Earth, Planets and Space

FULL PAPER

Open Access

Array-size dependency of the upper limit wavelength normalized by array radius for the standard spatial autocorrelation method
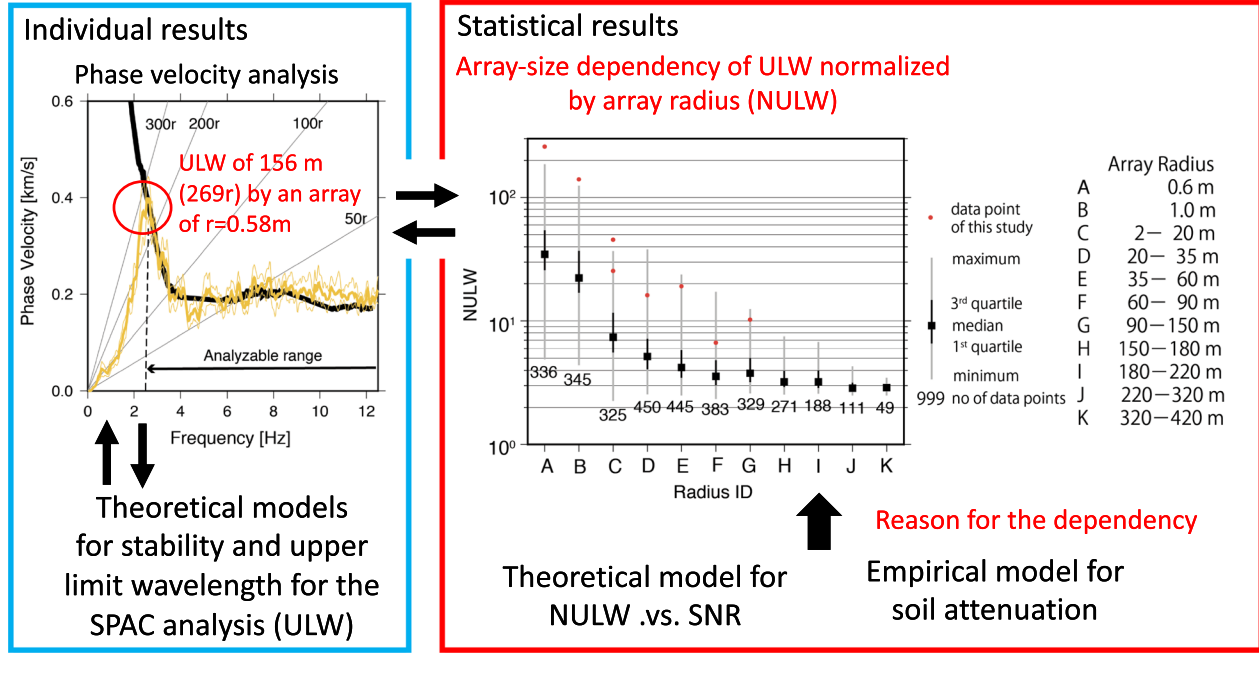
Ikuo Cho[*]

Abstract

This study has shown for the analysis of the standard spatial autocorrelation (SPAC) method that the upper limit wavelength (ULW) normalized by the array radius (normalized ULW, NULW) strongly depends on the array size if we include small (radius r less than a few tens of meters) and very small (r about 1 m or less) microtremor arrays in addition to conventional larger arrays. First, field data of microtremor arrays were analyzed to demonstrate the possible use of small/very small arrays. Specifically, it was shown that, (i) even in the case of a very small array, random errors in the analysis results for very long wavelengths relative to the array radius are kept in an acceptable range for practical use; (ii) the signal-to-noise ratio (SNR) is a crucial factor determining the NULW; and (iii) an equation determining the NULW applies, namely the relation $(\text{NULW}) \propto \sqrt{(\text{SNR})}$ holds through very small to large arrays. The field data used are those distributed for blind prediction (BP) experiments for an international symposium (BP data), which consist of high-quality microtremor array data with various radii from very small ($r = 0.58$ m) to large ($r = 555$ m). It was then shown that SNRs of the BP data, and consequently the NULWs, increase with a decrease of array radius. Statistical data obtained from a few hundred arrays in our previous research also exhibit a similar tendency. The BP data lie around the maximum values of these distributions, showing the high quality of the BP data as well as supporting the array-size dependency of the NULW. Finally, the BP data were processed to identify the characteristics of the soil attenuation. It was found that the array-size dependency of NULW, as well as the large variations in NULW, can generally be explained by soil attenuation. It is plausible that the SNR of small/very small arrays are generally determined by the soil attenuation if the self-noise of the recording system is excluded. A logical conclusion drawn from these results, and also empirically supported, is that the practicality of very small arrays increases as the soil gets softer.

Keywords: Surface geology, Blind prediction, Microtremor, SPAC method, Surface wave, S-wave, SNR, Noise, Phase velocity, Attenuation

*Correspondence: ikuo-chou@aist.go.jp

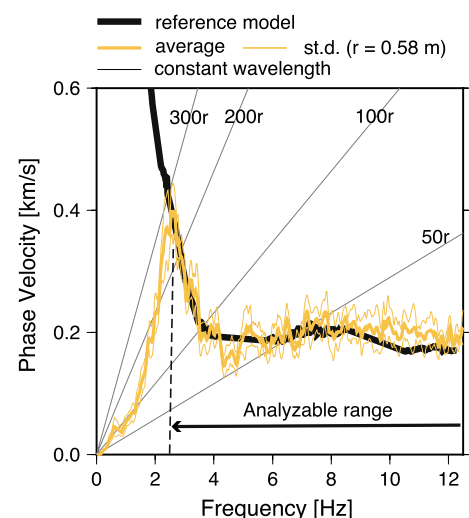Geological Survey of Japan, AIST, Tsukuba Central 7, Tsukuba 305-8567, Japan

Graphical Abstract

Introduction

For the 6th International Symposium on the Effects of Surface Geology on Seismic Motion (ESG6), data for blind prediction (BP) experiments were distributed to the participants (BP data). The BP data include high-quality microtremor array data recorded at Kumamoto city, Japan, with various radii (r) from 0.58 m to 555 m (see "Availability of data and material" Section for further information). After processing the BP data, I obtained an analysis result that appears to be unusual. In my analysis, I applied the standard SPAC method to the array data with $r = 0.58$ m and obtained phase velocities consistent with the reference model (see the later section for the creation of the reference model) to a low frequency of 2.33 Hz at a phase velocity of 364 m/s (Fig. 1). Readers can download the supplemental material containing the analysis results with parameter files and calculation logs (Cho 2022). Appropriate phase velocities were obtained up to a long wavelength of 156 m, or $269r$ (i.e., 269 times the array radius). Also, the phase velocities from an array of $r = 1.15$ m provided appropriate values up to a wavelength of $140r$, though this is not shown in the figure.

Foti et al. (2017) states as a guideline that "there is no clear maximum wavelength criterion" for the SPAC method and that it is "not recommended to try and extract wavelengths greater than 2–3 times the maximum aperture of the passive array" (i.e., their recommendation is NULWs from 4 to 6). This is one reason I described the analysis result shown

Fig. 1 Phase velocities obtained from an array with a radius of $r = 0.58$ m. The thick yellow line represents average values and the thin yellow lines show the standard deviations. A reference model (thick solid line) constructed based on arrays with radii from 5.8 m to 555 m (see the later section) is also plotted for reference. The four straight lines from the origin represent constant wavelengths of $50r$, $100r$, $200r$, and $300r$ (the straight lines from the origin represent constant wavelengths; (phase velocity) = (wavelength) × (frequency)). The lowest limit of the analyzable frequency range (indicated by the arrow) corresponds to a wavelength of $269r$

in Fig. 1 as being "unusual". In the figure, appropriate phase velocities have been obtained up to a wavelength that might be too long relative to the array radius.

