Filling the gap of seismic ambient noise taken from the earth by modification of the frequency content of the existing time series

Mohammad Reza Nameni[1] · SeyedMohammadSadegh Jafari[2] · Habib Rahimi[1]

Abstract
Seismic ambient noise is increasingly important in estimating velocity and attenuation structures, as well as understanding parameter variations. However, gaps in recorded time series create challenges in processing ambient noise. This study introduces a novel technique that fills these gaps by using time series with suitable frequency content. To achieve this, a gap was intentionally applied into a continuous time series, and a similar time series, of the same length as the gap, is selected as the primary signal. A moving periodogram, incorporating various windows, was applied to the time series before and after the gap. Stacking these periodograms creates a reference periodogram that indicates the frequency content of the gap. This reference periodogram was then used to improve the frequency content of the primary signal. The study shows promising results with a significant improvement in correlation between the primary signal and the gap through the modification of frequency content. Increasing the number of periodograms used in the stacking process improves the estimation of frequency content, and optimal window lengths have a significant impact on outcomes. This methodology ensures accurate recovery of high powers through modification operations. By addressing data gaps in seismic ambient noise, this advanced technique contributes to continuous progress in this field.

Keywords Seismic ambient noise · Gap · Periodogram · Window · Stacking

Introduction

Data collection is considered the most important part of any scientific work. If the collected data is a function of time, we can consider it as a time series. Time series data is a collection of data points gathered during a particular timeframe at regular intervals. One of the main characteristics of time series data is temporal dependence, meaning that each observation is influenced by the preceding ones (Shumway and Stoffer 2017). In the field of geophysics, a large portion of the gathered data can be categorized as time series, and it is anticipated that they will exhibit similarities to time series in terms of their characteristics. For instance, Ambrosino et al. (2019) conducted an analysis on the time series data of the earth's rotation rate and discovered correlations between detected anomalies in the rotation rate and the incidence of earthquakes. Similarly, Briestenský et al. (2021) found that radon activity concentrations varied across different caves but exhibited comparable seasonal trends, with higher values during summer and lower values during winter. Ansari et al. (2023a, b) examined seismic response at 200 sites and estimated amplification factors at different time intervals. They also used field data from geophysical testing to generate a vulnerability distribution for the area.

Ansari et al. conducted a study in 2022 to investigate the importance of comprehending earthquakes and historical information in evaluating how they impact structures. Their research specifically concentrates on utilizing data from the Himalayan region to create a mathematical model called the seismic tunnel damage prediction (STDP) model. In 2023, Ansari et al. further utilized fragility functions to measure the seismic performance of tunnels in various hazard zones. Fragility functions are used to assess the tunnels' vulnerability based on their location in different hazard zones and their depth (Ansari et al. 2023a, b).

Responsible Editor: Longjun Dong

✉ Habib Rahimi
rahimih@ut.ac.ir

[1] Institute of Geophysics, University of Tehran, Tehran, Iran

[2] SEISRAM, Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy

Another important time series in geophysics is ambient noise. The earth's surface is never at rest and is always subject to different forces and deformational energy sources such as human activities, atmospheric pressure, and tides (Nakata et al. 2019). Nowadays, seismic ambient noise has attracted lots of attention for many seismological studies especially to study the Earth's structure, e.g., in noise interferometry, Green's function, phase-velocity measurements, and tomography (Sadeghisorkhani 2017). With the advances in seismometers and the establishment of broadband seismic networks, the ambient fields of these forces are continually recorded and they are also called microseisms because of not having a dominant signal (Lay and Wallace 1995; Shearer 2009).

From the frequency point of view, ambient noise is generally categorized into three types: short periods (0.1–1 s), intermediate periods (1–30 s), and long periods (30–500 s). The first group is usually due to human activities, and machinery, while the second group is a product of microseisms, and the third group is a result of storms and oceanic tides (Nakata et al. 2019).

Missing data (referred to as a gap) is considered a crucial problem in any field of science, especially geophysics and seismology, which deals with data. Since the data are required for many processes, this challenge is always a source of concern and uncertainty for scientists. The gaps could be due to either failure in measurement instruments or errors in infrastructures related to storage, communication, etc. (Sidekerskiene and Damasevicius 2016; Xue et al. 2017). The gaps and, hence, uneven sampling can strongly affect the processing, analyses, and interpretation of the data, particularly the frequency content. Although there are some methods for uneven sampling, most of the standard data processing methods need even sampling; therefore, the missing data must be estimated or interpolated which is called gap filling (MacDonald 1989; Foster 1996).

To tackle the gap-filling problem, many scientists struggled and proposed several methods, some of which are used multiple times by many scientists while some are unique. For example, one of the most common and easiest solutions is an interpolation, in particular the linear type of it (Pascual-Granado et al. 2015; Reynolds and Smith 1994). However, this method usually produces some spectral artifacts and also not differentiable time series. The Lagrange form of interpolation is more prevalent among commonly used methods although it is not suitable for numerous points in time series since it will increase the order of the polynomial, causing fluctuations (Musial et al. 2011). To avoid this problem, the spline method was proposed which, while being computationally cost-effective and generally satisfying, has the problem of utilizing the spectral content of the whole of a time series (Hocke and Kämpfer 2008).

Furthermore, singular spectrum analysis (SSA) method is known as the more advanced and more popular method which is proposed by Kondrashov and Ghil (2006) based on the technique firstly developed by Broomhead and King (1986) and Broomhead et al. (1987). Even though it is suitable for anharmonic signals and nonlinear trends, this method has some drawbacks. For example, it needs heavy computation and is not appropriate for long time series (Vautard et al. 1992; Ghil et al. 2002).

Another approach is the CLEAN algorithm. Its main objective is to make the spectrum and frequency content of the signal undominated. The etymology of it is cleaning the signal from artifacts. It can implement the iterative deconvolution of the signal, and the reconstruction takes place by using inverse DFT (Baisch and Bokelmann 1999). Nonetheless, Hloupis et al. (2004) mentioned that this algorithm is only applicable for gap lengths up to 10% of the signal length.

Empirical mode decomposition (EMD) is another popular approach. In this algorithm, the signal is decomposed into its components which are called intrinsic mode functions (IMFs) obtaining the reconstruction. On the other side, this method is time-consuming and is not proper for large gaps (Moghtaderi et al. 2012; Huang et al. 1999).

In another research, Barnhart et al. (2011) developed the EMD method into a more advanced algorithm which is called ensemble empirical mode decomposition (EEMD) and claimed that it is locally more accurate.

As the final method in this list, the Lomb-Scargle method uses a periodogram of the signal by calculating its power spectrum by applying Hamming window. Then, with the help of inverse Fourier transform, it can reconstruct the signal (Hocke and Kämpfer 2008; Scargle 1989). However, Harris (1978) stated that Hamming window provides poor results and the Kaiser-Bessel window is a better alternative. It is also worth mentioning that this method has some problems with aperiodic fluctuations.

In all the introduced methods, the purpose was to create time series to fill the gap. Considering that the ambient noise recorded by the broadband sensors covers a wide range of frequencies, the aim of this study was not to create a new time series. Accordingly, the purpose was to use the existing signals with constructive modifications to fill the gap.

To achieve this goal, a close estimate of the gap frequency content was obtained by creating a gap on a continuous time series and analyzing the frequency spectrum around the gap using a periodogram. The time series taken after the gap was then replaced by the gap, and its frequency content was modified using estimates.

Data and data analysis

Seismic ambient noise used in this study was from 1-h records of 22 broadband stations belonging to the Iranian Seismological Center (IRSC), Institute of Geophysics, University of Tehran (IGUT), and 4 broadband stations belonged to the International Institute of Earthquake Engineering and Seismology (IIEES). The sampling rate of these records was 50 Hz. To eliminate very long periods, a 0.1 Hz high-pass filter was used. Considering the high-pass filter and sampling rate, the maximum period (wavelength) was 10 s. Therefore, the time series must be at least 500 samples to contain all frequencies, hereinafter referred to as "*minimum length.*"

$$L_{\mathrm{Min}} = T_{\mathrm{Max}} \times R \tag{1}$$

in which L_{Min} is the minimum length of the window, T_{Max} is the maximum period, and R is the sampling rate (Fig. 1).

