius cavern, in the model space and by a number of virtual seismic sensors. For the virtual mirror only the x -acceleration $a_{\text{true}}^{\text{NN}}$ is derived corresponding to the signal to be predicted. In addition the model contains two types of virtual 3C seismic displacement noise sensors, 132 geophone-based sensors and 3 seismometers. Both types of sensors provide the same seismic displacement data, but with different frequency-dependent sensor self-noise. For the geophone-based sensors, the self noise $n \text{ [m}(\sqrt{\text{Hz}})^{-1}]$ can be parameterized as

$$n \text{ [m}(\sqrt{\text{Hz}})^{-1}] = \frac{8.72 \times 10^{-10}}{f^{3.19}} - 2.45 \times 10^{-13}f + 4.63 \times 10^{-12}, \quad (11)$$

with frequency f in Hz [24]. For the underground seismometers the self noise is an order of magnitude less. The position of the sensors in the model space is shown in figure 12. Note that the network topology is by no means optimized, but solely serves to demonstrate the value of a Wiener filter or DNN for Newtonian noise subtraction. Each record contains seismic and Newtonian noise in a limited amount of time (few seconds) for a single frequency and for 136 sensor locations. The synthetic data must be preprocessed for suitability for the DNN. During preprocessing sensor-specific self-noise is added to the synthetic data. All data are in the frequency domain, and belong to a specific frequency band (1, 3, 5 and 10 Hz).

The neural network will be equally capable to reconstruct the acceleration of a single mirror when trained on this parameter as to subtract Newtonian noise in the case of three co-located detectors in an equilateral triangular layout when trained on combinations of the strains of these detectors.

This can be understood by considering three interferometers with the input and end mirrors of the Fabry–Perot cavities located in corner stations labeled 1, 2 and 3. The three interferometers provide three strain measurements S_1, S_2 and S_3 that measure the arm length differences $(\Delta A - \Delta C), (\Delta B - \Delta A), (\Delta C - \Delta B)$, respectively. We choose an orthonormal, right-handed coordinate system with the direction of the x -axis pointing from corner 1 towards corner 2 (along arm A) and the z -axis pointing downwards, towards the center of the earth. We can write the strains as differences in mirror displacements in the x and y horizontal directions (we ignore the small effects of displacements in the vertical direction); e.g.

$$\begin{aligned} S_1 &= (\Delta A - \Delta C) \\ &= (\delta x(2) - \delta x(1)) - \left(\frac{1}{2}(\delta x(3) - \delta x(1)) + \frac{\sqrt{3}}{2}(\delta y(3) - \delta y(1)) \right), \end{aligned} \quad (12)$$

where higher-order corrections of the order (δ^2/L) with δ the mirror displacement and L the arm length are neglected. It follows that the strain signal S_1 depends on 5 variables: the mirror displacements in the x - and y -directions for the mirrors at location 1, 2, and 3. This may seem a more difficult task for the neural network than finding the acceleration of a single mirror

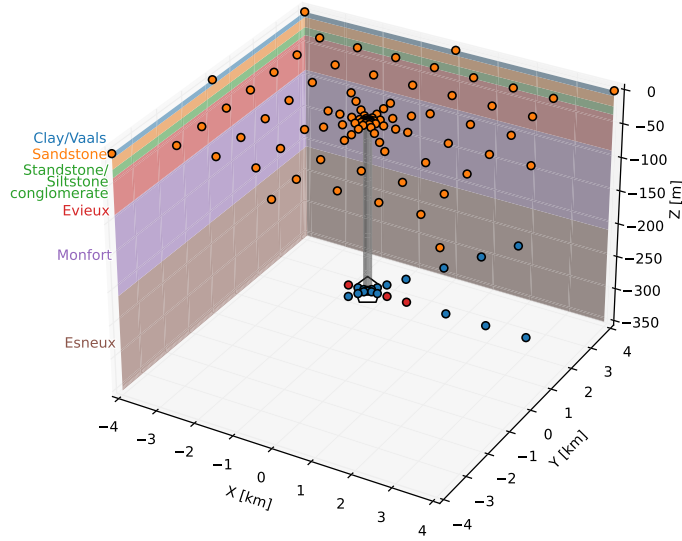


Figure 12. Surface and underground sensor arrays employed in the DNN model that is used to predict mirror acceleration due to Newtonian noise. Circles represent the 3D distribution of seismic sensors. An array of three-component geophones is placed on the surface (orange), and at 250 m depth a combination of geophones (blue) and three-component seismometers (red) follow the tunnels of Einstein Telescope. The pentagon shows the location of a mirror which acts as the witness sensor and is enclosed in a 10m radius cavern. The modeled soil layers are indicated by colored bands. The gray cylinder represents the boundary of the vertical access shaft of a corner station of Einstein Telescope.

in the x - and y -direction (2 independent parameters). However, these five variables are not independent; only one component of the displacement vector in each of the corners contributes. For S_1 for instance, the relevant displacements for mirror 1 are in the horizontal direction with an azimuthal angle of 120° (displacements with directions parallel to arm B, i.e. the combination $-\frac{1}{2}\delta x(1) + \frac{1}{2}\sqrt{3}\delta y(1)$), for mirror 2 in the direction of arm A, ($\delta x(2)$) and for mirror 3 in the direction of arm C, at an azimuthal angle of 60° (the combination of $\frac{1}{2}\delta x(3) + \frac{1}{2}\sqrt{3}\delta y(3)$). With a triangular lay-out, we have three different strain signals. Each of those strain signals depends on three independent variables, namely the displacements of the mirrors in a direction parallel to an arm; which arm depends on which strain signal is calculated. Thus, the accelerations in the corners contribute with different weights to the strains S_1 , S_2 , and S_3 .