Certainly, according to existing field research (e.g., Satoh et al. 2001; Kudo et al. 2002), the above guideline of Foti et al. (2017) is generally valid for arrays with radii larger than a few tens of meters, also as discussed in a later section. Is this still the case, however, even when we pay attention to smaller arrays? Many researchers know, albeit only qualitatively, that small arrays tend to have long analyzable wavelength ranges relative to the array radius.

On the other hand, I acknowledge that a general guideline should be somewhat conservative. Perhaps, we should agree with the cautious attitudes of Asten and Hayashi (2018) for the case when we unexpectedly obtain phase velocity results up to a wavelength that seems too long, if we cannot explain why we have obtained such results and consequently have no tools to diagnose the validity. With such a background, it is likely that many researchers believe that the SPAC method does not apply to such a very small array with a radius of about 1 m or less.

In this paper, it is first shown, based on the BP data, that the standard SPAC method can be applied to such a very small array for the determination of phase velocities at very long wavelengths relative to the array radius. Specifically, it is demonstrated that the analysis results of the very small array are not incidental, but are stable, and can be quantitatively explained based on the theoretical model of Cho and Iwata (2021) (hereafter, CI2021). This demonstration was required to finally discuss the array-size dependency of NULW, including the analysis results of such very small arrays, which is the goal of this study.

The following section outlines the data processing procedure of the SPAC method, so that readers can reproduce the analysis result shown in Fig. 1. The next sections describe the analysis results of the BP data to illustrate the theoretical basis for the stability of the

analysis results, even in such a long wavelength range. Also included is validation of the claim of CI2021 that the upper limit of the analyzable wavelength range is determined by the incoherent noise (i.e., SNR). Finally, it is shown that the NULW strongly depends on the array size when small and very small arrays are included, and that the array-size dependency can be explained by the intrinsic/scattering attenuation.
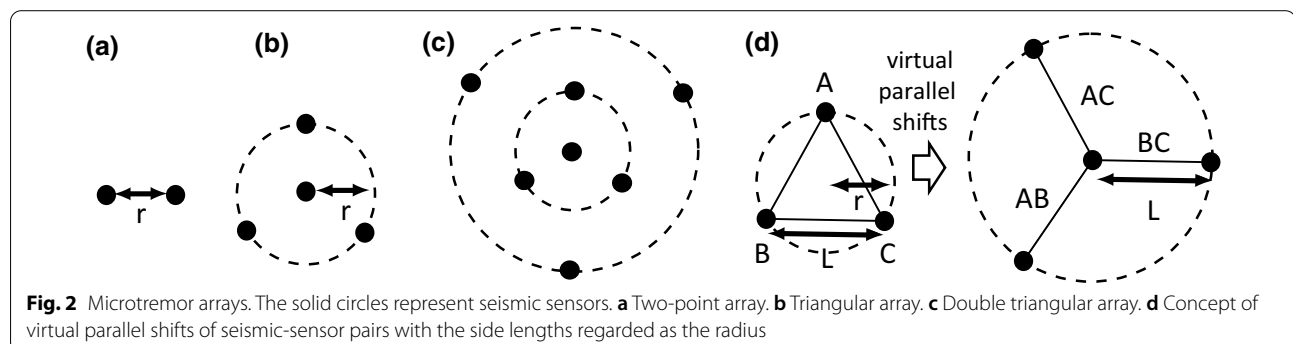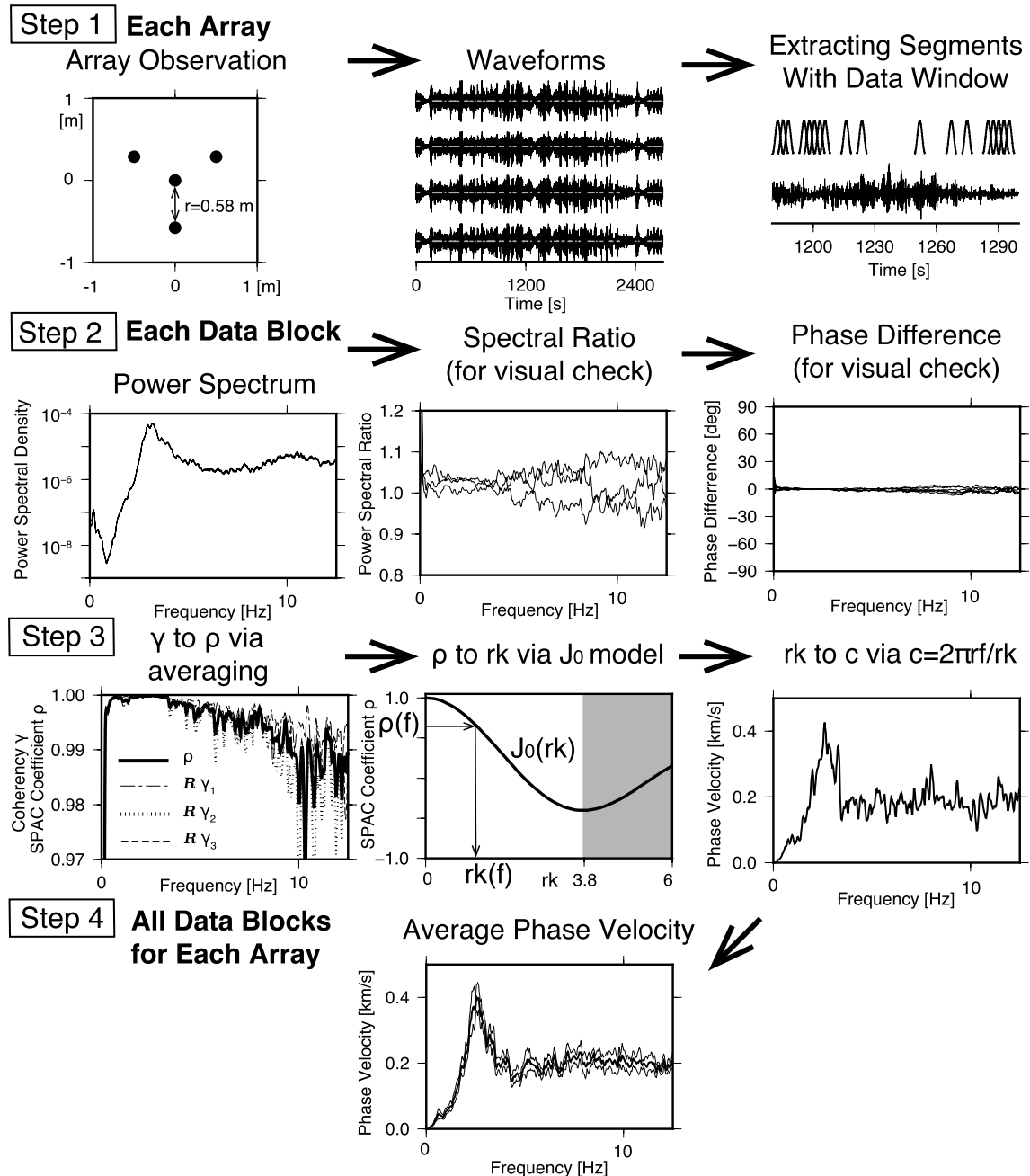
Standard SPAC method adopted in this study
Geometry of microtremor arrays
According to the formulation of the SPAC method (Aki 1957), we can use a two-point array (Fig. 2a) if the wavefield of microtremors is isotropic. However, a circular array is required to deal with a general wavefield. In practice, based on many field application tests (Okada 2003), it was found that a practical configuration for the SPAC method is a triangular array (Fig. 2b) consisting of a seismometer at the center point and three seismometers evenly spaced along the circumference of a circle.

The BP data consist of double triangular arrays (Fig. 2c). I have decomposed these double triangular arrays into two simple triangular arrays (Fig. 2b). Consequently, I have ten different simple triangular arrays with radii from 0.58 m to 555 m ($r = 0.58$, 1.15, 5.77, 11.6, 22.5, 45.0, 70.4, 140, 278, and 555 m).