Since the purpose of this study was to test the efficiency of the proposed method in different lengths of the gap and due to the limitation of the length of the used time series which was 1 h, it was decided to perform the test on 1, 5, 10, 15, 20, and 25-min gap, which this equates to 1.6 to 41.6% of the total time series available. Using these gaps makes the method practical for very short up to long gaps.

Methodology

In this study, to fill the gaps in the seismic noise time series, the proposed flowchart shown in Fig. 2 was used. For this purpose, after applying the gap in each 1-h time series, a part of the signal with a length equal to the gap was selected as the "initial signal." To select the initial signal before or after the existing gap, an evaluation was performed, and based on the results tabulated in Table 1, a slightly higher correlation coefficient in terms of frequency content was observed for the time series taken from after the gap. Therefore, we selected the initial signal after the gap. As shown in Table 1, this correlation decreases as the gap length increases. The frequency content of the time series is one of the most important factors for seismological studies. Thereby, to assess the efficiency of the applied method, it was decided to use the Pearson's correlation coefficient between the absolute value of the frequency content of the original and filled time series as a measure of goodness.

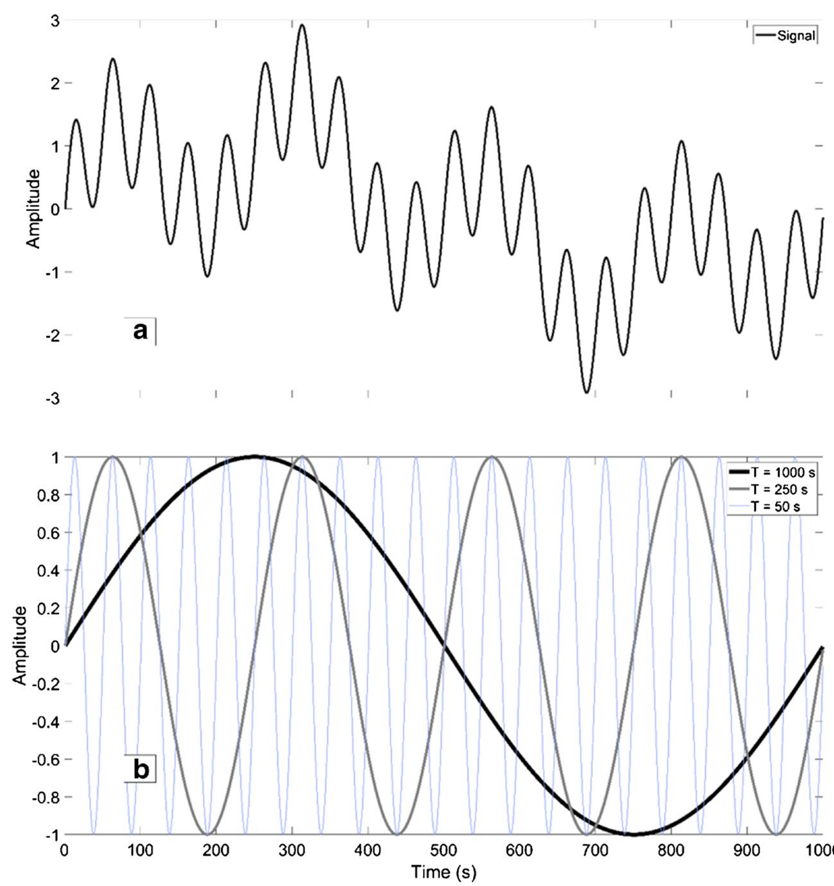


Fig. 1 Frequency decomposition, **a** A signal consists of three sinusoidal signals with periods: $T = 1000$ s, $T = 250$ s, and $T = 50$ s. **b** Decomposition of the top panel signal into its frequency components. Window length must be at least the same as the longest period

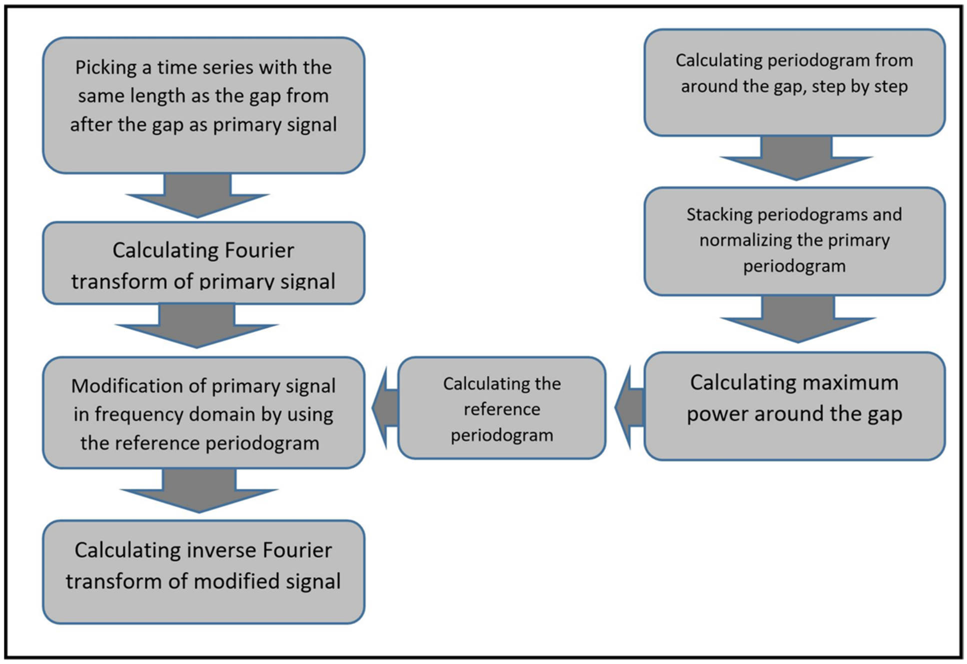


Fig. 2 The proposed flowchart in order to fill the gaps in the seismic noise time-series

Table 1 To select the initial signal before or after the existing gap, an evaluation was performed and the results tabulated

Gap length (min)	Correlation with before the gap	Correlation with after the gap
1	0.771335	0.772714
5	0.755883	0.759315
10	0.752165	0.756876
15	0.736163	0.751803
20	0.722147	0.748546
25	0.716045	0.748082

The formula for Pearson's correlation coefficient (ρ) is:

$$\rho_{X,Y} = \frac{\mathrm{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{2}$$

where:

E is the expectation, μ is the mean, and σ is the standard deviation.

Additionally, the parts of the time series with certain lengths were chosen before and after the gap which was used as a basis (from now on: "base signal") for the calculation of the power spectrum (Fig. 3).

To obtain the power spectrum, the periodogram was used. The periodogram is a spectral estimator with the following equation (Stoica and Moses 1997):

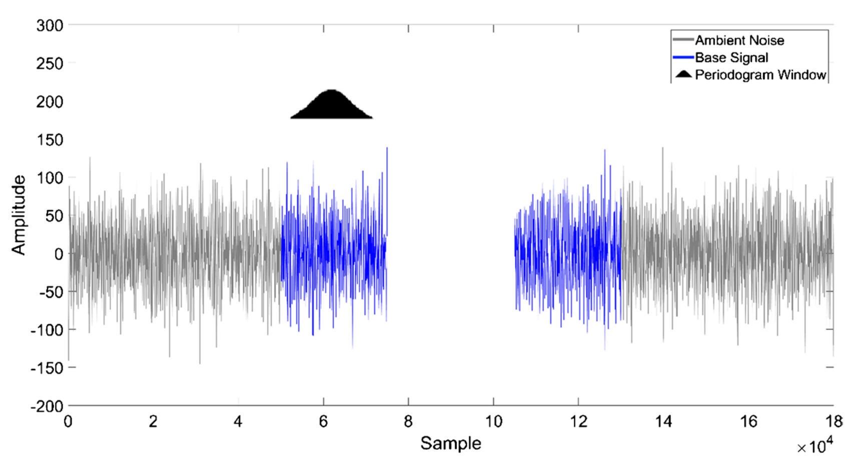


Fig. 3 Ambient noise (gray), base signals (blue), and moving window (window size was exaggerated for better realization)

$$\text{Periodogram} = \frac{|\text{Fourier transform (signal)}|^2}{N} = \frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{-i\omega t}\right|^2 \tag{3}$$

where N is the number of samples in the time domain and ω is the angular frequency.

