Improvement in the Empirical Green's Function Extraction Using Root Mean Square Ratio Stacking

Safarkhani, M.1* and Shirzad, T.2

1. M.Sc. Graduate, Department of Earth Physics, Institute of Geophysics, University of Tehran, Tehran, Iran
2. Assistant Professor, Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland

(Received: 17 June 2019, Accepted: 21 Jan 2020)

Abstract
Seismic interferometry is an efficient technique to extract the Empirical Green's Function (EGF) between station pairs when the source is considered at one of the stations. The geometry and energy flux of asymmetric noise sources have unavoidable impacts on the extracted EGFs, deduced from ambient seismic noise recorded in pairs of stations. In this study, to consider these effects, three methods of noise correlation functions stacking (linear, root mean square, root mean square ratio) are investigated using synthetic and real data processing. During synthetic data processing, effects of the noise sources geometry and energy flux inside and outside the Fresnel zone are examined. After separating stationary and non-stationary sources, the results have shown that the root mean square ratio contains the least effects of non-stationary signals compared to other methods of stacking. Moreover, comparison of the EGFs from the recorded data in Azerbaijan (NW Iran), indicates that the signal retrieved by root mean square ratio is more reliable than the other stacking methods' signals (e.g., linear, root mean square).

Keywords: Asymmetric distribution of noise energy flux, Empirical Green's functions, Fresnel zone, Non-stationary signals, Root Mean Square Ratio stacking.

1. Introduction
Seismic interferometry is a method that predicts the cross-correlations of all combinations of noise recorded yields to the Green's function between two receivers. In recent years, extracting empirical Green's functions (hereafter EGFs) from long-time recorded ambient seismic noise have been increasingly important (e.g. Lobkis and Weaver, 2001; Derode et al., 2003; Wapenaar, 2004; Snieder, 2004; Roux et al., 2005; Wapenaar et al., 2006; Snieder et al., 2007). These studies show that ambient seismic noise contains valuable information regarding wave propagation in the medium (Shapiro et al., 2005; Roux et al., 2005; Sabra et al., 2005). In the ideal uniform sources distribution, cross-correlations of ambient seismic noise recorded at two stations, yield an inter-station EGF (Weaver and Lobkis, 2001; Snieder, 2004; Gouédard et al., 2008; Tsai, 2009). The results produced by this cross-correlation technique have been applied in reconstruction of the Earth's velocity and anisotropy structures in different regions especially in the aseismic area (e.g. Shapiro and Campillo, 2004; Prieto et al., 2009). Uniform distribution of noise sources or energies in different azimuths is the main condition in determining the accurate EGF between the station pairs using the interferometry methods (Wapenaar, 2004; Stehly et al., 2006). Obviously, the condition for the homogeneous distribution of sources and energies is not always established in real-world conditions. In other words, the distribution of sources and energies around the station pairs is sufficiently symmetrical only in rare cases. Recent studies regarding noise source distribution show the dominant presence of them in oceanic regions (Stutzmann et al., 2009; Landes et al., 2010). Due to the presence of noise sources in oceanic areas and their severe seasonal variations (Sens-Schönfelder and Wegler, 2006; Meier et al., 2010), the distribution of sources is anisotropic and directional (Stehly et al., 2008). Failure to properly distribute these parameters leads to an inaccurate determination of the Green's function (Snieder and Sens-Schönfelder, 2015; Liu and Ben-Zion, 2016).

Many researchers (e.g. Schuster et al., 2004; Snieder, 2004; Wapenaar et al., 2004; Roux et al., 2005; Snieder et al., 2006) believe that the existence of sources located near the receiver line (inside the Fresnel zone) has a
major role in reconstructing the EGF. This result is obtained using the integral approximation of overall sources by the stationary phase method (Snieder, 2004). We refer to these sources as stationary sources, which are inside the Fresnel zone as outlined in Figure 1. Therefore, in a uniform distribution of sources in the medium (as demonstrated in Figure 1), the stationary sources play a major role in the reconstruction of the real EGF between the station pairs. Assuming full source coverage, the energy emitted by sources outside the Fresnel zone destructively interferes in this process. In this study, we refer to these sources as non-stationary sources.

In this paper, we are looking for a solution where the long-term average of cross-correlations converges towards the accurate EGF. In this way, three methods of stacking including linear (hereafter LIN, see Bensen et al., 2007), root mean square (hereafter RMS, see Shirzad and Shomali, 2013), root mean square ratio (hereafter RMS-R, see Safarkhani and Shirzad, 2019) has been investigated with synthetic and real data.