An extended SPAC method (e.g., Ohori et al. 2002), applicable to an irregular-shaped array, may be applied to the array in Fig. 2c without decomposing it into subarrays. However, such an approach is not suitable for this study because the meaning of "array radius" becomes ambiguous and as a result, the relationship between the upper limit wavelength and the array radius becomes ambiguous. In this study, to strictly examine the relationship between the upper limit wavelength and the array radius, I use simple arrays as shown in Fig. 2b.

Fig. 2 Microtremor arrays. The solid circles represent seismic sensors. **a** Two-point array. **b** Triangular array. **c** Double triangular array. **d** Concept of virtual parallel shifts of seismic-sensor pairs with the side lengths regarded as the radius

Fig. 3 Data processing workflow to obtain a phase velocity dispersion curve from each set of array data. The case of an array radius of 0.58 m of the BP data is shown as an illustration

Spectral analysis and statistics

I applied the standard SPAC method to the vertical-component data according to the procedure shown in Fig. 3. The data processing was conducted using the analysis software BIDO3.0 (Cho 2020).

(Step 1) The waveforms of the vertical-component microtremors were aligned and visually inspected. Data

segments were extracted using a Hanning window with a tapering ratio of 0.5 overlapped by half. The frequency distribution of the root-mean-square value of the amplitudes in a segment was examined before tapering; segments with abnormal values were considered to have nonstationary portions and were eliminated. Note that the data were fully utilized for recording durations of 45

to 120 min. The raw data were used as is (i.e., no instrumental correction, resampling, filtering, time differentiation/integration, etc.).

(Step 2) The remaining segments were used to estimate the spectral densities. The segment durations used for the FFT were set to 20.48 s and the spectral window (Parzen window) width was set to 0.1 Hz. Ten data segments (hereafter, called "data blocks") were collected for an ensemble average. Visual inspections were made of the power spectral ratios and phase differences thus obtained.

(Step 3) Using the spectral densities obtained above, I calculated the (complex) coherencies between the center point and a point on the circle, $\gamma_1(f), \gamma_2(f),$ and $\gamma_3(f)$, where f is frequency, of which the real parts were taken and averaged into the SPAC coefficients $\rho(f)$. The SPAC coefficients were modeled by the 0th order Bessel function of the first kind $J_0(\cdot)$ as

$$\rho(f) = J_0\big(rk(f)\big), \tag{1}$$

where k is the wavenumber (Aki 1957). At each frequency, the value of the normalized wavenumber rk was determined by inverting the model function in the range of rk between 0 and 3.8. The phase velocity $c(f)$ was determined via the relation $c(f) = 2\pi rf/rk(f)$. This process was carried out in each data block.

(Step 4) The numbers of the data blocks ranged from 10 to 39. The phase velocities obtained at each data block were gathered to calculate an average and standard deviation at each frequency. This is the final analysis result of the phase velocities.

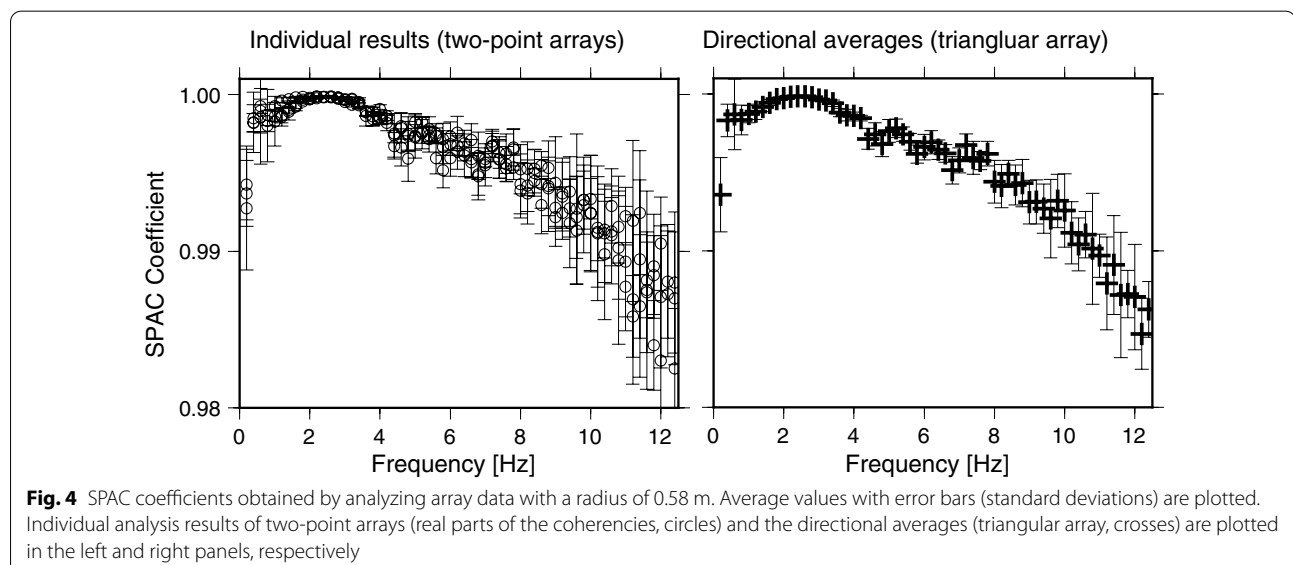## Stability of analysis results in long wavelength ranges

CI2021 claims that random error in the analysis results is not a factor in creating an analysis limit at long wavelengths. In the following, I examine the applicability of this claim to the BP data with $r = 0.58$ m by following the same procedure as CI2021. Readers are referred to this paper for the details.
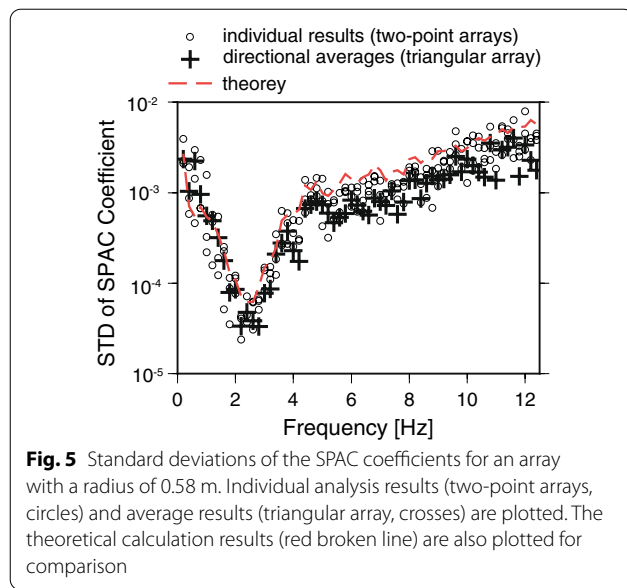
Coherency and the SPAC coefficient

An array with a radius of less than 1 m is like a huddle test for checking instrumental specifications; the coherencies should take values of almost 1. Indeed, the SPAC coefficients of both two-point arrays (i.e., the real parts of the coherencies between the center point and a vertex of the triangle) (Fig. 2a) and a triangular array (Fig. 2b) are greater than 0.99 in the frequency range lower than 10 Hz (Fig. 4). Many researchers may believe small variations in the SPAC coefficients within such a small range from 1 as random error not having a physical meaning. However, examining Fig. 4 carefully, we find that the error bars become smaller as the SPAC coefficients approach 1. This feature is not incidental, it is a universal trend predictable by spectral analysis theory.