The periodogram window was then moved on the base signal with certain steps, and a periodogram was calculated for each step. The best window length was used for each window (Fig. 3).

To apply a window on a signal in the time domain, the window is multiplied by the signal, and the frequency response of the window is convolved with the frequency response of the signal to do the same task in the frequency domain. Since the frequency response of windows varies, the type of window affects processing. The closer the window is to the rectangle, the narrower the main lobe of the frequency response and the larger the lateral lobe. In contrast, the larger the smooth window (more like Gaussian), the wider the main lobe and the smaller the sidelobe (Fig. 4).

Increasing the peak sidelobe causes noise in the signal, and the wider the main lobe, the smoother the signal will be. Therefore, the periodogram was computed by using 17 different windows, including:

Bartlett-Hann (Ha and Pearce 1999): The coefficients of window are computed from the following equation:

$$w(n) = 0.62 - 0.48\left|\left(\frac{n}{N} - 0.5\right)\right| + 0.38\cos(2\pi(\frac{n}{N} - 0.5)) \tag{4}$$

where $0 \le n \le N$ and the window length is $L = N + 1$.

Bartlett (Oppenheim and Schafer 1999)

$$w(n) = \begin{cases} \frac{2n}{N}.\ 0 \le n \le \frac{N}{2} \\ 2 - \frac{2n}{N}.\frac{N}{2} \le n \le N \end{cases} \tag{5}$$

The window length $L = N + 1$.

Blackman (Oppenheim and Schafer 1999)

$$w(n) = 0.42 - 0.5\cos(\frac{2\pi n}{L-1}) + 0.08\cos(\frac{4\pi n}{L-1}).\ 0 \le n \le M - 1 \tag{6}$$

where M is $N/2$ when N is even and $(N+1)/2$ when N is odd.

Rectangular (Oppenheim and Schafer 1999),

$$w(n) = \begin{cases} 1.\ -\frac{M-1}{2} \le n \le \frac{M-1}{2} \\ 0.\ \ otherwise \end{cases} \tag{7}$$

where M is the window length in samples (assumed odd for now).

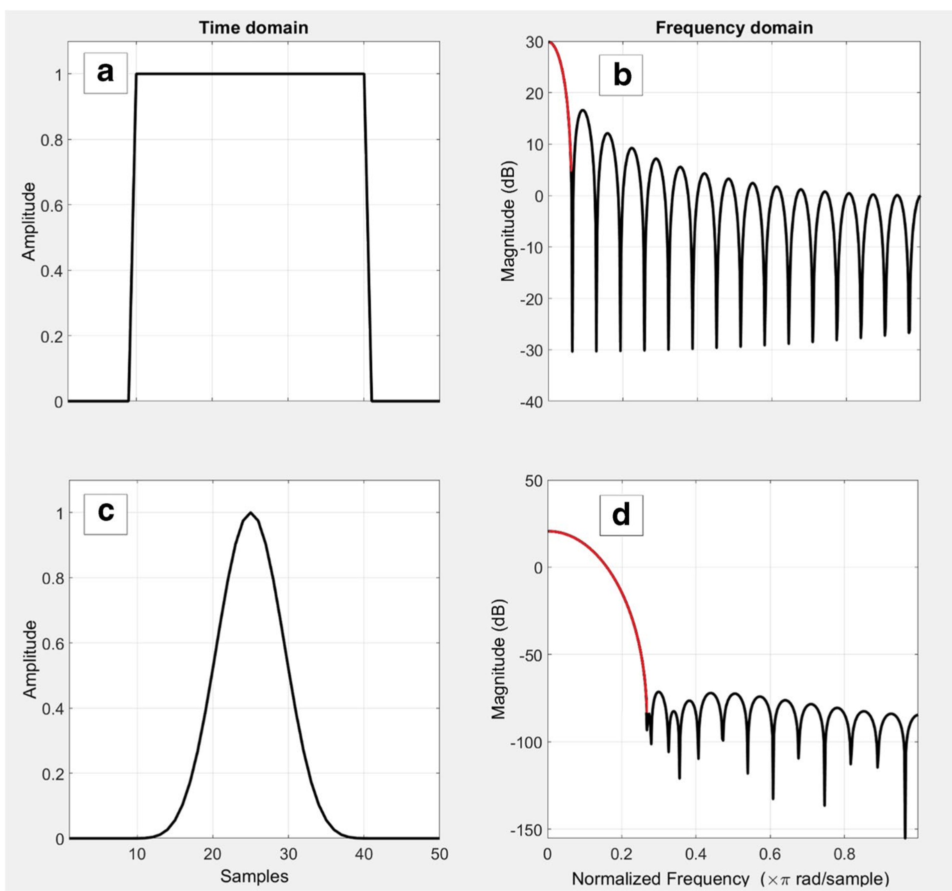


Fig. 4 Differences in the frequency response of windows. **a** Rectangular window. **b** Frequency response of rectangular window. Main lobe (red) is narrow, and peak sidelobe is high. **c** Blackman-Harris window. **d** Frequency response of Blackman-Harris window. Main lobe (red) is wide, and peak sidelobe is low

Blackman-Harris (Harris 1978).

The equation for the symmetric four-term Blackman-Harris window of length N is

$$w(n) = a_0 - a_1\cos\frac{2\pi n}{N-1} + a_2\cos\frac{4\pi n}{N-1} - a_3\cos\frac{6\pi n}{N-1}, 0 \le n \le N-1 \tag{8}$$

The equation for the periodic four-term Blackman-Harris window of length N is

$$w(n) = a_0 - a_1\cos\frac{2\pi n}{N} + a_2\cos\frac{4\pi n}{N} - a_3\cos\frac{6\pi n}{N}, 0 \le n \le N-1 \tag{9}$$

$$a_0 = 0.35875, a_1 = 0.48829, a_2 = 0.14128, a_3 = 0.01168$$

where N is the window length.

Bohman (Harris 1978)

$$w(n) = (1-|x|)\cos(\pi|x|) + \frac{1}{\pi}\sin(\pi|x|). -1 \le x \le 1 \tag{10}$$

Parzen (Harris 1978)

$$w(n) = \begin{cases} 1 - 6\left(\frac{|n|}{N/2}\right)^2 + 6\left(\frac{|n|}{N/2}\right)^3 0 \le |n| \le (N-1)/4 \\ 2\left(1-\frac{|n|}{N/2}\right)^3 (N-1)/4 < |n| \le (N-1)/2 \end{cases} \tag{11}$$

Tukey (Harris 1978)

$$w(n) = \begin{cases} \frac{1}{2}\left\{1 + \cos(\frac{2\pi}{r}\left[x - r/2\right])\right\}. & 0 \le x \le \frac{r}{2} \\ 1. & \\ \frac{1}{2}\left\{1 + \cos(\frac{2\pi}{r}\left[x - 1 + r/2\right])\right\}. & 1 - \frac{r}{2} \le x \le 1 \end{cases} \tag{12}$$

The parameter r is the ratio of cosine-tapered section length to the entire window length with $0 \le r \le 1$.