In the current study, we train the network to reproduce the acceleration of the test mass in the horizontal direction (a_x and a_y). In our model the noise in the x - and y -direction are uncorrelated, and the neural network must be able to minimize the uncertainty in 2 independent horizontal directions. This is tested by checking if the neural network is equally capable to predict the acceleration of an underground mirror in the x - or y -direction both separately and simultaneously on the basis of the seismic activity measured in our field. We expect that a neural network that is capable of reconstructing the independent accelerations of a mirror at a given location is equally capable of reconstructing three strain signals, each depending on the mirror accelerations at three different locations; although each strain signal depends on three degrees of freedom, these degrees of freedom are constrained by having three different

strain measurements (e.g. the mirror displacement $\delta x(1)$ contributes with different factors to the strains S_1 , S_2 , and S_3 and the neural network should be sensitive to this fact).

6. Newtonian noise and Wiener filtering

6.1. Analytical considerations

Newtonian noise depends in a complicated manner on the displacement fields in the surrounding soil and on the geometry of tunnels, caverns, hills and layer boundaries. For that reason, a numerical approach is warranted, both in predicting the noise and in subtracting the noise. It is not *a priori* clear that the displacement information is strongly correlated with the acceleration, or that a Wiener filter can capture the relevant information. To clarify this and to be better able to interpret the results we obtain in our study, we first proceed with a brief discussion of Newtonian noise in an analytical context.

Consider a cubical volume element with ribs of length Δ in an infinite homogeneous medium. The vector \mathbf{r} , points from the origin of our coordinate space to the center of the volume element. The normals to the surface \hat{n}_i for the six faces of the cube are aligned with the unit vectors $\pm\hat{e}_i$. The equilibrium density, the density of the undisturbed infinite medium, equals ρ_0 . The presence of a time-dependent displacement field $u_i(t)$ introduces mass currents with a size of $J_i = \rho \frac{du_i}{dt}$. From the continuity equation, we can relate the change in mass in an element of volume V to the current through the surface S of that element as

$$\frac{dm}{dt} = \int_V \frac{d\rho}{dt} dV = - \int_S \rho \frac{du_i}{dt} \hat{n}_i dS. \quad (13)$$

The Einstein summation convention to sum over repeated indices is used here and throughout this section. The mass fluctuation with respect to the mass at equilibrium density can be obtained by integrating the mass currents through the surface from the time $t_{u_i=0}$ where u_i equals zero (at the surface) to the present time t ,

$$\delta m = \int_{t_{u_i=0}}^t \frac{dm}{dt} dt = \int_V [\rho(t) - \rho_0] dV = -\rho_0 \int_S [u_i(t) - (u_i = 0)] \hat{n}_i dS. \quad (14)$$

The change in mass due to the integrated currents through the faces of the cube that are parallel to the i -direction thus equate to

$$\delta m = \rho_0 \Delta^2 \hat{e}_i \left[u_i \left(\mathbf{r} - \frac{1}{2} \Delta \hat{e}_i \right) - u_i \left(\mathbf{r} + \frac{1}{2} \Delta \hat{e}_i \right) \right]. \quad (15)$$

The change in density $\delta\rho$ is obtained by dividing the change in mass δm in the volume element by its volume Δ^3 as

$$\delta\rho = \frac{\delta m}{\Delta^3} = \rho_0 \frac{\hat{e}_i [u_i(\mathbf{r} - \frac{1}{2} \Delta \hat{e}_i) - u_i(\mathbf{r} + \frac{1}{2} \Delta \hat{e}_i)]}{\Delta} \approx -\rho_0 \frac{du_i(\mathbf{r})}{dx_i}, \quad (16)$$

where in the last step we take the limit that Δ is small compared to the wavelengths present in the displacement field.

One can decompose the seismic displacement field in terms of plane waves as

$$u_i(\mathbf{r}) = \sum_w A_{w,i} e^{i(k_{w,j} r_j - \omega t + \phi_0)}, \quad (17)$$

where $A_{w,i}$ represents the amplitude of the displacement field in the i -direction, $k_{w,j}$ the wave propagation vector (with length equal to 2π divided by the wavelength), ω the angular frequency, t the time and ϕ_0 the phase of the wave at $t = 0$ in the origin.

Inserting equation (17) into equation (16) yields for the density fluctuation associated with the displacement field at location \mathbf{r}

$$\delta\rho(\mathbf{r}) = \rho_0 \sum_w i k_{w,i} A_{w,i} e^{i(k_w r_j - \omega t + \phi_0)}. \quad (18)$$

One can observe that only longitudinal wave components with $k_{w,i} A_{w,i} \neq 0$ introduce density fluctuations inside the medium. Furthermore, the density fluctuation in a volume element is a scalar property; it depends on the total divergence of the displacement field, not on the direction of the field u_i or on the direction of the wave vector k_j (which are vector properties).

The contribution to the Newtonian noise from the density fluctuations in the cubical volume element with ribs of length Δ at location \mathbf{r} on a test mass located at the origin can be written as

$$\delta a_i = G\delta\rho\Delta^3 \frac{r_i}{|r|^3} = G\rho_0\Delta^3 \frac{r_i}{|r|^3} \sum_w k_{w,j} A_{w,j} e^{i(k_w r_k - \omega t + \phi_0 + \pi/2)}, \quad (19)$$

where δa_i represents the change in acceleration of the mirror in the i -direction and $\rho\Delta^3$ the mass of the volume element considered; we assume that Δ is small compared to the distance r here. The additional phase factor $e^{i\pi/2}$ at the end of equation (19) stems from the prefactor i in equation (18) that arose from taking the derivative of the displacement field.