2. Data set

Figure 1 shows the synthetic data for a pair of stations A and B with coordinates [-4, 0], [4, 0] and the inter-station distance of 8 km. Stations A [-4, 0] and B [4, 0] are surrounded with 144 sources located in a circle with a radius 40 km (see Figure 1). The distribution of 144 sources with a 2.5-degree interval on the circle (360-degree azimuthal coverage) around the station pair depicted in the right panel of the Figure 1. Based on the Figure 1, the numbering of the sources starts from the east part of the studied region and it continues in the counterclockwise direction along the line passes through between A and B station pair. For more regularity, sources are named with four-digit characters including S and counter (e.g. S001, S002 and etc.; see right panel of Figure 1).

All functions for sources are the negative normalized second derivative of a Gaussian function, called the Ricker wavelet, with 10 samples per second (SPS). Figure 2 (middle panel) shows the Ricker wavelet, \( y(t) = |1-2\pi f^2 t^2| \times \exp[-\pi f^2 t^2] \), where \( f \) and \( t \) are the frequency (Hz) and time (s), that would be propagated in a half-space and flat medium (without any topography) with constant velocity (1.0 km/s) and no reflector. The waveforms recorded by these source functions for stations A and B are shown on the right and left panel of Figure 2, respectively.

![Figure 1. Sources and receivers geometry with 2.5° evenly distributed sources (yellow stars) on a circle with radius 40 km. Light gray parabolas indicate Fresnel zone (stationary and non-stationary zones), stations (red triangles A [-4, 0] and B [4, 0]). The source names are started from [40, 0] in counterclockwise with S001 to S144.](image-url)
3. The Method and some results
In this section, we examine how well we can retrieve the accurate inter-station EGF with synthetic data in two cases including (1) fully and (2) selected coverage of sources (with gaps in between and also no difference in amplitude). In the first case, three groups of source combinations will be investigated in a circle with a radius of 40 km (Figure 1). These three groups include sources inside the Fresnel zone, sources outside the Fresnel zone and all sources inside and outside the Fresnel zone. Also, in order to simplify, all groups contain signals with the same amplitude. The cross-correlation of recorded waves from these source functions for stations A and B are shown in Figure 3a. This figure shows the collection of noise correlation functions (hereafter NCFs) for the station pair A-B as a function of source number. Based on sources located inside and outside the Fresnel zone, the NCF signals are separated by the stationary (blue traces), non-stationary (gray traces) signals. Therefore, the LIN stacking has been formulated by Equation (1):

\[ EGF \approx \sum_{i=1}^{n} NCF_i = \sum_{i=1}^{n} (NCF_{i}^{\text{stationary}} + NCF_{i}^{\text{non-stationary}}) = \sum_{i=1}^{n_1} NCF_{i}^{\text{stationary}} + \sum_{i=n_1+1}^{n_2} NCF_{i}^{\text{non-stationary}} \]

where \( n \) is the number of NCFs and \( n_1 + n_2 \) equal to \( n \). Therefore, LIN stacking of aforementioned three source group combinations is depicted in Figure 3b.

![Figure 2](image-url)  
Figure 2. Ricker wavelet source function (middle panel) and recorded data at stations A (left panel) and B (right panel) respectively.

![Figure 3](image-url)  
Figure 3. (a) Fully distribution of stationary and non-stationary NCFs are shown by blue and gray signals as a function of source number. (b) Indicates the LIN stacking NCF signals corresponding to sources located inside the Fresnel zone (blue signal), outside the Fresnel zone (gray signal) and all sources (black signal).
Since the NCFs of the stationary and non-stationary sources are completely independent of each other, Equation (1) can be expressed as:

$$ EGF \approx \sum_{i=1}^{n_1} NCF_{i}^{\text{stationary}} + \sum_{i=1}^{n_2} NCF_{i}^{\text{non-stationary}} $$  \hspace{1cm} (2)

In the full coverage of sources, a comparison of black and blue signals in Figure 3b indicates that the NCFs of stationary zone contribute constructively to retrieve inter-station EGF. Also, the non-stationary signals affecting the reconstructed inter-station EGF are almost eliminated by stacking as shown in this figure. Hence, Equation (2) can be summarized as a good approximation:

$$ EGF \approx \sum_{i=1}^{n_1} NCF_{i}^{\text{stationary}} $$

$$ \sum_{i=1}^{n_2} NCF_{i}^{\text{non-stationary}} \approx 0 $$ \hspace{1cm} (3)

The isotropic distribution of noise sources is seldom possible in the real earth model. In the second case, in order to consider a more realistic example, seven NCF signals excited by sources in the pure stationary (S005, S072), pure non-stationary (S105) and mixed stationary and non-stationary (S030+S105+S072, S005+S030+S105, S005+S030, S105+S072) zones are shown in Figure 4. The sources are located at a radius of 40 km and are randomly selected. With this accidental selection, the random nature of the sources/random wavefield/random excitation is taken into account for calculating the Green's function. Notably, in this case, it is not possible to estimate the effect of each segment on the retrieved signal by changing all parameters of the ideal condition simultaneously. Thus the amplitudes (energy) of all signals are equal.