In Fig. 5, the variability of the observed data (i.e., the random errors in the SPAC coefficients of two-point arrays) (circles) is explained by the theoretical values (red broken line), represented as

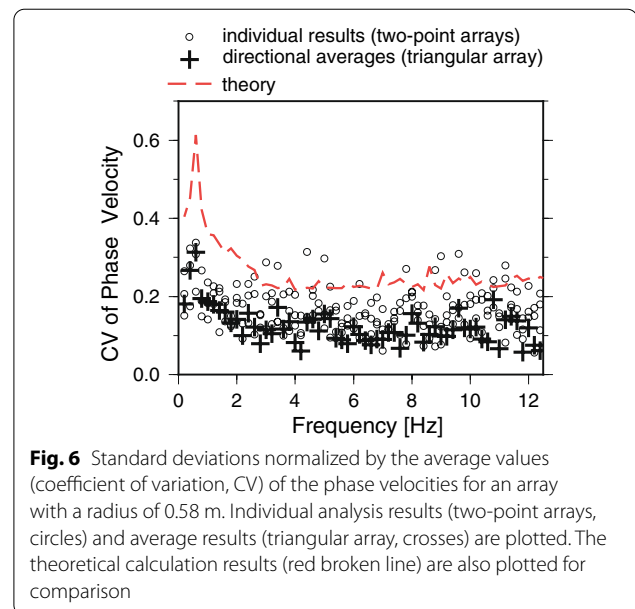$$s.d.\big\{\widehat{\rho}(f)\big\} \approx \frac{1 - \rho^2(f)}{\sqrt{2n_{\mathrm{d}}}}, \tag{2}$$

Fig. 4 SPAC coefficients obtained by analyzing array data with a radius of 0.58 m. Average values with error bars (standard deviations) are plotted. Individual analysis results of two-point arrays (real parts of the coherencies, circles) and the directional averages (triangular array, crosses) are plotted in the left and right panels, respectively

Fig. 5 Standard deviations of the SPAC coefficients for an array with a radius of 0.58 m. Individual analysis results (two-point arrays, circles) and average results (triangular array, crosses) are plotted. The theoretical calculation results (red broken line) are also plotted for comparison

Fig. 6 Standard deviations normalized by the average values (coefficient of variation, CV) of the phase velocities for an array with a radius of 0.58 m. Individual analysis results (two-point arrays, circles) and average results (triangular array, crosses) are plotted. The theoretical calculation results (red broken line) are also plotted for comparison

where a hat above a symbol indicates that it refers to an estimate. The parameter n_d corresponds to the number of data segments for the FFT extracted with a non-overwrapping boxcar window (i.e., half the degrees of freedom for spectral estimation). It means that the magnitude of the random error in the real part of the coherency, and consequently in the SPAC coefficient and phase velocity, depends on the number of independent samples (n_d) used in the spectral analysis. In this study, the parameter n_d is about 11 for all arrays.

Note that we can decrease the magnitude of random error to less than those shown in Figs. 4 and 5 by increasing the value of n_d. This can be done either by increasing the number of segments for the ensemble average or by increasing the width of the spectral window or both. (Readers can refer to Appendix D of CI2021 for the calculation of n_d in practical data processing.) Thus, physically meaningful information can be extracted from the SPAC coefficients, not affected by random errors, even at long wavelengths exceeding a hundred times the array radius (i.e., SPAC coefficients very close to 1).

Incidentally, since the SPAC coefficients of a triangular array are the average values of two-point arrays, their random errors should be smaller than those of the individual values of the real parts of the coherencies. The results in Fig. 5 (crosses) show that this is indeed the case. This can be seen directly in Fig. 4: the error bars in the right panel are much smaller than those in the left panel. It should also be noted, on the other hand, that the differences between the average values of the left and right panels in Fig. 4 are

slight, demonstrating the significance of the averaging process.

Phase velocity

As shown in Fig. 6, according to the error propagation theory, the magnified random errors of the SPAC coefficients in estimating phase velocities are kept within a practically acceptable range. According to the figure, the theoretical values of the random errors in phase velocity estimates give upper limits for the variations in the analysis results of observed data. The theoretical values are calculated using the following equation:

$$\frac{s.d.\{\hat{c}\}}{c} \approx \frac{1+\varepsilon}{\sqrt{2n_d}rk|J_1(rk)|}\left[1 - \left(\frac{1}{1+\varepsilon}J_0(rk)\right)^2\right],$$

(3)

where $J_1(\cdot)$ is the first-order Bessel function of the first kind and ε is the frequency-dependent noise-to-signal ratio (NSR) (reciprocal of SNR).

Installation positions of seismometers

One might consider that a 1% error of the installation location of a seismometer would have a devastating impact on the analysis results of phase velocities because all of the information on phase velocity is packed between 0.99 and 1 of the coherency (Fig. 4). However, this is not a concern because according to error propagation theory, the relative error of the phase velocity due to errors in the seismometer location will have only the same magnitude as the relative errors of the array radius. For example, if an array radius is set to 1.01 m mistakenly when it is installed for an array with a radius of 1 m, the

phase velocity will only be underestimated by 1%. Practically, this will not be a serious problem.

Crucial factor determining the upper limit of analyzable wavelength ranges

According to CI2021, incoherent noise determines the analysis limit of the SPAC method in the long wavelength range. This means that the fact that we can obtain phase velocities up to very long wavelengths relative to the array radius from the array of $r=0.58$ m in the BP data indicates that this array data includes very low powers of incoherent noise. In this section, I investigate the validity of this idea by following the same procedure as CI2021. Readers are referred to this paper for the details.

Construction of a reference model of the phase velocity dispersion curve

In the standard SPAC method, we estimate the SPAC coefficient $\rho(f)$ based on array data, then generally determine the value of $rk(f)$ through a noise-free model of $\rho(f) = J_0(rk(f))$ (Fig. 7a). In practice, however, incoherent noise inevitably exists and the magnitude of the SPAC coefficient is underestimated following the relation

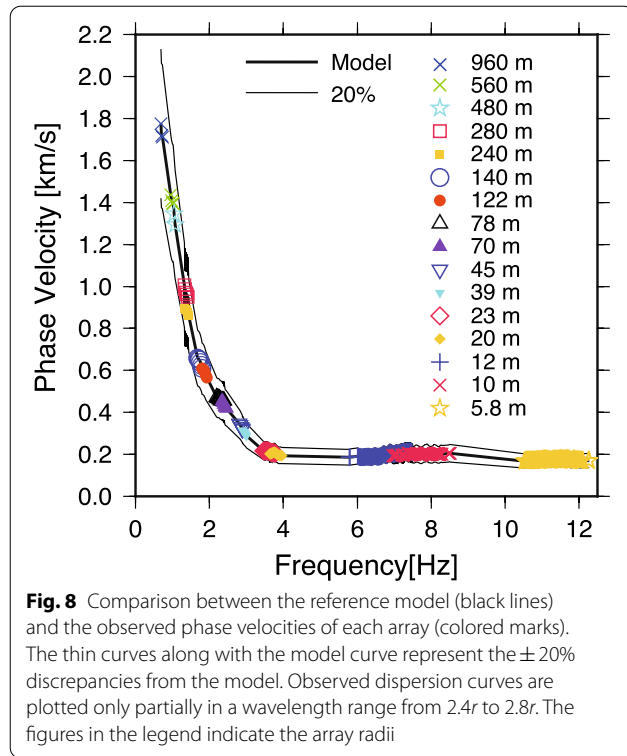$$\rho(f) = \frac{1}{1 + \varepsilon(f)} J_0(rk(f)). \tag{4}$$

If we adopt a noise-free model with incoherent noise present, we will overestimate the value of $rk(f)$ (Fig. 7b). Consequently, through the relation $c(f) = 2\pi rf/rk(f)$, we will underestimate the value of the phase velocity.

At the zero-crossing point of the Bessel function (the square in Fig. 7b), where the wavelength corresponds to 2.6r, the incoherent noise does not affect the phase velocity estimate since the amplitude is originally zero. This means that we can more accurately estimate the phase velocities if we estimate them at only the zero-crossing points for various sizes of microtremor arrays. Also, using only the first zero-crossing point gives the additional benefit that the first zero-crossing point is free from the effects of the wavefield (i.e., free from directional aliasing errors) (e.g., Cho et al. 2021).