Chebyshev (Harris 2004)

$$w(n) = \frac{T_N(\beta\cos(\frac{\pi k}{N+1}))}{T_N(\beta)}. 0 \le k \le N \tag{13}$$

$T_n(x)$ is the nth Chebyshev polynomial of the first kind evaluated in x, which can be computed using

$$T_n(x) = \begin{cases} \cos\left(n\cos^{-1}(x)\right) & if\ -1 \le x \le 1 \\ \cosh(n\cosh^{-1}(x)) & if\ x \ge 1 \\ (-1)^n\cosh\left(n\cosh^{-1}(-x)\right) & if\ x \le -1 \end{cases}$$

and

$$\beta = \cosh(\frac{1}{N}\cosh^{-1}(10^\alpha))$$

Flat-Top (D'Antona and Ferrero 2006; Gade and Herlufsen 1987)

$$w(n) = a_0 - a_1\cos(\frac{2\pi n}{L-1}) + a_2\cos\left(\frac{4\pi n}{L-1}\right) - a_3\cos\left(\frac{6\pi n}{L-1}\right) + a_4\cos(\frac{8\pi n}{L-1}) \tag{14}$$

where $0 \le n \le L-1$. The coefficient values are:

$$a_0 = 0.21557895, a_1 = 0.41663158, a_2 = 0.277263158, a_3 = 0.083578947, a_4 = 0.006947$$

Gaussian (Roberts and Mullis 1987)

$$w(n) = e^{\frac{1}{2}(a\frac{n}{(L-1)/2})^2 = e^{-n^2/2\sigma^2}} \tag{15}$$

where $-(L-1)/2 \le n \le (L-1)/2$, and α is inversely proportional to the standard deviation, σ, of a Gaussian random variable. The exact correspondence with the standard deviation of a Gaussian probability density function is $\sigma = (L-1)/(2\alpha)$.

Hamming (Oppenheim and Schafer 1989)

$$w(n) = 0.54 - 0.46\cos\left(2\pi\frac{n}{N}\right).0 \le n \le N \tag{16}$$

The window length $L = N + 1$.

Hanning (Oppenheim and Schafer 1989)

$$w(n) = 0.5\left(1 - \cos\left(2\pi\frac{n}{N}\right)\right).0 \le n \le N \tag{17}$$

The window length $L = N + 1$.

Triangular (Oppenheim and Schafer 1989).

The coefficients of a triangular window are:

For L odd:

$$w(n) = \begin{cases} \frac{2n}{L+1} 1 \le n \le \frac{L+1}{2} \\ 2 - \frac{2n}{L+1}\ \frac{L+1}{2} + 1 \le n \le L \end{cases} \tag{18}$$

For L even:

$$w(n) = \begin{cases} \frac{2n-1}{L} 1 \le n \le \frac{L}{2} \\ 2 - \frac{2n-1}{L}\ \frac{L}{2} + 1 \le n \le L \end{cases} \tag{19}$$

Kaiser (Kaiser 1974): The coefficients of a Kaiser window are computed from the following equation:

$$w(n) = \frac{I_0(\beta\sqrt{1 - (\frac{n-N/2}{N/2})^2})}{I_0(\beta)}. 0 \le n \le N \tag{20}$$

$$\beta = \begin{cases} 0.1102(a - 8.7). & a \ge 50 \\ 0.5842(a-21)^{0.4} + 0.7886(a-21). & 50 \ge a \ge 21 \\ 0 & a < 21 \end{cases}$$

Nuttall (Nuttall 1981).

The equation for the *symmetric* Nuttall defined four-term Blackman-Harris window is

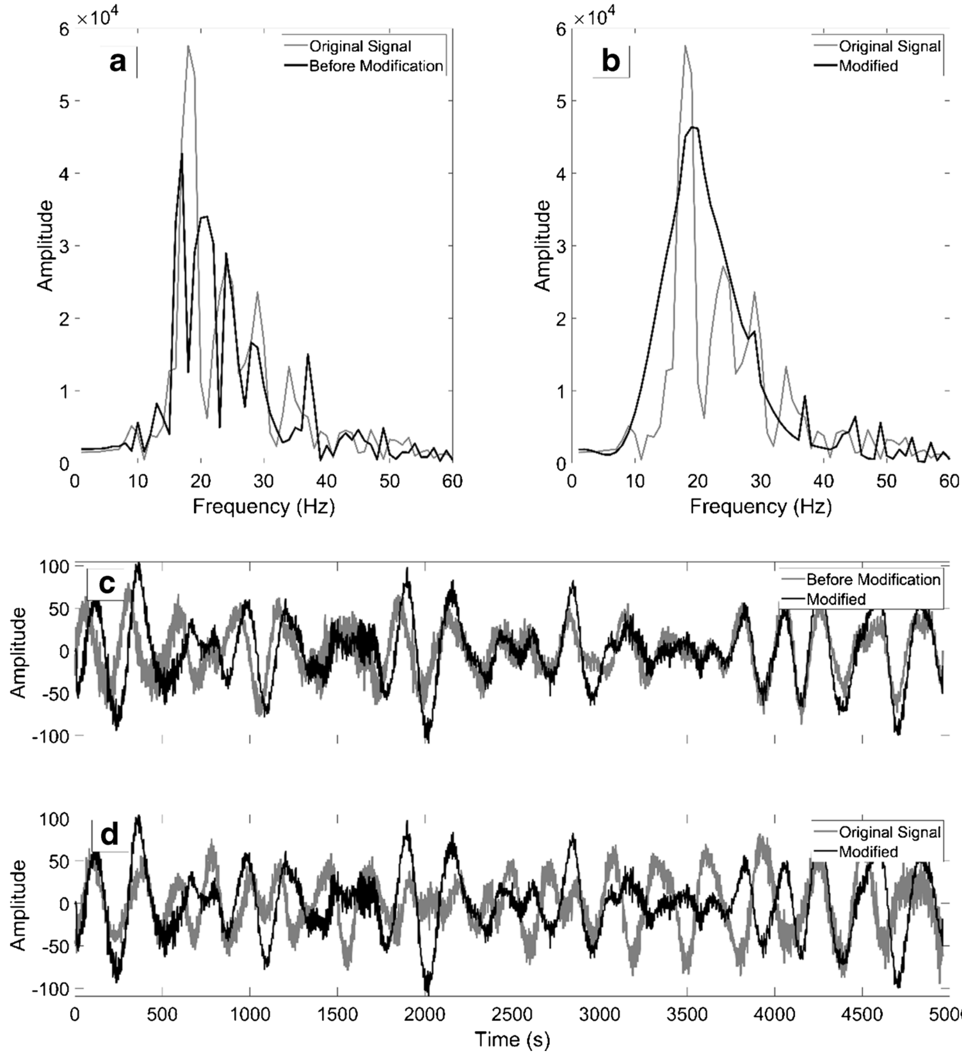


Fig. 5 Effects of modifications in the time and frequency domains. **a** Frequency content of the original and the primary signal before modification. **b** Frequency content of the original and the primary signal after modification. **c** Primary signal before modification and after modification. **d** Original signal and modified primary signal in the time domain