![Figure 4](image-url)

**Figure 4.** Selected distribution of sources within pure stationary (a), pure non-stationary (b, c), mixed stationary and non-stationary (d, e, f, g).
Because of stationary and non-stationary sources existence in a signal (e.g. \textbf{S105+S072}), Equation (2) does not apply in this case. Therefore, in order to retrieve reliable inter-station EGF, the NCF signals are investigated by the LIN, RMS (Shirzad and Shomali, 2013) as well as the RMS-R stacking method (Safarkhani and Shirzad, 2019). The RMS stacking method, which was first introduced by Shirzad and Shomali (2013) includes the following steps: first, the RMS values should be calculated for expected signal windows associated to each positive and negative elapse-time sides of NCFs. In this example, the expected signal window is between 5 and 9 s that is depicted by green column in Figure 5a. Then, after sorting the RMS values, the RMS curves are defined. In the next procedure, a threshold value is selected based on the change in the gradient of RMS curves.

In this study, we improve RMS stacking method by defining new constraint of data selecting and call it RMS-R stacking method. According to this issue, zero-lag's RMS value of NCFs is computed in addition to positive and negative elapse-time sides. The zero-lag is defined from zero to start of the signal window in this study. Then, RMS curve was defined based on the ratio of RMS values in positive and negative elapse-time sides to zero-lag's ones for each inter-station individually.

In other words, this NCF selection (RMS-R= RMS_{expected\_signal\_window}/RMS_{zero\_lag\_window} of NCF signal) defined as a ratio of the RMS value of stationary and non-stationary signals (Safarkhani and Shirzad, 2019). Based on this illustration, Equation (1) can be simply rewritten as:

$$EFG \approx \sum_{i=1}^{m} \{NCF_i^{\text{stationary}} + NCF_i^{\text{non-stationary}}\} \quad (4)$$

where \(m\) is the number of selected NCFs based on RMS-R condition, and always \(m \leq n\). The result of LIN, RMS, and RMS-R stacking of seven (stationary and non-stationary) signals are shown in Figure 5b.

In order to compare stacking methods on the real dataset, we used continuous data from seven seismic stations deployed in the Azerbaijan region (NW Iran) as part of the permanent array of the Iranian Seismological Center (IrSC). Seven stations in the inter-station distance range between 47 and 198 km depicted in Figure 6a. These stations are equipped with short period sensor (Kinematics SS-1) and the continuous data were recorded with 50 SPS. The vertical component of continuous seismic noise (data) was considered from December 2011 to December 2012.

![Figure 5](image_url)

Figure 5. (a) Selected distribution of stationary and non-stationary NCF signals are shown as a function of source number. In right side: the name of participating sources; and in the left side: participating signals in RMS, RMS-R, and LIN stacking methods. Expected and zero-lag windows are illustrated by green and yellow columns respectively. (b) Indicates the stacked NCF signals corresponding to (a) distribution.
Data preparation was applied based on a standard common low-frequency method, as described by Bensen et al. (2007). This method includes segmentation of the continuous raw data into 10 min time windows (Safarkhani and Shirzad, 2017), removing the mean and trend, one-bit time domain normalization, frequency whitening (0.1-0.3 Hz) to suppress the influence of earthquake signals, removing instrument irregularities and industrial noise related to the human activity (Bensen et al., 2007). Thus, after the preparation of the raw data, the prepared 10 min data were cross-correlated between all possible combinations of the station pairs. Figure 6b-f depicts the raw data recorded in SHB-SRB station pair and the steps of data processing on Dec. 28, 2012 at the same time.

![Figure 6](image)

**Figure 6.** (a) Depicts region of study (NW of Iran) and data processing steps including (b) removing mean and trend, (c) time domain normalization, (d) left panel: before, right panel: after frequency (0.1-0.3 Hz) whitening (e) cross-correlation, (f) stacking for SHB-SRB station pair with inter-station 186 km on Dec. 28, 2012.
4. Discussion

One of the main assumptions in most applications of seismic interferometry is that sources should be fairly well distributed around the station pair to recover accurate EGFs (e.g. Wapenaar and Fokkema, 2006). Each deficiency in the complete azimuthal distribution of sources directly leads to non-symmetry and deviations in the retrieved inter-station EGF signal. The simple classical LIN stacking of NCFs (from all sources) will not solve this problem, because the energy (sources) is generally non-uniform. In many studies (e.g. Stehly et al., 2006; Snieder, 2004), the researchers increase the duration of the recorded ambient seismic noise to overcome this effect. Although this simple condition would be able to enhance the SNR of retrieved EGFs in some studies, it is not a suitable approach in the presence of dependent sources or dominant energy fluxes from the non-stationary zone. However, some researchers used different types of stacking to decrease the effect of non-stationary signals on the retrieved EGF. In the synthetic example, the sources that give the main contribution to the causal and anti-causal EGFs are those in the Fresnel zone around these station pairs.