We notice that first of all, the change in acceleration is in the direction of the volume element, but it is not correlated with the direction of the displacement field. For example P-waves in x , y , and z directions all introduce similar density fluctuations. Secondly, the factor $\pi/2$ in the phase in equation (19) represents a 90-degree phase shift between the displacement field component and the density fluctuation $\delta\rho$ in the center of the volume element. The maximal/minimal density is reached when the displacement amplitude crosses zero; $\delta\rho(t)$ is zero for the maxima/minima in $u(t)$. Thirdly, for transverse waves, $u_i k_i$ equals zero (the displacements in the volume element do not contribute then), whereas for a P-wave, the factor $u_i k_i$ changes sign for waves travelling in opposite directions, leading to a non-linear contribution of the displacement field in the volume element to the acceleration of the mirror, and to vanishing cross-correlations in a Wiener-filter approach in the case of isotropically-distributed waves. In that case a Wiener filter is not suited to predict the contributions of Newtonian noise from a homogeneous medium with isotropic fields on a test mass. It can estimate the magnitude of the contribution, but not the sign; filter coefficients for the cross-correlation will approach zero.

In practice, the test mass is not embedded in a homogeneous medium. It is located in a cavity in the rock, there are tunnels for the beams, and also the medium does not continue above the surface; there rock is replaced with air. In order to calculate the full Newtonian noise, the boundary conditions of the geometry have to be taken into account. From the last part of equation (14) one can derive that mass fluctuations associated with the movement of a surface element $r^2 d\Omega$ introduce changes in acceleration amounting to

$$\delta a_i = G\delta m \frac{r_i}{|r|^3} = G\rho \frac{r_i}{|r|} u_k \hat{n}_k d\Omega, \quad (20)$$

where \hat{n}_k represents a unit vector normal to the surface and r_i points towards the surface element. Note that for *this* term, only displacements in the direction normal to the surface contribute. These may be due to S-waves or P-waves and indeed a large contribution to the Newtonian noise in our model calculations are due to vertical displacement fields at the surface. The direction of δa stems from the direction of the location of the surface element. Whereas the contributions from volume elements scale with $1/r$, the surface contribution in a given solid angle element $d\Omega$ does not depend on the distance to the test mass.

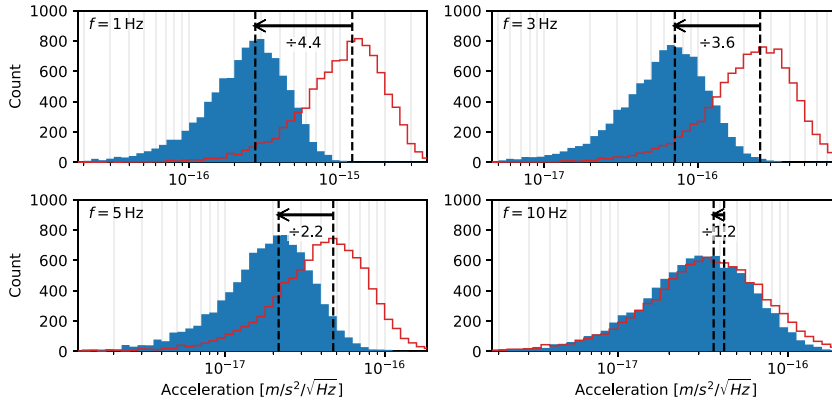


Figure 13. Wiener filter validation record data for Newtonian noise at frequencies of 1, 3, 5 and 10 Hz. Histogram of the original x -direction mirror acceleration noise distribution $a_{\text{true}}^{\text{NN}}$ (red outline) and the reduced x -direction acceleration noise distribution ($a_{\text{true}}^{\text{NN}} - a_{\text{pred}}^{\text{NN}}$) (blue filled). The mean of both distributions is shown as a vertical dashed line. The reduction factor $\text{mean}(a_{\text{true}}^{\text{NN}} - a_{\text{pred}}^{\text{NN}}) / \text{mean}(a_{\text{true}}^{\text{NN}})$ is provided below the arrow in each plot.

Surface contributions arise from the excavations around the mirror and from the soil-air boundary and depend on the geometry of these boundaries. In contrast to volume contributions, a Wiener filter approach may accurately predict surface contributions, due to the fixed phase relation between sensor displacement and mirror acceleration. Depending on the geometry and seismic field properties, the relative importance of surface and volume contributions may differ.

6.2. Wiener-filter simulation results

The synthetic data sets as described in section 5.1 have been applied to a Wiener filter. These sets contain 134 tri-axial sensors and any witness channel that we calculated (we can generate synthetic data for accelerations and displacements at any location). First, we tested whether the seismic sensor network was dense enough. We calculated the displacement field at the position of the test mass (in absence of the cavern around the mirror) and verified how well this field can be reproduced from the synthetic data of the 134 surrounding sensors by using the Wiener filter. As expected, the Wiener filter performs excellently here, since the problem of reproducing the displacement field is linear. The coherence between the most nearby sensor and the displacement field is above 0.99 for the frequencies 1, 3, and 5 Hz and still 0.987 for the frequency at 10 Hz. We use 40 000 different noise trials to build the correlation matrix and determine the Wiener filter coefficients, and use 10 000 new trials to see how these coefficients perform on new data. In all cases, the noise can be reduced with more than 40 dB when the projection is made for the sum of the 40 000 training trials, i.e. if the measured correlations are applied to the full training set. For new trials (with different sources and different noise then in the training set) the results are still excellent; here the reduction factors equal 21 dB at 1 Hz, 37 dB at 3 Hz, 32 dB at 5 Hz and 24 dB at 10 Hz.

Next, we apply the Wiener filter on the acceleration of the test mass in the x -direction. We use 40 000 trials to obtain the filter coefficients and 10 000 new trials to establish the performance. The results for Newtonian noise are shown in figure 13.