Based on the above benefits of the first zero-crossing points of the SPAC coefficients, I analyzed arrays of each size included in the BP data and extracted only the part corresponding to 2.4–2.8r from the obtained phase velocity dispersion curves, which are then combined into a single phase velocity dispersion curve. Figure 8 overplots this single dispersion curve, which we call the "reference model" in Fig. 1, and the constituent data (i.e., phase velocities obtained from each array). A label is assigned to each phase velocity data indicating the "radius" of the microtremor array, which represents either the radius in an ordinary sense or the side length of a triangle array (i.e., r and L in Fig. 2d). For example, the label "960 m" corresponds to the side length L of a triangle array. We can see from the comparison between Figs. 1 and 8 that the validity of the phase velocities of the array with $r=0.58$ m is supported by the phase velocities of 2.4–2.8r obtained from the arrays with radii up to 70 m.

Fig. 7 **a** Schematic diagram of determining the value of *rk* from the SPAC coefficient through the model function. The grey curve represents the model function of $J_0(rk)$. **b** SPAC coefficient curve affected by incoherent noise (broken line). *rk* is overestimated (i.e., *rk′* in the figure) in the *rk* range less than 2.6 when incoherent noise is present. The shaded regions in each diagram represent the *rk* range larger than 3.8 (i.e., out of the search range for *rk*)

Fig. 8 Comparison between the reference model (black lines) and the observed phase velocities of each array (colored marks). The thin curves along with the model curve represent the ±20% discrepancies from the model. Observed dispersion curves are plotted only partially in a wavelength range from 2.4r to 2.8r. The figures in the legend indicate the array radii

Relation between NULW and NSR

First, I determine the dispersion curves of the Rayleigh wave phase velocities for each array of the BP data. Then, I define the upper limit frequency (ULF) and the ULW as the frequency and wavelength at which the divergence of the observed phase velocity dispersion curve from the reference model reaches ±20%. Next, the NSRs ε are estimated for each array using the following equation:

$$\varepsilon = \frac{N[(\rho_{CCA} + 2)(1 - \rho) - 1]}{N(\rho_{CCA} + 2)\rho - \rho_{CCA} + 1},\tag{5}$$

where N is the number of sensors along the circumference of a circle (i.e., $N = 3$) and ρ_{CCA} is the spectral ratio used in the centerless circular array method of Cho et al. (2004). Then, the value of NSR at the ULF is read.
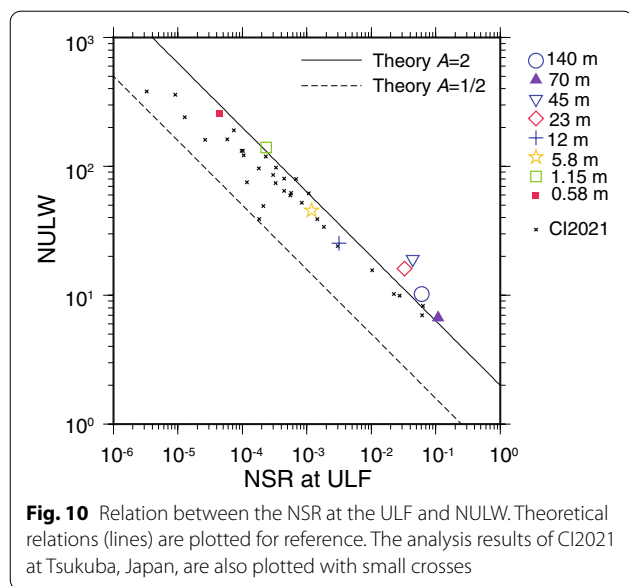
The dispersion curves for arrays with $r = 0.58$, 11.6, and 140 m are shown in Fig. 9 as an illustration. The ULFs of these dispersion curves are 2.33, 1.68, and 0.83 Hz, respectively (circles in the left panel). The corresponding values of the ULWs are 156, 294, and 1430 m ($269r$, $25r$, and $10r$), respectively, obtained by dividing the corresponding phase velocities by the ULFs. By reading the corresponding values of the NSRs at those ULFs from the right panel, we can obtain the sets of values (NSR, ULW) = $(3.7 \times 10^{-5}, 269r)$, $(3.2 \times 10^{-3}, 25r)$, and $(6.1 \times 10^{-2}, 10r)$.

Figure 10 shows the relationship between the NSR and NULW thus obtained for the eight different arrays ($r \leq 140$m) for which the analysis limits do not exceed the ends of the reference model. In the figure the relation obtained by CI2021 is also shown, which is expressed as:

$$(\text{NULW}) = A\varepsilon^{-\frac{1}{2}}.\tag{6}$$

Fig. 9 Readings (circles) of the phase velocities (left) and NSRs (right) at the upper limits of the analyzable frequency range. Three cases of array radius of 0.58 m (yellow), 11.6 m (purple), and 140 m (sky blue) are shown as an example. In the left panel, the reference model (black line) and the ±20% levels (thin black lines) are also plotted

Fig. 10 Relation between the NSR at the ULF and NULW. Theoretical relations (lines) are plotted for reference. The analysis results of CI2021 at Tsukuba, Japan, are also plotted with small crosses

This is a theoretical model based on the error propagation theory. The parameter A is a constant given by $A = \sqrt{2b\pi}$, where b represents the threshold of the divergence of the observed phase velocity dispersion curve from the reference model (Eq. (15) in CI2021; a threshold value is needed to quantitatively define NULW). The parameter b is set to 0.2 in this study, as well as in CI2021, and thus A is calculated to be 2. CI2021, however, empirically treated A to have a range from 1/2 to 2, as it was necessary to take into account fluctuations in the phase velocities (e.g., ripples appearing in the dispersion curve) and variations caused by modeling errors in the velocity structure. Figure 10 shows that the analysis results of the BP data are generally distributed in accordance with this

equation. This means that the formation of the NULW for the BP data can be explained by the effect of incoherent noise. The trend of this line is well represented by the theoretical value of $A = 2$, suggesting that the BP data were obtained under very favorable geoenvironmental conditions.

Figure 10 also plots the data points obtained in CI2021 (i.e., the case of Tsukuba city) for reference. Readers can refer to Appendix E of Cho et al. (2021) for the plots of 665 data points as a further validation of the relationship between NSR and NULW. The point that should be emphasized in Fig. 10 is that the "unusual" analysis result shown in Fig. 1 is not so special. This means that the red square in Fig. 10 (i.e., Fig. 1) is just one of a series of data points aligned in a straight line. We can conclude that the analysis result in Fig. 1 was obtained because, simply enough, the intensities of the incoherent noise were very small relative to the intensities of the signal (microtremors).

Impacts of array size on the NSR and NULW

According to either Fig. 10 or Eq. (6), the NSR should be kept smaller than 10^{-4} for the NULW to exceed 100. A recording system with sufficiently low self-noise is required to realize such values of NSR. Also, in my experience, very small arrays ($r \leq 1$ m) are often able to provide very small NSRs, which are followed by the NULW exceeding a few tens of times the array radius.

Figure 11 shows the dependency of NULW on array size, taken from a statistical study based on a few hundred microtremor arrays deployed within the Kanto Plain, Japan (Cho et al. 2021). The analysis results of this study, taken from Fig. 10, are overplotted in the figure (red circles). In the previous results in Fig. 11, the median

Fig. 11 Relationship between array radius and NULW. Modified from Fig. 11a of Cho et al. (2021)

value of the NULW for the microtremor arrays with a radius of 0.6 m is 35. This means that half of the microtremor arrays with a radius of 0.6 m provide phase velocities with relative wavelengths longer than 35. As well, the median values of NULWs for very small to small arrays of $r = 1.0, 2–20$, and $20–35$ m are 22, 7, and 5, respectively. The NULW decreases with an increase in the radius.