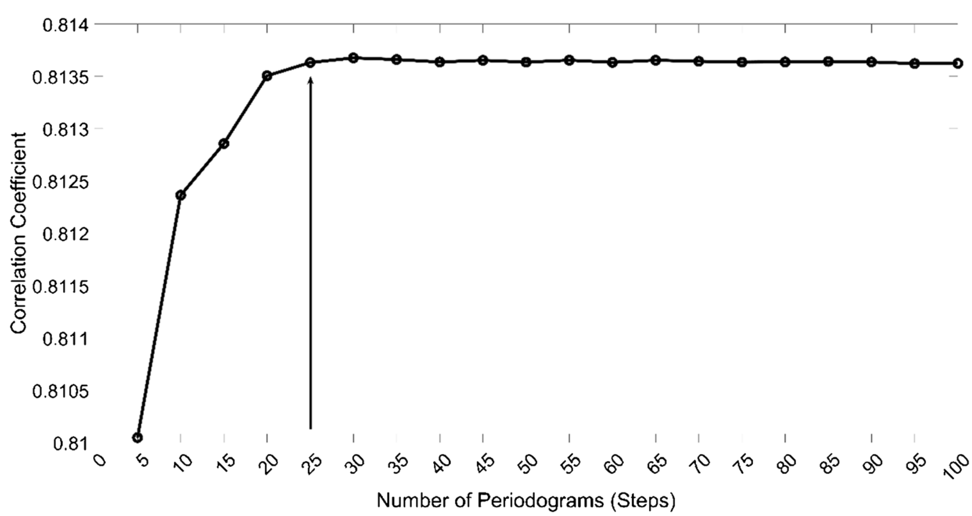
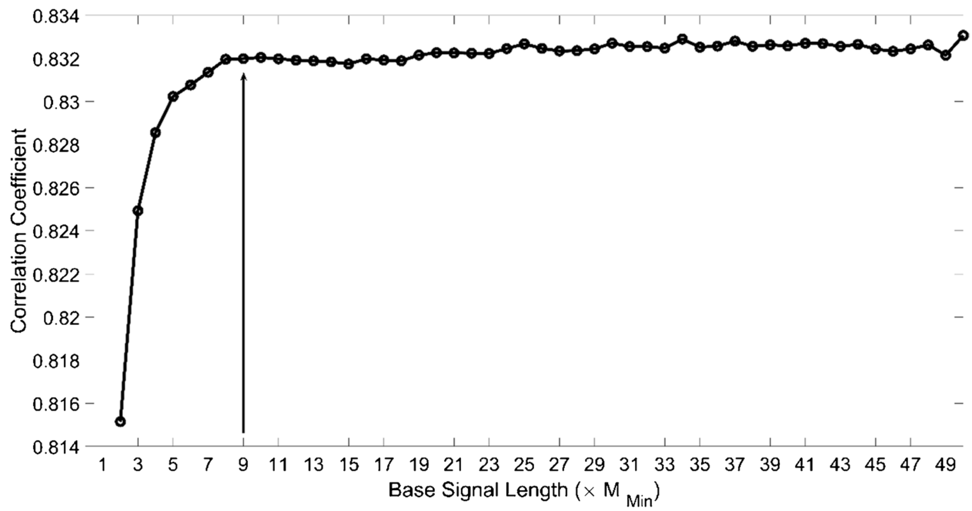


Fig. 6 Performance of different number of periodograms (steps) (the arrow indicates the optimal correlation coefficient)

Fig. 7 A sample of achieved results of analyzing different ratios of the minimum length for base signal length in a 1-min gap (the arrow indicates the optimal correlation coefficient)



$$w(n) = a_0 - a_1\cos\frac{2\pi n}{N-1} + a_2\cos\frac{4\pi n}{N-1} - a_3\cos\frac{6\pi n}{N-1}, 0 \le n \le N-1 \quad (21)$$

where $n = 0,1,2, \ldots N-1$.

The equation for the *periodic* Nuttall defined four-term Blackman-Harris window is

$$w(n) = a_0 - a_1\cos\frac{2\pi n}{N} + a_2\cos\frac{4\pi n}{N} - a_3\cos\frac{6\pi n}{N}, 0 \le n \le N-1 \quad (22)$$

where $n = 0,1,2, \ldots N-1$. The periodic window is N periodic.

The coefficients for this window are a_0=0.3635819, a_1 = 0.4891775, a_2 = 0.1365995, and a_3 = 0.0106411.

Taylor (Carrara et al. 1995; Brookner 1991).

Taylor windows are similar to Chebyshev windows. A Chebyshev window has the narrowest possible mainlobe for a specified sidelobe level, but a Taylor window allows you to make tradeoffs between the mainlobe width and the sidelobe level. The Taylor distribution avoids edge discontinuities, so Taylor window sidelobes decrease monotonically. Taylor window coefficients are not normalized. Taylor windows are typically used in radar applications, such as weighting synthetic aperture radar images and antenna design.

The periodograms were then stacked, resulting in the production of the primary periodogram. Stacking is one of the most important processes in many seismological studies and means summing and averaging traces (Lee, Kanamori, Jennings, and Kisslinger, 2002) (Stein and Wysession, 2003). After that, by normalizing the primary periodogram (dividing to its maximum), the values become between 0 and 1 that were considered as coefficients for each frequency. In the next step, the Fourier transform of the base signal before and after the gap was calculated. The average of maximum absolute values of these two Fourier transforms was considered as the maximum power of the area around the gap. This maximum power was multiplied by the coefficients of the primary periodogram. Consequently, the "*reference periodogram*" was achieved which was an estimate of the frequency content of the gap.

As mentioned already, the purpose of this study was to improve the frequency content of the primary signal using the reference periodogram. To achieve this, using the Fourier transform, the primary signal was converted into the frequency domain. As a result, complex numbers for each frequency were obtained. Using the absolute value of these complex numbers, the corresponding amplitude was assigned to each frequency. The lengths of the reference periodogram and Fourier transform of the primary signal must become the same. Regarding complex numbers as the outputs of the primary signal Fourier transform, and considering the suggested power obtained from the reference periodogram, complex numbers were modified by changing the real part without changing the imaginary part. For instance, at a certain frequency, the complex number $4+3i$ whose absolute value is equal to 5 was obtained from the Fourier transform. Hence, if the suggested power for this frequency is 6, by changing the real part, the absolute value will be modified to 6.

Table 2 The length of time-series around the gap affects the results. Base signal length for each gap scenario is shown

Gap length (min)	Base signal length ($\times L_{min}$)
1	9
5	10
10	11
15	11
20	19
25	30

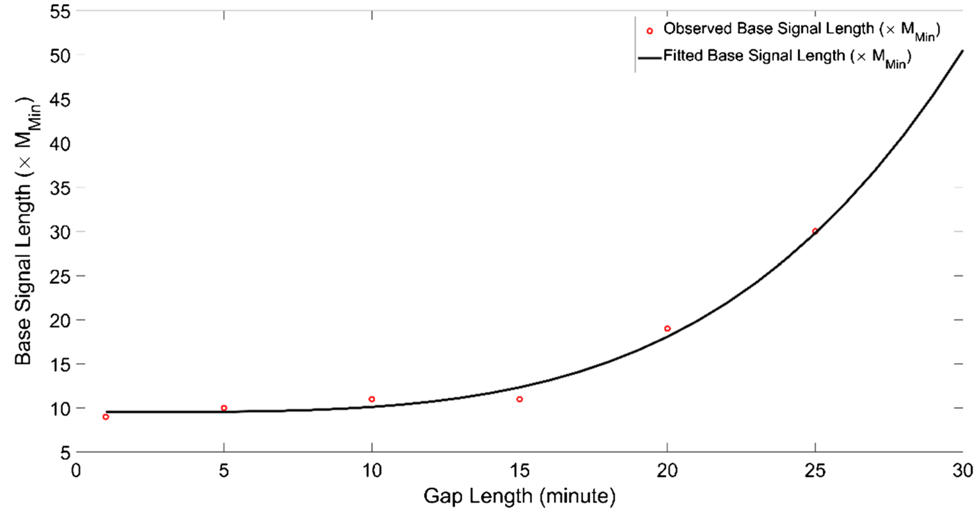
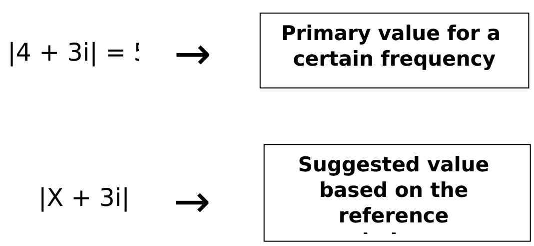


Fig. 8 Fitted curve for optimal ratios of minimum length for the base signal in the different scenario



As a result, the complex number changes to $5.1 + 3i$, and by inverse Fourier transform, the signal returns to the time domain.

