An Automated Segment Selection Algorithm Based on the Assessment of Surface Wave Characteristic in Ambient Noise Recordings

Ya Liu¹, Jianghai Xia¹, Tianyu Dai², Chaoqiang Xi¹, Ling Ning¹, Huaixue Xing³, Xiaojun Chang³

¹ Key Laboratory of Geoscience Big Data and Deep Resource of Zhejiang Province, School of Earth Sciences, Zhejiang University, Hangzhou, China
² Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China
³ Nanjing Center of Geological Survey, China Geological Survey, Nanjing, China
E-mail: liuya@zju.edu.cn

Abstract. Passive surface wave methods have been widely used to reconstruct velocity structures on different scales. The methods are becoming increasingly popular in urban areas because of the urgent needs of urban development. Unlike years of recordings on continental scales, the duration of recordings in urban areas usually lasts only for hours, even minutes. Thanks to huge amounts of cultural noise, especially from traffic, passive surface wave methods have proved their efficiency in urban areas, even with relatively short-time recordings. However, data needs to be processed more carefully for obtaining valuable information with limited recordings. Based on the characteristic of surface waves in − domain, we propose a formula to assess the intensity of surface waves in a linear array, thereby forming an automated data selection algorithm by selecting those segments with energetic surface waves. Results of a real-world example demonstrated the procedures of the proposed algorithm, and its effectiveness and reliability are therefore validated.

1. Introduction

Passive surface wave methods using ambient noise have been developed since the 1950s [1]. They do not require an active source, therefore, becoming more and more popular, especially in urban areas. Some researchers utilize a linear array, which is convenient to deploy, to image passive surface waves [2-4]. Cheng et al. (2016) proposed a multichannel analysis of passive surface waves (MAPS) [5], which combines cross-correlation and multichannel analysis of surface waves [6-8] to handle the effects of directional noise sources.

Passive surface wave methods usually require to preprocess the raw data, which is vital for obtaining a good dispersion measurement [9]. Data selection is a promising alternative to improve the imaging of passive surface wave methods. Data selection of single station records usually aims to reduce the effects of records anomalies, for example, gaps, transient signals. Array-based data selection utilizing the spatial-time characteristics of signals is also frequently used. It proves to be effective in suppressing non-stationary noise and improving the stacked result [10-12]. Nevertheless, traditional data selection methods generally work with already split segments, therefore being less sensitive to changes of ambient noise over time.
In this paper, we propose a simple approach to characterize the surface-wave intensity of a linear array in continuously recorded ambient noise. Therefore, an automated data selection algorithm is proposed by selecting time windows with energetic surface waves in ambient noise records. A field example demonstrates the applicability and reliability of the proposed algorithm.

2. Method

$\tau - p$ transform (or called slant-stacking) is commonly used for plane wave decomposition based on different characteristics of waves in the x-t domain. The surface waves always appear in the x-t domain as straight lines with different slopes. This is due to their dispersive characteristic in velocity varying medium. Therefore, based on the idea of slant-stacking, we propose to characterize the surface-wave intensity in the wavefield $s(x, t)$ by

$$I(\tau) = \frac{\max |u(p, \tau)|_{p_{\min} < p < p_{\max}}}{\sum_{n=1}^{N} \text{RMS}(s(x_n, t + x_n p_{\min} < t < t + x_n p_{\max}))}$$

(1)

where $N$ is the number of traces in the wavefield and $\text{RMS}(s(x_n, t + x_n p_{\min} < t < t + x_n p_{\max}))$ represents the root-square-mean (RMS) value of the $n$-th trace in a “signal window” (defined by a slowness range, $p_{\min} < p < p_{\max}$, with the offset $x_n$ of the $n$-th trace); $u(p, \tau)$ is the value of slant-stacking:

$$u(p, \tau) = \sum_{n=1}^{N} s(x_n, t = \tau + px_n)$$

(2)

where $\tau$ is the intercept on the time axis. Using equations 1 and 2, we can obtain the distribution of surface-wave intensity over time. In the preprocessing of passive surface wave methods, the long-period records are usually split into short-period segments. Therefore we can split the records based on the surface-wave intensity to obtain surface-wave-rich segments, thereby benefiting the imaging of surface wave.

In order to better reflect the distribution of surface waves in ambient noise, we calculate the maximum values of $I(\tau)$ in a sliding window, whose length is equal to the longest period of the analyzed surface waves, to obtain $E(\tau)$, where higher values represent higher possibility of the existence of surface waves. After setting the length of segments, we can calculate the moving mean of a sliding window, in which segments with high values will be automated selected. In the next section, we will use a field example to prove the applicability and reliability of the proposed algorithm.

3. A field example

We deployed 48 three-component ZLand wireless geophones (called nodes) along Qianchao Road in Hangzhou, China, to acquire ambient noise data with a 1000 Hz sampling rate and a 5 m space interval. We recorded the ambient noise for 1 hour. Figure 1 shows the $E(\tau)$ curve (with $+p$) representing the distribution of forward propagating surface waves over 1 hour. Note that the length of the sliding window was set to 1 s because we wanted to analyze surface waves above 1 Hz. As we can see, the distribution was time-variant, and lots of time windows had relatively low energy of surface waves, which needed to be thrown out, or they might deteriorate the stacked dispersion measurement. To ensure enough stacking of dispersion measurements, we kept the number of selected segments equal to half of the number of traditional split segments by setting a proper threshold, which was adjustable with a different number of required segments.