On the other hand, the median values of the NULW stay within 5 for larger arrays of $r = 35–420$ m in Fig. 11 (i.e., radius IDs E–K in Fig. 11). These values are consistent with the guideline of Foti et al. (2017). It is considered that for large arrays the NSR increases too much with the increase in array size and so we can obtain an appropriate estimate of phase velocity only around the zero-crossing point (i.e., wavelength of $2.6r$) of the SPAC coefficient curve, as described in the above section.

Note that for the radius IDs J and K in Fig. 11 (i.e., the largest arrays of r from 220 to 420 m), the NULWs are almost $3r$ for most of the data used in the statistics. This does not contradict the guidelines of Foti et al. (2017), which recommend not trying to extract wavelengths greater than $4–6r$. However, prior experience suggests that it is too small (e.g., Satoh et al. 2001; Kudo et al. 2002). Presumably, there are strong influences of either microtremors of low intensity relative to the self-noise level, or incoherent noise generated by general oscillation sources in the vicinity (i.e., industrial activity), or both. This may be related to the fact that all the data used in the previous study in Fig. 11 were measured with accelerometers during the daytime (Cho et al. 2021).

All data points obtained in this study lie around the maximum of the previous study, showing the high quality of the BP data as well as supporting the existence of the array-size dependency. The reason for the high quality of the BP data for small/very small arrays is considered in the following section. Concerning large arrays (the radius IDs E–G in Fig. 11), the high quality of the BP data may partly be due to the fact that the observations were conducted in the nighttime (i.e., 2:00 to 6:00).

By applying a rule of thumb for the penetration depth, given as (penetration depth) = (ULW)/3 (p. 93 of SEGJ 2008), to the statistical data shown in Fig. 11, we can obtain a rough estimate of the penetration depth for each radius. If we regard the statistical results for the radius IDs of A, B, C, G, and K as those for the representative radii of 0.6, 1, 10, 100, and 400 m, the calculation results are as listed in Table 1, where the depth range corresponds to an interval between the first and third quartiles (i.e., an interval containing half of the data). In the case of microtremor array surveys in the Kanto Plain, Japan, the penetration depths for radii (i.e., the distance between seismometers) of 0.6, 1, 10, 100, and 400 m are about $9–18r$, $6–12r$, $2–4r$, $1–2r$, and $0.9–1r$, respectively. This

Table 1 Relation between array radius and penetration depth

Radius, m	Penetration depth, m
0.6	5–11
1	6–12
10	19–39
100	110–170
400	360–410

Fig. 12 Relation between array radius and NSR at ULF. The data of this study (Fig. 10) are plotted

may be a guideline for SPAC users. Note that the penetration depth for the array with $r = 0.58$ m of the BP data is 52 m (i.e., $90r$) based on this calculation, considerably higher than the representative ranges shown in Table 1, indicating the very good quality of the BP data.

Discussion

What causes the array-size dependency for small/very small arrays

As described in the above section, the array size of a recording system is an important factor affecting the NSR, and consequently the NULW. Then, one may consider why the array size affects the NSR. To describe this, we first examine the trend of NSRs of the BP data as a function of the distance (i.e., array radius) (Fig. 12). In the figure, the NSR increases with distance even at a small distance of about 1 m. If the only source of incoherent noise is the self-noise of the recording system, the NSRs are unlikely to be changed by a difference in the distances on the order of 1 m or less. Furthermore, the observed changes are systematic: as the distance gets smaller, the NSR becomes smaller. It is natural to consider intrinsic/

scattering attenuation of soil as a universal factor increasing NSRs on such a distance scale.

I converted the NSRs obtained from the BP data to surface wave attenuation coefficients through the same procedure taken in CI2021. Attention was paid to the shorter side of the analysis limits (Appendix C of CI2021). I finally obtained the frequency characteristics of the attenuation coefficients (Fig. 13), which are similar to those obtained from the Tsukuba data in CI2021. As shown in the figure, this analysis result has frequency characteristics similar to those obtained by surface wave surveys at high frequency and by seismic-wave analysis at low frequency at various sites around the world (Mitchel 1973; Foti 2000; Lai et al. 2002; Prieto et al. 2009; Badsar et al. 2010; Xia et al. 2012; Lu 2015; Gao et al. 2018; Mun and Zeng 2018), supporting the validity of the analysis results of this study.

The attenuation coefficient shown in Fig. 13 is for Rayleigh waves propagating along the surface of a layered medium and thus depends on the overall soil structure. To specify the soil attenuation of each stratum (i.e., Q factor; "Q" in Aki and Richards 2002) at each observation site, we should consider the small differences between the frequency characteristics in the figure. For example, Prieto et al. (2009) obtained frequency-dependent attenuation coefficients from microtremor analyses and then determined depth-dependent Q factors based on the relation between the attenuation coefficient and the Q factor, as per Eq. (6) in Mitchel (1995). At the current stage of the study, however, I only pay attention to the overall similarity between the frequency characteristics

Fig. 13 Frequency characteristics of the attenuation coefficients. The analysis results of this study are indicated by red circles with error bars (standard errors). The solid black line indicates the analysis result of CI2021 at Tsukuba, Japan. The gray lines and marks indicate the analysis results obtained by other researchers (Fig. 12 of CI2021)

of the attenuation coefficients in Fig. 13 for the following discussion in this paper. Further specification of the soil attenuation structure at each observation site is left for future study.

Overall, Fig. 13 shows that the attenuation coefficients increase with increasing frequency. The strong array-size dependency of NULW for small/very small arrays shown in Fig. 11, as well as the large variations in NULW, can be explained by this simple feature, as described in the following. Let us assume for simplicity that $A = 1$ in Eq. (6) and the recording system is ideal (i.e., no self-noise). By substituting the relation (Prieto et al. 2009) between NSR $\varepsilon(f)$ and the attenuation coefficient $\alpha(f)$,