Results and discussion

As already mentioned, this technique was applied to the 1-h data of 26 seismic stations with different gap scenarios. The results showed that these modifications increase the absolute value of the frequency content correlation coefficient of the primary signal with the original signal in different scenarios by 6.7 to 8.7% and cause changes in the amplitude and phase of the signal in the time domain (Fig. 5 c).

The frequency content of the gap is estimated by stacking the periodograms obtained from the signals around the gap. Stacking reduces random data and assigns high power to the correct frequencies, but reduces low power accuracy (Fig. 5 b). It is worth mentioning that stacking sometimes reduces the data containing negative values to zero. Since the periodogram is obtained from the square values of the power spectrum, smoothing in the frequency domain does not cause a loss of a certain frequency and only weakens or amplifies the frequency. Therefore, the more periodograms produced, the better the results will be. However, the results demonstrated that values more than 25 for the numbers of periodograms (steps) for each side of the base signal did not make much difference in the results (Fig. 6).

Using the frequency content of the time series around the gap (as a base signal) is one of the main parts of this technique, the length of which affects the results. The length should be long enough to cover the frequency content; however, it should not be too long to be unavailable. To investigate this issue further, the rate of changes in correlation coefficients for different lengths was considered, and the first point with a minimum change rate (derivative $\rightarrow 0$) was chosen as the best length. Results for each gap scenario are shown in Fig. 7 (a sample) and tabulated in Table 2 (the base signal length is a coefficient of L_{Min} in Eq. 1). Based on these results, an equation was fitted which can be a help for other gap lengths (Fig. 8):

$$f(x) = a \times x^b + c \tag{23}$$

Coefficients (with 95% confidence bounds):

$$a = 7.877 \times 10^{-5},\ b = 3.87,\ c = 9.534$$

Table 3 Performance of the type of window affects for each gap scenario. The results of the best performance of windows for each gap scenario are shown

Gap length (min)	Number of tests	Increase of correlation coefficient (%)	Best window
1	1456	7.003	Chebyshev
5	260	7.0455	Chebyshev
10	104	6.7148	Nuttall
15	52	7.7348	Blackman-Harris
20	26	8.6991	Blackman-Harris
25	26	8.7951	Blackman-Harris

According to Eq. 23, as the gap length increases, the base signal length is as well. This value becomes too large for a long gap, which indicates that this method does not work well in such cases.

As described, the type of window affects the subsequent processing and produces different results. The results of the best performance of windows for each gap scenario are shown in Table 3 and Fig. 9. In addition, the length of the window also affects the results. Given that the main purpose was to assign high powers to the correct frequency range by stacking the periodograms, the length of the window should be determined in such a way that provides the appropriate number of periodograms from the base signal. By reducing the length of the window, more periodograms can be prepared, but this length cannot be less than the minimum length. As the window length increases, the main lobe of the frequency response becomes narrower but the peak sidelobes increase slightly. In contrast, as the window length decreases, the main lobe widens, but the peak sidelobes decrease slightly (Fig. 10). Ideally, the main lobe should be extremely narrow and the peak sidelobes extremely small (i.e., impulse). In this case, when it is convolved with a

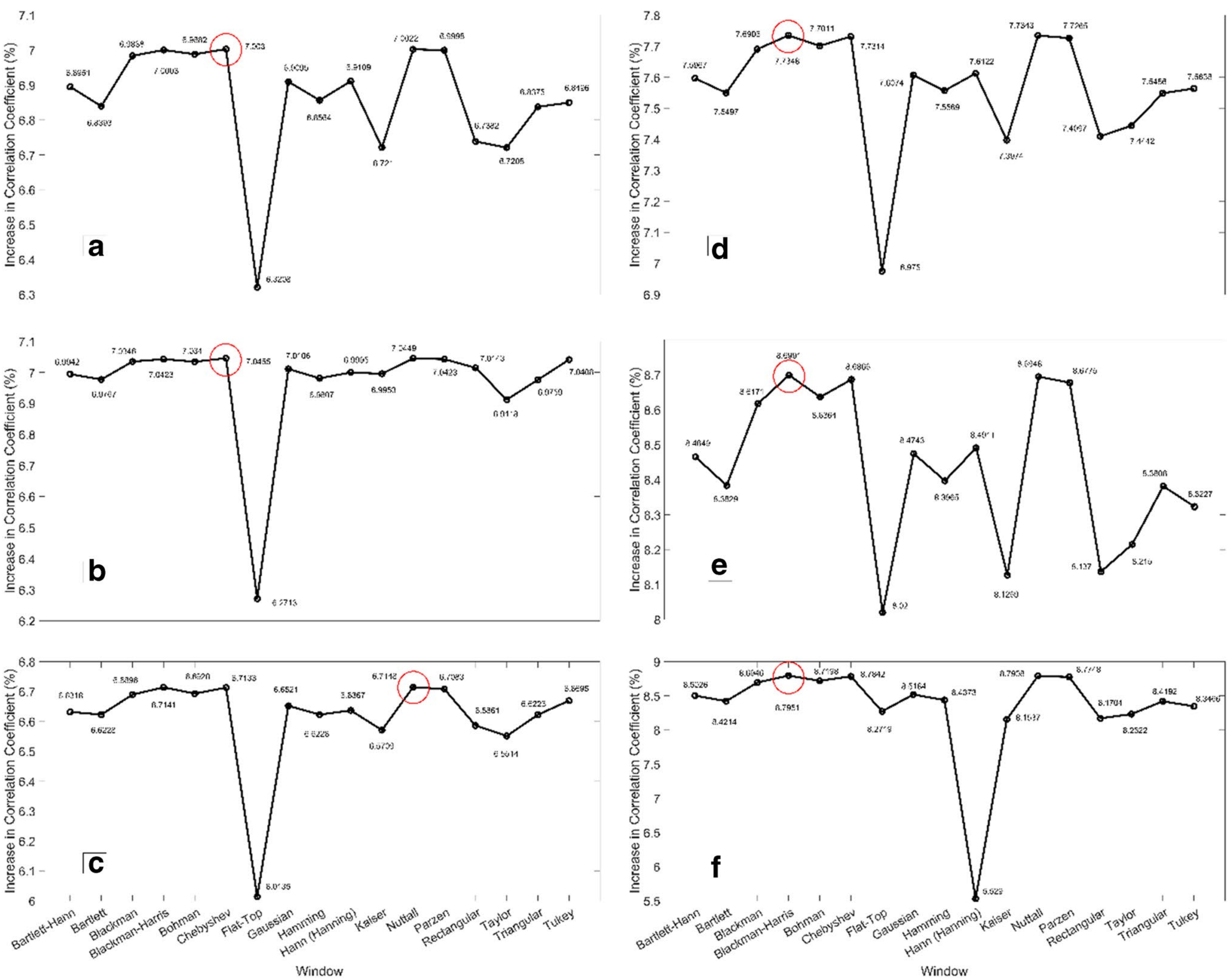


Fig. 9 Performance of different windows. **a** For the 1-min gap. **b** For the 5-min gap. **c** For the 10-min gap. **d** For the 15-min gap. **e** For the 20-min gap. **f** For the 25-min gap. (Red circle indicates the best window)

Fig. 10 The effect of window length on frequency response. **a** Blackman-Harris window with a length of 64 samples. **b** Frequency response of Blackman-Harris window with a length of 64 samples. **c** Blackman-Harris window with a length of 15 samples. **d** Frequency response of Blackman-Harris window with a length of 15 samples. The main lobe (red) is wider than the window with a length of 64 samples, and the peak sidelobe is a bit lower