We observe that the Wiener filter is capable of reproducing the noise at 1 Hz quite well, but that the performance deteriorates for higher frequencies and it performs marginally at 10 Hz. This can be understood from the size and nature of the relative contributions to the Newtonian noise. Since for the Wiener filter we determined all normalized cross-correlations, we can inspect which sensors contribute most. At 1 Hz, the wavelength is long, typically kilometer scale for Rayleigh waves, and the soil moves largely coherently in the vicinity of the mirror. This implies that the cavern wall displacements and to a lesser extent the surface displacements (which are coherent over a long distance at 1 Hz) contribute relatively more than the local density fluctuations in the surrounding soil. The maximal correlation (0.605 in amplitude) occurs with the sensor in the x -tunnel 250 m downstream from the mirror, where the x, y, z coordinates are defined in figure 12. This is not the most nearby sensor, it seems that at this position the displacement contribution from the cavern wall and the density contribution from the P-wave component of the wave in x -direction add coherently. Indeed, the coherence between the displacement field in the y -direction and the mirror acceleration in x -direction is also substantial for this sensor (0.346). We interpret this as due to events with P-wave components traveling under an angle with respect to the x -axis; for some waves the displacement of the cavern can be close in phase with the density fluctuation in the soil at the sensor position 250 m in x -direction (which is 90 degrees out of phase with the displacement field at the same location). It seems to help if the contraction is partly due to an y -component in the wave. The correlations for both x and y displacements with the acceleration in x -direction are close to π radians out of phase; as should be expected from equations (19) and (20).

At 10 Hz, the wavelengths for surface waves are rather short (in the order of 25 m). The sizes of the regions that contribute coherently are small, leading to small correlations between the contributions from any given sensor and the acceleration of the mirror. The maximal correlation between displacement field and acceleration at 10 Hz is only 0.14 in amplitude. It occurs between the test mass acceleration and the vertical displacement at the surface, measured by a sensor located at -400 m in x -direction. The next-highest correlation is also with the vertical channel of a sensor at the surface, this time at -200 m in x -direction. Both correlations show a phase close to zero, as expected from equation (20) for the terms due to surface displacement. The largest correlations for an underground sensor occur for a sensor 13 m away, at $x = 6.7$ m and $y = -11.5$ m, with coherences of 0.099 and 0.089 for the displacements in x - and y -direction, respectively. That the largest correlations occur due to surface sensors can be understood when one realizes that at 10 Hz, the displacement power spectral densities are four orders of magnitude larger than at 250 m depth. Even though the surface is relatively far away, it still dominates in the integral.

In conclusion, Wiener filters may be applied to subtract Newtonian noise, and are most effective at long wavelengths and in cases where the density fluctuations induced by P-waves in the volume of the surrounding rock are not a dominant contribution. For the Terziet geology, at frequencies of 10 Hz, we find that a Wiener filter approach is performing poorly.

7. DNN for Newtonian noise prediction

For the Newtonian noise prediction, a different implementation for Seismonet is required which only operates on records generated for a single frequency $f \in \{1, 3, 5, 10\}$ Hz. Figure 14 shows that this DNN has two pipelines, one for linear correlations (bottom) and one for non-linear correlations (top). The synthetic input data consists of the real and imaginary part of the spectra of 136 sensors that measure in the x, y, z direction as described in section 5.1.

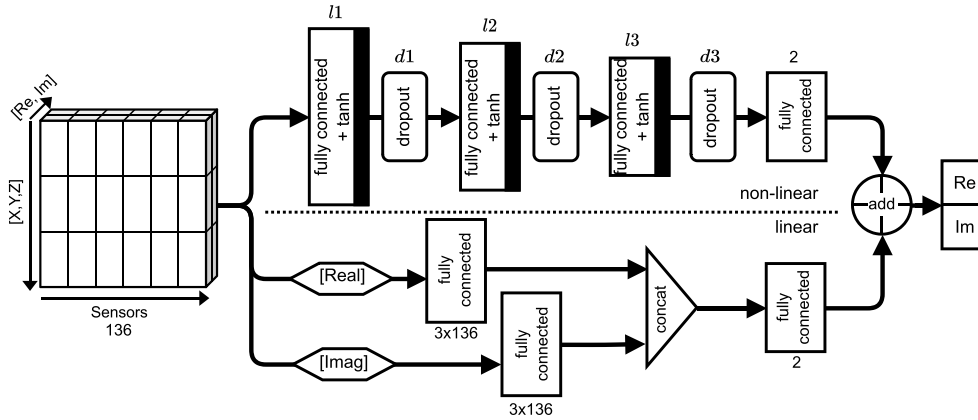


Figure 14. The DNN model employed for the Newtonian noise prediction from synthetic data consists of a linear and a non-linear pipeline. For the linear layers the number of neurons is shown, while for the non-linear layers the numbers of neurons and the dropout values differ depending on frequency. Input consists of a 3D array of sensor-id, direction (x, y, z), and complex amplitude (Re, Im); output is the predicted mirror acceleration (Re, Im) in the x -direction.

Table 1. Parameters used for the non-linear pipeline of the DNN model shown in figure 14. The values have been determined using hyperparameter optimization. For each frequency $f \in \{1, 3, 5, 10\}$ Hz, the number of neurons is provided for the fully connected layers $l1 \dots l3$, and the dropout rate for dropout layers $d1 \dots d3$.

f (Hz)	Neurons in layer			Dropout rate		
	$l1$	$l2$	$l3$	$d1$	$d2$	$d3$
1	1536	1228	982	30%	27%	24%
3	3072	2457	1965	10%	4%	2%
5	3072	1536	768	40%	20%	10%
10	3072	1843	1105	30%	27%	24%

The linear pipeline optimizes through linear combination of channels, effectively similar to a Wiener filter after optimization. This pipeline has no hyperparameters. The non-linear pipeline consist of 4 fully connected layers with three activation functions, and three dropout layers. This pipeline has been optimized for each frequency separately with HPO. The number of neurons for the fully connected layers $l1 \dots l3$ and the dropout value for dropout layers $d1 \dots d3$ are provided in table 1. During HPO, the activation function was also optimized, and the best models all utilized the hyperbolic tangent (tanh) activation function.