The main scope of RMS stacking is to remove the deterministic signals (e.g., earthquakes), and also to extract the coherent part of NCF signals within the signal window (Fresnel zone) as mentioned in Shirzad and Shomali (2015). However, the non-stationary energy will not be completely canceled out using RMS stacking, when the distribution of emanated energy is not uniform, and stationary and non-stationary sources are dependently excited. In this case, although the RMS stacking method can extract a signal with higher SNR than the LIN method, the artifact effects will appear in retrieved EGF as shown in Figure 7a.

Thus, to alleviate this problem, we employed an additional constraint that is typically used in the RMS stacking procedure. We have considered the source geometry as one of the main NCF selection criteria before stacking procedure to retrieve optimum EGFs and to enhance the SNR value of this signal, because the stationary energy in the NCF signals is characterized by coherency and high RMS value within the expected signal window. By contrast, non-stationary energy is characterized by incoherently high RMS value within the zero-lag window (from zero to start of the signal window). These characteristic differences were used to separate the two parts of the energy in the NCFs to obtain more accurate inter-station EGF estimations for cases including non-ideal sources coverage.

Figure 7. (a) LIN (gray), RMS (dark gray) and RMS-R (black) stacking method for SHB-SRB station pair with inter-station distance (186 km). Gray shadow shows the expected signal windows in positive and negative elapsed time sides. (b) All possible EGFs between station pairs using RMS-R stacking method in the period band of 3-10 s. The velocity range of fundamental mode Rayleigh waves is highlighted by dash-lines for positive and negative lag times.
As shown in Figure 5b, although EGF signals are approximately similar within the signal window, the RMS values within the zero-lag window are very different. For more scrupulously explanation, the retrieved EGF using RMS stacking avoids only one signal ($S_{105}$), which is indicated in the left side of Figure 5a. Whereas the RMS-R method rejects three signals ($S_{105}$, $S_{05}+S_{030}+S_{105}$, $S_{030}+S_{105}+S_{072}$) that they have RMS$_{\text{expected signal window}}$/RMS$_{\text{zero lag window}}$ less than 1. RMS amplitude within the expected signal window gives information about the energy of emitted signals by sources. Our investigation indicated that the RMS value ratio depends directly on frequency contents and the energy level in the study area. Moreover, the RMS-R method can lead to the enhancement of the stationary signal and also cancel out sufficiently non-stationary signals as it is clearly shown in Figure 7a. The figure illustrates the retrieved EGFs using three stacking methods (LIN, RMS, RMS-R) in the period band of 3-10 s for SHB-SRB station pair. The expected Rayleigh wave fundamental mode is shown by gray window. Figure 7b shows all available inter-station EGFs retrieved by RMS-R stacking method wherein the gray lines indicate velocity range of Rayleigh waves between 1.5 and 2.7 km/s.

5. Conclusion
This study compares different stacking methods (LIN, RMS and RMS-R) for extracting EGFs within a relatively small regional array in NW of Iran. The result (Figure 5b) demonstrate that in the study area, RMS-R method improved the quality and stability of the inter-station EGF significantly and decreased the effects of non-stationary NCF signals in comparison with LIN and RMS stacking methods. In addition, the RMS-R stacking method can be used for dependent sources, whereas the LIN and RMS stacking methods are more suitable for independent sources. To extract reliable EGFs and apply an automatic method to minimize the effect of energy flux and stations geometry that may lead to some artifacts on final resultant signals, RMS-R stacking method would be considered as an easy and fast way in the related data processing.

Acknowledgments
T.S. thanks the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), São Paulo, Brazil [grant 2016/20952-4], and S4CE [grant 764810]. All plots were also made using the Generic Mapping Tools (GMT) version 4.2.1 (Wessel and Smith, 1998; www.soest.hawaii.edu/gmt, last accessed May 2019).

References
Prieto, G. A., Lawrence, J. F. and Beroza, G. C., 2009, An elastic Earth structure from the coherency of the ambient seismic
Improvement in the Empirical Green's Function Extraction Using Root …