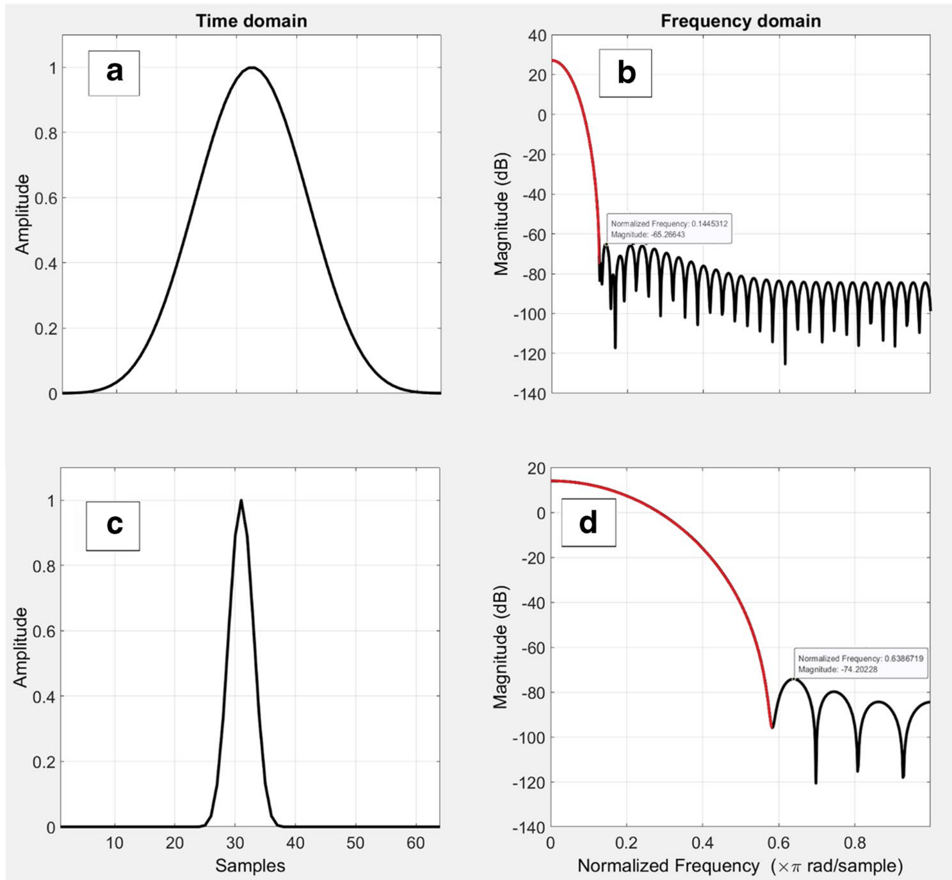
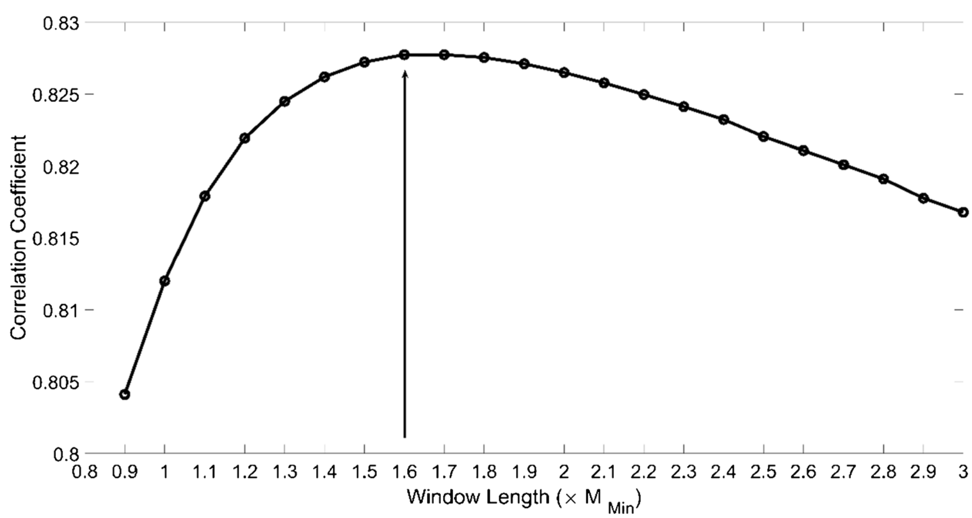


Fig. 11 A sample of achieved results of analyzing different ratios of the minimum length for window length (the arrow indicates the maximum correlation coefficient)



signal, it makes very small changes in it (Oppenheim and Schafer 1999).

Larger windows are closer to ideal, but increasing the length of the window makes it impossible to obtain the required number of periodograms from the base signal. Different coefficients of the minimum length were investigated to find the appropriate length of the windows. The results exhibited that in all windows, lengths less than the minimum length

Table 4 The most suitable length for each window based on achieved results

Window	The most suitable length ($\times L_{\text{Min}}$)
Bartlett-Hann	1.2
Bartlett	1.1
Blackman	1.5
Blackman-Harris	1.6
Bohman	1.5
Chebyshev	1.6
Flat-Top	3
Gaussian	1.2
Hamming	1.2
Hann (Hanning)	1.2
Kaiser	1.1
Nuttall	1.6
Parzen	1.6
Rectangular	1.1
Taylor	1.1
Triangular	1.1
Tukey	1.1

showed poor performance and each window had its optimal length, which is shown in Fig. 11 and tabulated in Table 4.

Conclusion

In this study, we aimed to fill the gap of seismic ambient noise recorded by broadband sensors by introducing a novel technique using time series around the gap. Our findings demonstrate that by modifying the frequency content of the time series taken from after the gap, we were able to create an appropriate time series to fill the gap. The analysis conducted in this study revealed that applying these modifications to the primary signal generates frequency content that is more correlated with the gap. However, it is important to note that there were rare cases (less than 1% of the total analysis) where this technique resulted in reduced correlation. Our technique estimates the frequency content of the gap through stacking the periodograms calculated from signals around the gap. The results indicate that increasing the number of periodograms improves the estimation, but after reaching a certain number, further improvement will not have a significant effect. Moreover, we observed that different windows used in calculating the frequency content using the periodogram yield varying results. It is crucial to consider that the window length significantly impacts the outcomes, and each window performs optimally at its specific length. Our practical results demonstrate that the Blackman-Harris, Chebyshev, and Nuttall windows exhibited higher efficiency compared to other windows in this technique. Given the wide frequency range covered by broadband sensors, utilizing the existing time series taken from after the gap enables us to restore the embedded time series effectively, ensuring that high powers are more accurately recovered through modification operations.

Author contribution Mohammad Reza Nameni: writing and edition, original draft preparation, methodology, computation, and programming. SeyedMohammadSadegh Jafari: writing and edition, original draft preparation, methodology, computation, and programming. Habib Rahimi: conceptualization, methodology, and review and edition.

Data availability All data applied to accomplish this research will be available under reasonable demand.

Code availability • Name of code: FillGap.

Developers and contact address: Mohammad Reza Nameni, Institute of Geophysics, University of Tehran (IGUT), nameni.mreza@ut.ac.ir + 989,156,553,989.

• Seyedmohammadsadegh Jafari, SEISRAM, Department of Mathematics and Geosciences, University of Trieste, seyedmohammadsadegh.jafari@phd.units.it + 393,792,420,222.

• Habib Rahimi, Professor, Institute of Geophysics, University of Tehran (IGUT), rahimih@ut.ac.ir (+ 98) 21–61,118,240.

• Year first available: 2022.

• Hardware required: PC with 8 GB of RAM.

• Software required: MATLAB 2020a or higher to run the open-source code; the free MATLAB Compiler Runtime R2021a (9.10)—Windows OS 64-bit for the compiled code.

• Program language: MATLAB (R2021a).

• Program size: 8 Kb.