Training was done with the same hardware and optimizer configuration as discussed in section 3.3. Convergence is reached after validation loss does not decrease for 90 epochs.

7.1. Results for Newtonian noise prediction

The performance of the neural network is shown in figure 15. The figure shows validation data for frequencies of 1, 3, 5 and 10 Hz. For each frequency the red histogram shows the true mirror

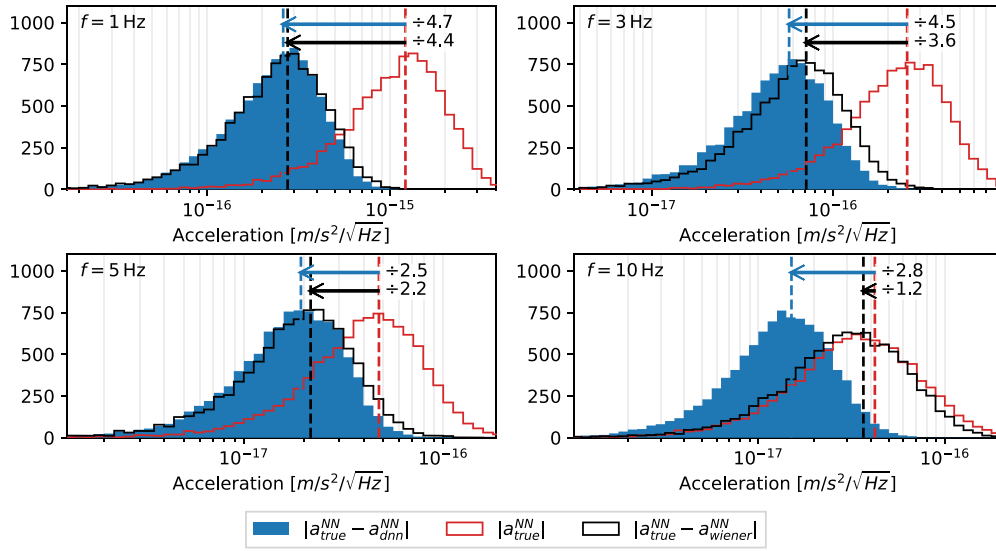


Figure 15. Validation record data for frequencies at 1, 3, 5 and 10 Hz. The original x -direction mirror acceleration noise distribution $a_{\text{true}}^{\text{NN}}$ is shown as outline and the reduced x -direction acceleration noise distribution ($a_{\text{true}}^{\text{NN}} - a_{\text{dnn}}^{\text{NN}}$) as step-filled histogram. For reference the reduction when applying a Wiener filter to the same data $a_{\text{true}}^{\text{NN}} - a_{\text{wiener}}^{\text{NN}}$ has also been provided. The mean of all distributions is shown as a vertical dashed line. The reduction factor $\text{mean}(a_{\text{true}}^{\text{NN}} - a_{\text{dnn}}^{\text{NN}})/\text{mean}(a_{\text{true}}^{\text{NN}})$ and $\text{mean}(a_{\text{true}}^{\text{NN}} - a_{\text{wiener}}^{\text{NN}})/\text{mean}(a_{\text{true}}^{\text{NN}})$ is provided after each arrow in each plot.

acceleration noise $a_{\text{true}}^{\text{NN}}$ for the x -direction. The blue histogram shows the Newtonian noise subtracted x -direction acceleration noise distribution $a_{\text{true}}^{\text{NN}} - a_{\text{pred}}^{\text{NN}}$. The mean of both distributions is shown as a vertical dashed line.

The figure shows that subtracting $a_{\text{pred}}^{\text{NN}}$ from the true acceleration $a_{\text{true}}^{\text{NN}}$ reduces the amplitude of the Newtonian noise by a factor of 2.8 at 10 Hz. The predictions are better at low frequencies, where seismic wavelengths are larger and coherence lengths are longer.

The DNN was successfully applied to minimize a_x and a_y separately, and to minimize both horizontal accelerations simultaneously. We expect that the neural network can be equally well applied to reconstruct the acceleration of a single mirror when trained on this parameter as to subtract Newtonian noise in the case of three co-located detectors in a equilateral triangular layout when trained on combinations of the strains of these detectors; see section 5.1.

Figure 16 shows the predicted contribution of Newtonian-noise to the sensitivity of Einstein Telescope at the EMR candidate site with and without the use of seismic sensor arrays in combination with a DNN. The Newtonian noise contains contributions from seismic surface waves and a body-wave background. The prediction assumes the strain sensitivity for a single interferometer with 10 km arm length, and test masses at 250m depth enclosed in 10m radius caverns. Solid curves show the mean, while the colored bands show the 10–90 percentile. The ET-D sensitivity is represented by the dashed-black curve. Our results show that a DNN can effectively predict mirror accelerations from seismic displacement data obtained with a network of tri-axial sensors. This allows predicting Newtonian noise and our study indicates that a significant contribution can be subtracted from the strain output, especially at low frequency.