$$\varepsilon(f) = \alpha(f)r, \tag{7}$$

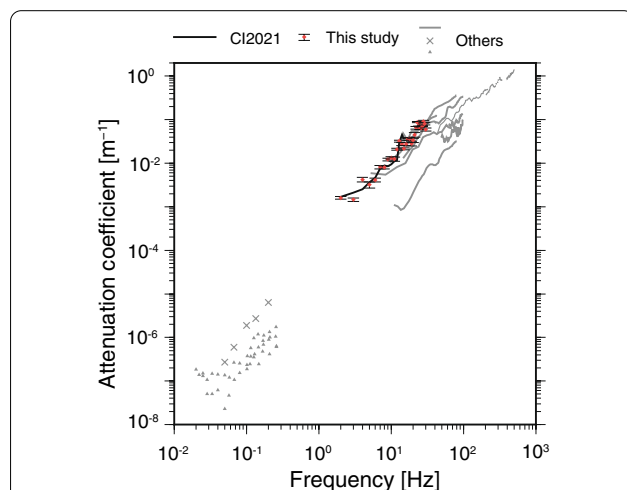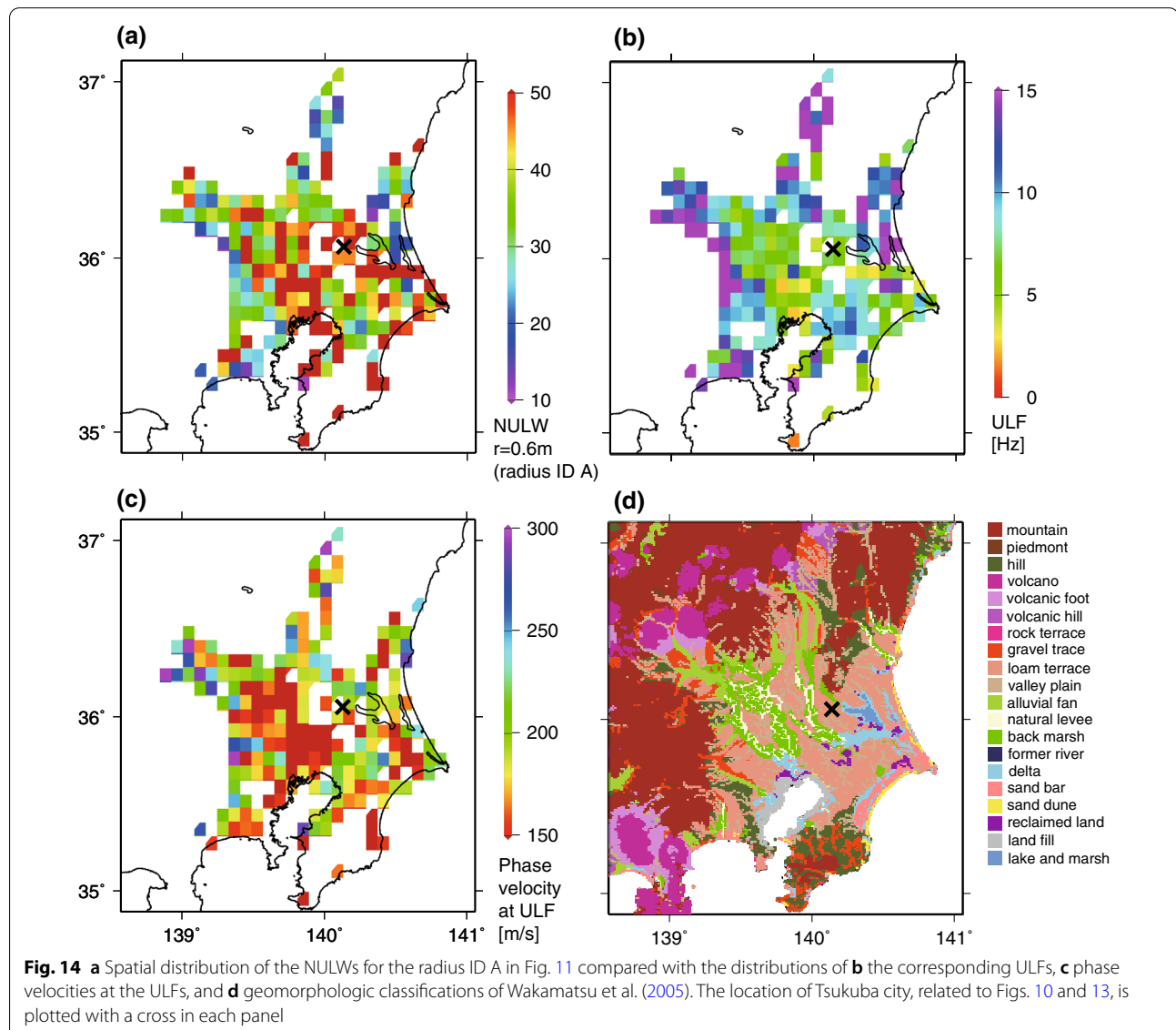into Eq. (6), we obtain the following relation:

$$(\text{NULW}) = [\alpha(f)r]^{-\frac{1}{2}}. \tag{8}$$

Note that CI2021 suggests the use of another equation instead of Eq. (7) because Eq. (7) is just an approximation. However, the simple equation of Eq. (7) is adequate at this point for rough estimates.

According to Eq. (8), the NULWs for $r = 10$ and 25 m, for example, are 32% and 20% of the NULW for $r = 1$ m when the ULFs are the same irrespective of array size, and therefore the attenuation coefficients have the same values. In Fig. 11, the median values of B, C, and D are 22, 7, and 5, respectively, of which values C and D divided by B are 32% and 23%, quite consistent with the above calculation results using Eq. (8).

Let us further assume that the attenuation coefficient can be modeled as $\alpha(f) = f^2/10^4$ from a visual inspection in Fig. 13 and that the ULFs depend on the surface geology for a microtremor array with a fixed size. Indeed, the ULF will depend on the velocity structure at a target site because the frequency with a certain wavelength is high at a hard-soil site (i.e., a site with either very thin surface soft sediments or no soft sediments) and low at a soft-soil site. If we set the ULFs to be 1, 3, 10, and 30 Hz for $r = 0.6$ m, for example, the corresponding values of NULWs are calculated to be 130, 43, 13, and 4, respectively, based on Eq. (8). It is therefore expected that when the soil is hard such that the ULF becomes high exceeding 10 Hz, the NULW takes a small value (e.g., less than 10) affected by high NSRs due to the soil attenuation. On the other hand, when the soil is soft so that the ULF becomes low close to a few hertz, the NULW takes a large value (e.g., more than 100). These calculation results can explain why the NULW takes such a large variation, as exhibited in the statistical result for radius ID A in Fig. 11.

Thus, the strong array-size dependency of NULW characterized by small/very small arrays, as well as the large variations in NULW, can be generally explained by soil attenuation. As the logical consequence of this idea, we can expect that a very small array can be more effectively used at a soft-soil site than at a hard-soil site. Indeed, I obtained the spatial distributions of NULWs and ULFs using the original data for radius ID A in Fig. 11 to demonstrate that the high and low values of NULWs (panel (a)) correspond to the low and high values of ULFs (panel (b)), respectively (Fig. 14). Also, the high and low values of NULWs correspond to the sites expected to be soft-soil (e.g., back marshes, delta, land fill, natural levee, and reclaimed lands) and hard-soil (e.g., alluvial fans, gravelly terrace, and hills) sites, respectively (panel (d)). In

panel (c) in Fig. 14, the phase velocities at the ULFs take reasonable values that are consistent with the geomorphological classifications, supporting the validity of the analysis results.

Effects of incoherent noises generated by other than soil attenuation

The self-noise of the recording system is excluded in the above discussion, but is a factor in the real system. If the frequency characteristics of the self-noise are flat or exhibit a negative trend with frequency (i.e., 1/f noise), considering the balance with the frequency characteristics of the attenuation coefficients, we can generally expect that soil attenuation is dominant in the high frequency range, whereas self-noise is dominant in the low

Fig. 14 **a** Spatial distribution of the NULWs for the radius ID A in Fig. 11 compared with the distributions of **b** the corresponding ULFs, **c** phase velocities at the ULFs, and **d** geomorphologic classifications of Wakamatsu et al. (2005). The location of Tsukuba city, related to Figs. 10 and 13, is plotted with a cross in each panel

frequency range (e.g., CI2021). Thus, in the context of the above discussion, self-noise may limit large values of NULW at soft-soil sites. From this perspective, the array-size dependence and large variations in NULW seen in small/very small arrays are possibly influenced not only by soil attenuation, but also the self-noise of the recording system.

It should be emphasized, however, this does not imply that impractically expensive equipment is required to obtain large NULWs at soft-soil sites. Usual equipment for microtremor array surveys can be applied to very small arrays for the standard SPAC method. This is because the argument presented above regarding the stability of the analysis results at long wavelengths (e.g., Eq. (2)) is independent of the equipment. For example, CI2021 demonstrated that usual geophones can provide high values of NULW ranging from about 60 to 100 (i.e., Array no. 31 and 32 in Tables 1 and 3 of CI2021).

As described in "Impacts of array size…" Section (i.e., the discussion for the radius IDs E–K in Fig. 11), for arrays larger than a few tens of meters, incoherent noise generation due to general oscillation sources in nearby areas could additionally affect the NULW. Therefore, in summary, it is plausible that for small/very small arrays, the high quality (i.e., very large NULW) of the BP data is due to the use of seismometers with low self-noise at a not-so-hard soil site, while the high quality for middle-to-large arrays is also due to the low intensity of incoherent noise from general oscillation sources in the vicinity. With regard to the three kinds of incoherent noises, i.e., soil attenuation, self-noise of the recording system, and general oscillation sources in the vicinity, the readers are referred to CI2021 for deeper considerations.