• How to access to source code: The open-source FillGap Matlab code is available on the GitHub link https://github.com/mreza-nameni/FillGap

Declarations

Competing interests The authors declare no competing interests.

References

Ambrosino F, Thinová L, Briestenský M, Sabbarese C (2019) Anomalies identification of earth's rotation rate time series (2012–2017) for possible correlation with strong earthquakes occurrence. Geodesy Geodynam 10(6):455–459

Ansari A, Rao KS, Jain AK (2022a) Seismic vulnerability of tunnels in Jammu and Kashmir during post-seismic functionality. Geotech Geol Eng 40(11):1–26. https://doi.org/10.1007/s10706-022-02341-0

Ansari A, Rao KS, Jain AK, Ansari A (2022b) Deep learning model for predicting tunnel damages and track serviceability under seismic environment. Model Earth Syst Environ 8(4):1–20. https://doi.org/10.1007/s40808-022-01556-7

Ansari A, Rao KS, Jain AK (2023a) Seismic response and fragility evaluation of circular tunnels in the Himalayan region: implications for post-seismic performance of transportation infrastructure projects in Jammu and Kashmir. Tunn Undergr Space Technol 137:1–13. https://doi.org/10.1016/j.tust.2023.105118

Ansari A, Zahoor F, Rao KS, Jain AK (2023b) Seismic response and vulnerability evaluation of Jammu Region (Jammu and Kashmir).

Indian Geotechnical Journal 53(3):509–522. https://doi.org/10.1007/s40098-022-00694-0
Baisch S, Bokelmann GH (1999) Spectral analysis with incomplete time series: an example from seismology. Comput Geosci 25:739–750. https://doi.org/10.1016/S0098-3004(99)00026-6
Barnhart BL, Nandage HK, Eichinger W (2011) Assessing discontinuous data using ensemble empirical mode decomposition. Adv Adapt Data Anal 3(4):483–491. https://doi.org/10.1142/S179353691100091X
Briestenský M, Ambrosino F, Smetanová I, Thinová L, Šebela S, Stemberk J, Pristašová L, Pla C, Benavente D (2021) Radon in dead-end caves in Europe. J Cave Karst Stud 84(2):41–50. https://doi.org/10.4311/2021es0101
Brookner, E. (1991). Practical Phased Array Antenna Systems. Lex Book.
Broomhead DS, King GP (1986) Extracting qualitative dynamics from experimental data. Physics D 20:217–236. https://doi.org/10.1016/0167-2789(86)90031-X
Broomhead DS, Jones R, King GP (1987) Topological dimension and local coordinates from time series data. J Phys A 20:563–569. https://doi.org/10.1088/0305-4470/20/9/003
Carrara WG, Majewski RM, Goodman RS (1995) Spotlight Synthetic Aperature Radar Signal Processing Algorithms. Artech House
D'Antona G, Ferrero A (2006) Digital Signal Processing for Measurement Systems Theory and Applications. Springer
Foster G (1996) Wavelets for period analysis of unevenly sampled time series. Astron J 112:1709–1729. https://doi.org/10.1086/118137
Gade S, Herlufsen H (1987) Use of weighting functions in DFT/FFT analysis (part I). Brüel & Kjær, Windows to FFT Analysis (Part I) Technical Review, 3, 1–28
Ghil M, Allen MR, Dettinger MD, Ide K, Kondrashov D, Mann ME ... Yiou P (2002) Advanced spectral methods for climatic time series. Rev Geophys, 40, 1–41. https://doi.org/10.1029/2000RG000092
Ha Y, Pearce JA (1999) A new window and comparison to standard windows. IEEE Trans Acoust Speech Signal Process 37(2):298–301. https://doi.org/10.1109/29.21693
Harris FJ (1978) On the use of windows for harmonic analysis with the discrete Fourier transform. Proc IEEE 66(1):51–83. https://doi.org/10.1109/PROC.1978.10837
Harris FJ (2004) Multirate Signal Processing for Communication Systems. Prentice Hall
Hloupis G, Moisidi M, Vallianatos F, Stonham J, Makris J, Triantis D (2004) Application of CLEAN algorithm in incomplete mictrotremors recordings. Bull Geol Soc Greece. https://doi.org/10.12681/bgsg.16471
Hocke K, Kämpfer N (2008) Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram. Atmos Chem Phys Discuss 8:4603–4623. https://doi.org/10.5194/ACP-9-4197-2009
Huang NE, Shen Z, Long SR (1999) A new view of nonlinear water waves: the Hilbert spectrum. Annu Rev Fluid Mech 31:417–457. https://doi.org/10.1146/ANNUREV.FLUID.31.1.417
Kaiser, J. (1974). Nonrecursive Digital Filter Design Using the I0- sinh Window Function. IEEE Symposium Circuits and Systems.
Kondrashov D, Ghil M (2006) Spatio-temporal filling of missing points in geophysical data sets. Nonlinear Process Geophys 13:151–159. https://doi.org/10.5194/npg-13-151-2006
Lay T, Wallace TC (1995) Modem Global Seismology. Academic Press
MacDonald GJ (1989) Spectral analysis of time series generated by nonlinear processes. Rev Geophys 27:449–469. https://doi.org/10.1029/RG027i004p00449
Moghtaderi A, Borgnat P, Flandrin P (2012) Gap-filling by the empirical mode decomposition. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). https://doi.org/10.1109/ICASSP.2012.6288750
Musial JP, Verstraete MM, Gobron N (2011) Technical Note: comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series. Atmos Chem Phys 11:7905–7923. https://doi.org/10.5194/acp-11-7905-2011
Nakata N, Gualtieri L, Fichtner A (2019) Seismic ambient noise. Cambridge University Press. https://doi.org/10.1017/9781108264808
Nuttall AH (1981) Some windows with very good sidelobe behavior. IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-29(1), 84–91. https://doi.org/10.1109/TASSP.1981.1163506
Oppenheim A, Schafer RW (1989) Discrete-Time Signal Processing. Prentice Hall
Oppenheim A, Schafer RW (1999) Discrete-Time Signal Processing. Prentice Hall
Pascual-Granado J, Garrido R, Suárez JC (2015) MIARMA: a minimal-loss information method for filling gaps in time series Application to CoRoT light curves. Astronomy & Astrophysics, 575(A78). https://doi.org/10.1051/0004-6361/201425056
Reynolds RW, Smith TM (1994) Improved global sea-surface temperature analysis using optimum interpolation. J Clim 7:929–948. https://doi.org/10.1175/1520-0442(1994)007/3C0929:IGSSTA/3E2.0.CO;2
Roberts RA, Mullis CT (1987) Digital Signal Processing. Addison-Wesley
Sadeghisorkhani H (2017) Analyses and Application of Ambient Seismic Noise in Sweden: Source, Interferometry. Thesis, Faculty of Science and Technology, Uppsala University, Tomography
Scargle JD (1989) Studies in astronomical time series analysis. III. Fourier transforms, autocorrelation and cross-correlation functions of unevenly spaced data. Astrophys J 343:874–887. https://doi.org/10.1086/167757
Shearer PM (2009) Introduction to Seismology. Cambridge University Presshttps://doi.org/10.1017/CBO9780511841552
Shumway RH, Stoffer DS (2017) Time series analysis and its applications: with R examples (3rd ed.). Springer
Sidekerskiene T, Damasevicius R (2016) Reconstruction of Missing Data in Synthetic Time Series Using EMD
Stoica P, Moses RL (1997) Introduction to Spectral Analysis. Prentice Hall
Vautard R, Yiou P, Ghil M (1992) Singular-spectrum analysis: a toolkit for short, noisy chaotic signals. Physica D 58:95–126. https://doi.org/10.1016/0167-2789(92)90103-T
Xue J, Nie B, Smirni E (2017) Fill-in the Gaps: Spatial-Temporal Models for Missing Data. 13th International Conference on Network and Service Management (CNSM). https://doi.org/10.23919/CNSM.2017.8255983