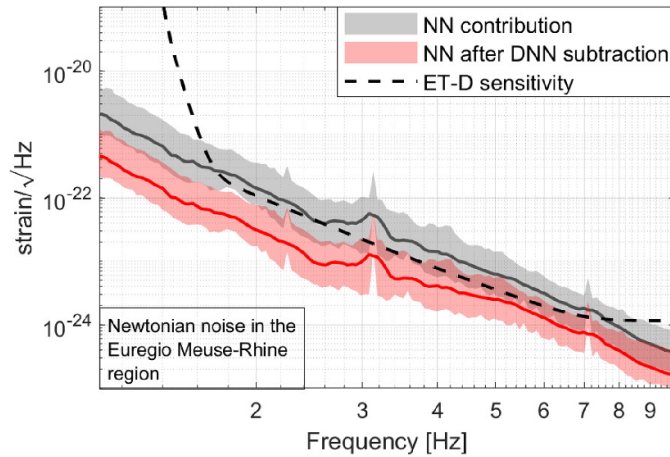


Figure 16. Prediction for the Newtonian-noise contribution to the sensitivity of Einstein Telescope at the Euregio Meuse-Rhine candidate with and without the use of seismic sensor arrays in combination with a deep neural network. The Newtonian noise contains contributions from seismic surface waves and a body-wave background. The prediction assumes three interferometers in an equilateral triangle with 10 km arm length, and test masses at 250 m depth enclosed in 10 m radius caverns. Solid curves show the mean, while the colored bands show the 10–90 percentile. The ET-D design sensitivity is represented by the dashed-black curve.

8. Summary and outlook

A DNN that employs a network of surface and underground sensors as input is a promising tool for subtracting Newtonian noise from gravitational-wave strain signals. Such a network is specifically effective in the band where Newtonian noise strongly contributes to the detector sensitivity.

The applied model includes the full solution of the elasto-dynamic wave equation for seismic surface excitation, while body waves are modeled as harmonic plane waves. Studies of underground geology and source distribution remain important to obtain a more realistic layered 3D geology and enhanced characterization of both local geology and seismic wave field, mostly notably of body waves. Additionally, the accuracy of a DNN for Newtonian noise prediction may benefit from the use of three-component sensors. Also the displacement noise generated by the facility infrastructure will be an important factor in the actual Newtonian noise experienced at the observatory. These effects need to be numerically studied as well.

For Einstein Telescope with three low-frequency interferometers the synthesis problem is complex. For each interferometer the strain signal depends on the accelerations of the 4 Fabry–Perot cavity mirrors at the ends of 2 arms. For uncorrelated noise, we show that our neural network can perform equally well in subtracting Newtonian noise in the case of three co-located interferometers, where three strain signals are present. Each corner hosts the input mirror of one interferometer and the end mirrors of the other two interferometers. This causes the seismic noise to be partly correlated, especially at low frequency. To have a realistic test would require better knowledge of the body-wave background, implementation of the different geologies at the corner stations, the exact lay-out of tunnels, excavations, and underground infrastructure, and the local environmental noise. Such detailed modeling will be the subject of future work.

Data availability statement

The data cannot be made publicly available upon publication because no suitable repository exists for hosting data in this field of study. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

We acknowledge Netherlands Science Organisation NWO for their role in facilitating this research.

Appendix. Seismonet

As discussed in section 2.1 Seismonet operates in the frequency domain and takes the complex spectra of reference channels for a specific segment of time as input and estimates the complex spectrum for a witness channel.

Seismonet has a sequential DNN architecture, illustrated in figure A1, where we follow the nomenclature of Keras [19]. Each layer in this architecture takes the input data from the previous layer, performs a specific operation, and passes the result to the next layer. The number of times a section is repeated (G and H in figure A1) is configurable by hyperparameters. Additionally each repetition does not need to have the same number of neurons. Seismonet provides customization through hyperparameters, allowing for example adjustment of the number and width of bands in the band correlation section, the number of neurons in the connected layers, and the deactivation ratio in the dropout layers.

The most important layers are the fully connected and locally connected layers which are labeled respectively Dense and LocallyConnected in figure A1. These are the only layers which can be optimized and involve one or more matrix multiplications of type

$$\mathbf{o} = \mathbf{W}\mathbf{i} + \mathbf{b}, \quad (\text{A.1})$$

where \mathbf{i} represents the vector of input values and \mathbf{o} denotes the output vector. Vector \mathbf{b} is optional and contains bias weights, while \mathbf{W} represents the matrix of trainable weights and contains $\dim(\mathbf{o}) \times \dim(\mathbf{i})$ values.

The neural network is trained by using back-propagation learning. This is an iterative process involving neural network prediction, computing the error between the predicted and expected data by using a loss function, and adjusting the weights of the network layers such that the error is reduced. Generally training ends when the error no longer decreases substantially or when a specific number of training iterations has been completed. For a comprehensive review of DNNs see [25].

A fully connected layer performs a single matrix multiplication over *all* the input data, while a locally connected layer performs matrix multiplications on *subsections* of the data. This can be visualized as sliding a window over the input data, where each window position has its own corresponding \mathbf{W} , and only the data within the window is used as input \mathbf{i} .

Figure A2 provides a simplified example of our use of this sliding window property to generate distinct frequency bands from the input data. The example assumes an input for 6 frequency bins and from 6 reference channels. This input is separated into 3 bands by using a window of 3×6 (frequency bins \times channels) with a stride of 2 in frequency. Since each spectrum has a real and an imaginary component, the input for each band contains $2 \times 3 \times 6 =$

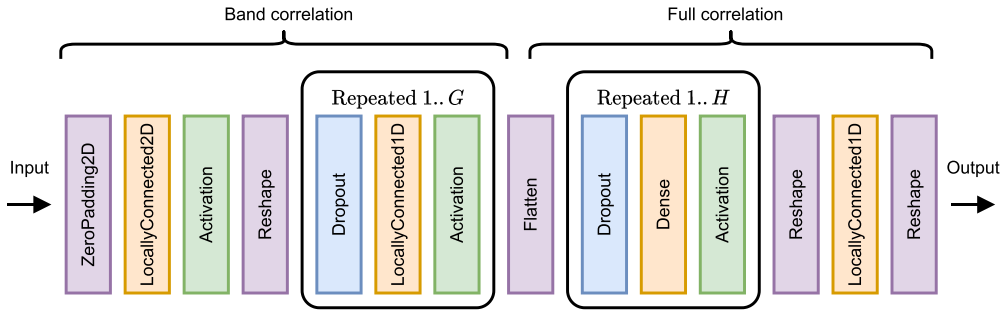


Figure A1. Seismonet consists of a sequence of layers, each serving a specific purpose. The purple layers modify the shape of the data while preserving its content. Orange layers contain neurons that perform a linear transformation on (a subset of) the data. The green activation layers apply a non-linear function to introduce non-linearity into the network. Blue dropout layers are incorporated to prevent over-fitting and improve the generalization. Both the band correlation and full correlation layers contain sections that can be repeated a number of times.