Summary

The following claims of our recent paper (Cho and Iwata 2021) were validated with specific examples based on the data for the blind prediction experiments provided at the 6th International Symposium on the Effects of Surface Geology on Seismic Motion (BP data). The claims are that (i) random errors in the SPAC coefficients and phase velocities are generally kept small even for very long wavelengths relative to the array radius; (ii) consequently, the signal-to-noise ratio (SNR) becomes a crucial factor determining the upper limit wavelength normalized by array radius (NULW) for the analysis of the SPAC method; (iii) the relation $(\mathrm{NULW}) \propto \sqrt{(\mathrm{SNR})}$ holds. The key point emphasized in this study is that the above claims are established irrespective of array size.

Based on the statistical data obtained in our previous study by Cho et al. (2021), together with the BP data, it was shown that the NULW strongly depends on the array-size if we take into account either small (radius r less than a few tens of meters) or very small (r about 1 m or less) arrays or both, the NULWs of which are dramatically larger than those of both middle (r about 100 m) and large (r about 500 m or more) arrays. The existing guideline is valid for arrays with radii larger than several tens of meters.

Based on the statistical results for microtremor array surveys in the Kanto Plain, Japan, rough estimates of the penetration depths for radii of 0.6, 1, 10, 100, and 400 m are about $9-18r$, $6-12r$, $2-4r$, $1-2r$, and $0.9-1r$ (i.e., $5-11$, $6-12$, $19-39$, $110-170$, and $360-410$ m), respectively, where the intervals correspond to the first and third quartiles. This can be helpful for planning a survey.

Finally, it was demonstrated that the empirical array-size dependency of NULWs for small/very small arrays, as well as the large variations appearing in NULWs, can be explained by the effects of soil attenuation, although it is considered that the self-noise of a recording system also plays an important role. As a logical consequence, also shown empirically, the practicality of very small arrays increases as the soil gets softer (i.e., the S-wave velocity gets lower).

Declarations

References
Aki K (1957) Space and time spectra of stationary stochastic waves, with special reference to microtremors. Bull Earthquake Res Inst Univ Tokyo 35:415–456

Aki K, Richards PG (2002) Quantitative seismology, 2nd edn. Univ. Science Books, Sausalito, p 700

Asten MW, Hayashi K (2018) Application of the spatial auto-correlation method for shear-wave velocity studies using ambient noise. Surv Geophys 39:633–659

Badsar SA, Schevenels M, Haegeman W, Degrande G (2010) Determination of the material damping ratio in the soil from SASW tests using the half-power bandwidth method. Geophys J Int 182:1493–1508. https://doi.org/10.1111/j.1365-246X.2010.04690.x

Cho I (2020) BIDO. https://staff.aist.go.jp/ikuo-chou/bidodl_en.html. Updated on November 19, 2020

Cho I (2022) Archive of the analysis results of the microtremor data obtained from a seismic array with a radius of 0.58 m distributed to the participants of the blind prediction experiments for the ESG6 symposium. Available via ZENODO repository. Uploaded on April 27, 2022. https://doi.org/10.5281/zenodo.6496789

Cho I, Iwata T (2021) Limits and benefits of the spatial autocorrelation microtremor array method due to the incoherent noise, with special reference to the analysis of long wavelength ranges. J Geophys Res Solid Earth 126:e2020JB019850. https://doi.org/10.1029/2020JB019850

Cho I, Tada T, Shinozaki Y (2004) A new method to determine phase velocities of Rayleigh waves from microseisms. Geophysics 69:1535–1551. https://doi.org/10.1190/1.1836827

Cho I, Senna S, Wakai A, Jin K, Fujiwara H (2021) Basic performance of a spatial autocorrelation method for determining phase velocities of Rayleigh waves from microtremors, with special reference to the zero-crossing method for quick surveys with mobile seismic arrays. Geophys J Int 226:1676–1694. https://doi.org/10.1093/gji/ggab149

Foti S (2000) Multistation methods for geotechnical characterization using surface waves. Ph.D. Dissertation, Politecnico di Torino (Italy)

Foti S, Hollender F, Garofalo F, Albarello D, Asten M, Bard PY, Comina C, Cornou C, Cox B, Giulio DG, Forbriger T, Hayashi K, Lunedei E, Martin A, Mercerat D, Ohrnberger M, Poggi V, Renalier F, Sicilia D, Socco V (2017) Guidelines for the good practice of surface wave analysis: a product of the InterPACIFIC project. Bull Earthquake Eng 16:2367–2420. https://doi.org/10.1007/s10518-017-0206-7

Gao L, Pan Y, Gang T, Xia J (2018) Estimating Q factor from multi-mode shallow-seismic surface waves. Pure Appl Geophys 175:2609–2622. https://doi.org/10.1007/s00024-018-1828-7

Kudo K, Kanno T, Okada H, Özel O, Erdik M, Sasatani T, Higashi S, Takahashi M, Yoshida K (2002) Site-specific issues for strong ground motions during the Kocaeli, Turkey, earthquake of 17 August 1999, as inferred from array observations of microtremors and aftershocks. Bull Seism Soc Am 92:448–465. https://doi.org/10.1785/0120000812

Lai CG, Rix GJ, Foti S, Roma V (2002) Simultaneous measurement and inversion of surface wave dispersion and attenuation curves. Soil Dyn Earthq Eng 22:923–930. https://doi.org/10.1016/S0267-7261(02)00116-1

Lu Z (2015) Self-adaptive method for high frequency seismic surface wave method. J Appl Geophys 121:128–139. https://doi.org/10.1016/j.jappgeo.2015.08.003

Mitchel BJ (1973) Surface-wave attenuation and crustal anelasticity in central North America. Bull Seism Soc Am 63:1057–1071

Mitchel BJ (1995) Anelastic structure and evolution. Rev Geophys 33:441–462

Mun SC, Zeng SS (2018) Estimation of Rayleigh wave modal attenuation from near-field seismic data using sparse signal reconstructions. Soil Dyn Earthq Eng 107:1–8. https://doi.org/10.1016/j.soildyn.2018.01.002

Ohori M, Nobata A, Wakamatsu K (2002) A comparison of ESAC and FK methods of estimating phase velocity using arbitrarily shaped microtremor arrays. Bull Seismol Soc Am 92:2323–2332. https://doi.org/10.1785/0119980109

Okada H (2003) The microtremor survey method (Geophysical monographs series no. 12). Society of Exploration Geophysicists, Tulsa, OK. p 135

Prieto GA, Lawrence JF, Beroza GC (2009) Anelastic earth structure from the coherency of the ambient seismic field. J Geophys Res 114:B07303. https://doi.org/10.1029/2008JB006067

Satoh T, Kawase H, Iwata T, Higashi S, Sato T, Irikura K, Huang HC (2001) S-wave velocity structure of the Taichung Basin, Taiwan, estimated from array and single-station records of microtremors. Bull Seism Soc Ame 91:1267–1282. https://doi.org/10.1785/0120000706

Society of Exploration Geophysicists of Japan (SEGJ) (2008) Guide to the application of geophysical surveys, new edition: civil engineering Geophysical survey manual 2008, SEGJ, p 540 [in Japanese]

Wakamatsu K, Kubo S, Matsuoka M, Hasegawa K, Sugiura M (2005) Japan engineering geomorphologic classification map (product serial number: JEGM0936), University of Tokyo Press [in Japanese with English manual]

Wessel P, Smith WHF (1998) New, improved version of Generic Mapping Tools released. Eos Trans 79: p. 579. American Geophysical Union. https://doi.org/10.1029/98EO00426

Xia J, Xu Y, Miller RD, Ivanov J (2012) Estimation of near-surface quality factors by constrained inversion of Rayleigh-wave attenuation coefficients. J Appl Geophys 82:137–144. https://doi.org/10.1016/j.jappgeo.2012.03.003

Publisher's Note