36 components. These values are transformed into a flattened vector \mathbf{i}_n with n as band index. By employing the band-specific weight matrix \mathbf{W}_n and bias weight vector \mathbf{b}_n , this input vector is mapped to an output vector \mathbf{o}_n , whose elements are denoted filters. In this example we assume that HPO yields that 4 filters are needed to optimize this layer. Consequently, the four elements of the output vector \mathbf{o}_n for a specific band are now linear combinations of all the input values, and can generally no longer be attributed to a specific reference channel or frequency bin. Subsequently, the output vectors of the different band are concatenated to form a new output comprising $\text{bands} \times \text{filters} = 3 \times 4 = 12$ components.

The advantage of using bands in the first section of the network is twofold. First it reduces the number of weights required, compared to a fully connected layer with the same number of inputs and outputs, thereby reducing model complexity. Second it guides the optimization process to take into account bands with a lower signal to noise ratio compared to that of the full spectrum, allowing for more effective signal extraction. The addition of overlap between bands, achieved by setting the stride to be less than the width of the band, in bins, enables the network to effectively capture frequency information at band transitions. Omitting this overlap leads to reduced accuracy in determining frequencies near the edges of the bands.

The way Seismonet applies the LocallyConnected2D decreases the number of dimensions of the data from three (channels \times frequencies $\times \mathbb{R}(R_x), \mathbb{I}(R_x)$) to two (bands \times filters). For this reason subsequent layers only require a LocallyConnected1D layer to process data in separated bands. This layer is equivalent to a LocallyConnected2D, but operates on data with two dimensions: bands and filters.

The output of one layer is fed into the next with an activation function A

$$i_k^{(n)} = A(o_k^{(n-1)}), \quad (\text{A.2})$$

where $i_k^{(n)}$ and $o_k^{(n-1)}$ represent the value of element k of respectively the input of layer n and of the output vector of the previous layer. The activation function enables the network to perform non-linear functions and operates element-wise, thus preserving the shape of the input. With

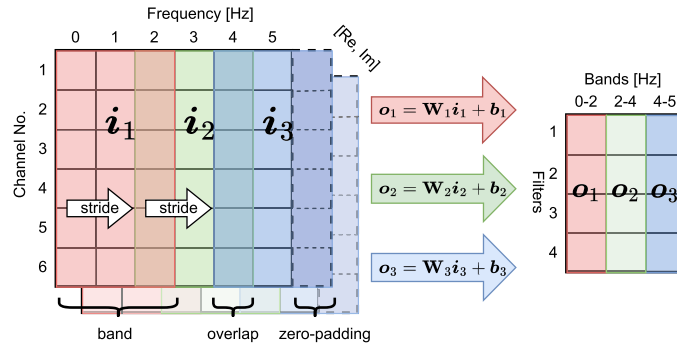


Figure A2. Example of how the LocallyConnected2D layer is applied to separate an array of 6 frequency bins and six channel data into three separate frequency bands. A stride of 2 in frequency results in an overlap of 1 frequency bin per band. The output vector for band n is found by $\mathbf{o}_n = \mathbf{W}_n \mathbf{i}_n + \mathbf{b}_n$ where \mathbf{i}_n denotes a vector of the flattened input values in band n , and \mathbf{W}_n and \mathbf{b}_n the weights specific for that band. In this example zero padding is applied first as the number of bands would otherwise not fit the data. Hyperparameter optimization is assumed to indicate that 4 filters are optimal for this layer, leading to $4 \times 36 + 4 = 148$ trainable parameters for each band.

HPO we determined that for Seismonet the ELU [12] activation function produces the most accurate networks. ELU is given by

$$\text{ELU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha(e^x - 1), & \text{otherwise,} \end{cases} \quad (\text{A.3})$$

where for Seismonet $\alpha = 1$.

The dropout layer is only active during training and effectively randomly disables a number of neurons in the previous layer. This is achieved by zeroing their output. The ratio of neurons which are disabled can be configured, but the specific set of neurons changes per iteration. Dropout forces the network to generalize over multiple inter-layer connections and prevent over-fitting.

The complexity of Seismonet can be appreciated by considering some details for the DNN implementation used in section 3 to estimate ambient surface noise. Here, Seismonet was configured with the sequential model, with layer configuration, data dimensionality and number of trainable weights provided in table A1.

As an example, the derivation of the number of weights will be explained for the LocallyConnected2D layer. This layer is employed to partition the input data into 18 bands, with each band comprising 16 frequency bins and 66 input channels. Thus, the total number of inputs per band for this layer can be calculated as

$$\text{Number of inputs per band} = 16 \times 66 \times 2 = 2112.$$

Here, the factor 2 arises from the real and imaginary part of the spectrum of each reference channel.

The next step involves mapping these inputs to 1000 filters by using a weight matrix of dimension 2112×1000 . Additionally, each filter is accompanied by an extra bias weight. Hence, the total number of weights per band can be calculated as

$$\text{Number of weights per band} = 2112 \times 1000 + 1000 = 2113000.$$

Table A1. Configuration of layers, data dimensionality and trainable weights for estimating ambient surface noise (see section 3) as provided by the Keras model summary function of Tensorflow. Dropout layers are omitted for brevity. See figure A1 for a visual representation of the layers and the network.

Layer	In shape	Out shape	Weights	Remark
ZeroPadding	$66 \times 257 \times 2$	$66 \times 271 \times 2$		Required for next layer
LocallyConnected2D	$66 \times 271 \times 2$	$1 \times 18 \times 1000$	38 034 000	16 input bins per band
Reshape	$1 \times 18 \times 1000$	18×1000		Drop first dimension
LocallyConnected1D ($G = 1$)	18×1000	18×600	10 810 800	
LocallyConnected1D ($G = 2$)	18×600	18×360	3894 480	
Flatten	18×360	6480		
Dense($H = 1$)	6480	1915	12 411 115	
Dense($H = 2$)	1915	514	826 512	
LocallyConnected1D	514	514	1028	Scaling and translation
Reshape	514	257×2		
Total			65 977 935	

Considering that there are 18 bands, the overall number of weights in the LocallyConnected2D layer can be calculated as

$$\text{Number of weights in LocallyConnected2D} = 2113000 \times 18 = 38034000.$$

This value aligns with the weight count provided by Tensorflow and shown in table A1.

ORCID iDs

Vincent van Beveren  <https://orcid.org/0000-0001-8397-9725>

Jo van den Brand  <https://orcid.org/0000-0003-4434-5353>

Xander Campman  <https://orcid.org/0000-0001-6482-1391>

Soumen Koley  <https://orcid.org/0000-0002-5793-6665>

References

- [1] Punturo M *et al* 2010 The Einstein Telescope: a third-generation gravitational wave observatory *Class. Quantum Grav.* **27** 194002
- [2] van den Brand J 2019 *100 Years of General Relativity: Advanced Interferometric Gravitational-Wave Detectors* vol 5 (World Scientific) ch 6
- [3] Bader M, Koley S, van den Brand J, Campman X, Bulten H, Linde F and Vink B 2022 Newtonian-noise characterization at Terziet in Limburg—the Euregio Meuse-Rhine candidate site for Einstein Telescope *Class. Quantum Grav.* **39** 025009
- [4] Zhang H and Wang D 2021 Deep ANC: a deep learning approach to active noise control *Neural Netw.* **141** 1–10

- [5] Vajente G, Huang Y, Isi M, Driggers J C, Kissel J S, Szczepańczyk M J and Vitale S 2020 Machine-learning nonstationary noise out of gravitational-wave detectors *Phys. Rev. D* **101** 042003
- [6] Ormiston R, Nguyen T, Coughlin M, Adhikari R X and Katsavounidis E 2020 Noise reduction in gravitational-wave data via deep learning *Phys. Rev. Res.* **2** 033066
- [7] Badaracco F *et al* 2020 Machine learning for gravitational-wave detection: surrogate Wiener filtering for the prediction and optimized cancellation of Newtonian noise at Virgo *Class. Quantum Grav.* **37** 195016
- [8] Liu Q, Fu L and Zhang M 2021 Deep-seismic-prior-based reconstruction of seismic data using convolutional neural networks *Geophysics* **86** V131–42
- [9] Chollet F *et al* 2015 Keras (available at: <https://keras.io>)
- [10] Abadi M *et al* 2015 TensorFlow: large-scale machine learning on heterogeneous systems (available at: www.tensorflow.org/)
- [11] Hinton G E, Srivastava N, Krizhevsky A, Sutskever I and Salakhutdinov R R 2012 Improving neural networks by preventing co-adaptation of feature detectors (arXiv:1207.0580 [cs])
- [12] Clevert D-A, Unterthiner T and Hochreiter S 2016 Fast and accurate deep network learning by exponential linear units (ELUs) (arXiv:1511.07289 [cs])
- [13] Yu T and Zhu H 2020 Hyper-parameter optimization: a review of algorithms and applications (arXiv:2003.05689 [cs, stat])
- [14] Koley S, Bader M, Brand J van den, Campman X, Bulten H J, Linde F and Vink B 2022 Surface and underground seismic characterization at Terziet in Limburg—the Euregio Meuse–Rhine candidate site for Einstein Telescope *Class. Quantum Grav.* **39** 025008
- [15] Peterson J R 1993 *Observations and Modeling of Seismic Background Noise (USGS Numbered Series 93-322)* (U.S. Geological Survey)
- [16] Geophone-based seismic sensor developed by Innoseis (available at: www.inovageo.com/products/quantum)
- [17] Dozat T 2016 *Incorporating Nesterov Momentum into Adam* (available at: <https://openreview.net/forum?id=OM0jvwB8jlp57ZJjtNEZ>) (Accessed 24 May 2022)
- [18] Li L, Jamieson K, DeSalvo G, Rostamizadeh A and Talwalkar A 2018 Hyperband: a novel bandit-based approach to hyperparameter optimization (arXiv:1603.06560 [cs, stat])
- [19] Keras Team 2019 *Keras Documentation: KerasTuner* (available at: https://keras.io/keras_tuner/)
- [20] Wiener N 2013 *Extrapolation, Interpolation and Smoothing of Stationary Time Series, With Engineering Applications* (Wiley)
- [21] Bader M K M 2021 Seismic and Newtonian noise modeling for Advanced Virgo and Einstein Telescope *PhD Thesis* VU University Amsterdam
- [22] KNMI—Koninklijk Nederlands Meteorologisch Instituut 1993 Data are publicly (available at: <http://rdsa.knmi.nl/dataportal>) for the sensors at Heimansgroeve (HGN) and at Terziet (TERZ)
- [23] Schevenels M, François S and Degrande G 2009 EDT: an ElastoDynamics toolbox for MATLAB *Comput. Geosci.* **35** 1752–4
- [24] Beker M *et al* 2016 Innovations in seismic sensors driven by the search for gravitational waves *Leading Edge* **35** 590–3
- [25] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (MIT Press)