The segment length was set to 10 s with a 50% overlap. The preprocessing parameters included remove-means, remove-trends, and spectral whitening. Note that we used both $+p$ and $-p$ to represent two propagation directions along the road. Therefore two $E(\tau)$ curves were used for data selection, and the selected segments were used to image by using MAPS with causal and acausal cross-correlation functions, respectively. The two kinds of dispersion measurements were stacked to form the final measurement. Figure 2 shows the dispersion measurement of MAPS of different data durations with and without data selection. Note that, the left panel shows the results of MAPS using the mean of causal and acausal cross-correlation functions without data selection. The dispersion image was distorted around 5
Hz before data selection for 5 min data shown in Figure 2a and the improvement were very limited with 10 min (Figure 2c) or 20 min (Figure 2e) data. The results with data selection, however, showed a significant improvement compared with those without data selection. The distortion around 5 Hz was mitigated, and the dispersion image was getting more and more pure and obvious with increasing data durations.

![Figure 1. The distribution curve of forward propagating surface waves over 1 hour.](image)

![Figure 2. Dispersion images of different data durations without and with data selection. (a), (c), and (e) are the results without data selection and the data durations are 5 min, 10 min, and 20 min, respectively, and (b), (d), and (f) are similar, but with data selection.](image)

To validate the accuracy of the dispersion images above, we plotted the dispersion images of MAPS using the whole 1 hour data. Figure 3a and Figure b are the results without and with data selection,
respectively. They are very similar to each other and are in good agreement with Figure 2d and Figure f, which proves the efficiency and accuracy of the proposed data selection algorithm.

![Dispersion images of 1 hour data. (a) is the result without data selection, and (b) is the result of data selection.](image)

**Figure 3.** Dispersion images of 1 hour data. (a) is the result without data selection, and (b) is the result of data selection.

4. **Discussion**
Data selection is not a panacea. When the data quality is dramatically high or extremely poor, we should not expect data selection to improve the result. For example, after an hour of stacking (Figure 3), the result without data selection was already good enough, and the data selection cannot improve it much. Incoherent noise may challenge imaging quality; however, when the recording time is limited. Therefore the data selection has the potential to make a significant improvement, as shown in Figure 2.

The proposed algorithm aims to select those segments with strong surface-wave characteristics in $x-t$ domain. Therefore, it is suitable for passive surface wave methods that use raw ambient noise recordings. Meanwhile, it is also effective for interferometry methods like MAPS in this paper because energetic surface waves would contribute to the retrieval of coherent signals.

There is no “silver bullet” for data selection [9]. Based on different definitions of “good” ambient noise, different data selection methods might select different data under complex noise environments. Our algorithm selects data in $x-t$ domain, therefore ignoring the quality differences of data in different frequencies. Frequency-dependent data selection might be a better way to evaluate the quality of surface waves in ambient noise and it will be our future work.

5. **Conclusions**
We developed an automated data selection algorithm based on the assessment of surface-wave characteristic. The procedures of the proposed algorithm were introduced through a field example and the improvement of dispersion measurement proves its applicability and reliability.

**Acknowledgements**
This study is supported by the National Nature Science Foundation of China under Grant No. 41830103 and the project of Nanjing Center of China Geological Survey under Grant No. DD20190281.

**References**
[1] Aki K 1957 Space and time spectra of stationary stochastic waves, with special reference to microtremors Bull. Earthq. Res. Inst. 35 415–456
[2] Louie J N 2001 Faster, better: shear-wave velocity to 100 meters depth from refraction microtremor arrays Bulletin of the Seismological Society of America 91 347–364
[3] Park C B, Miller R, Laflen D, Neb C, Ivanov J, Bennett B and Huggins R 2004 Imaging dispersion curves of passive surface waves In SEG Technical Program Expanded Abstracts 2004 1357–1360 Society of Exploration Geophysicists
[4] Le Feuvre M, Joubert A, Leparoux D and Côte P 2015 Passive multi-channel analysis of surface waves with cross-correlations and beamforming. Application to a sea dike *Journal of Applied Geophysics* **114** 36–51

[5] Cheng F, Xia J, Luo Y, Xu Z, Wang L, Shen C, Liu R, et al. 2016 Multichannel analysis of passive surface waves based on crosscorrelations *Geophysics* **81** EN57–EN66

[6] Song Y Y, Castagna J P, Black R A and Knapp R W 1989 Sensitivity of near-surface shear-wave velocity determination from Rayleigh and Love waves in *SEG Technical Program Expanded Abstracts 1989* 509–512 Society of Exploration Geophysicists

[7] Park C B, Miller R D and Xia J 1999 Multichannel analysis of surface waves *Geophysics* **64** 800–808

[8] Xia J, Miller R D and Park C B 1999 Estimation of near-surface shear-wave velocity by inversion of Rayleigh waves *Geophysics* **64** 691–700

[9] Cheng F, Xia J, Xu Z, Hu Y and Mi B 2018 Frequency–wavenumber (fk)-based data selection in high-frequency passive surface wave survey *Surveys in Geophysics* **39** 661–682

[10] Cheng F, Xia J, Behm M, Hu Y and Pang J 2019 Automated Data Selection in the Tau–p Domain: Application to Passive Surface Wave Imaging *Surveys in Geophysics* **40** 1211-1228

[11] Zhou C, Xi C, Pang J and Liu Y 2018 Ambient noise data selection based on the asymmetry of cross-correlation functions for near surface applications *Journal of Applied Geophysics* **159** 803–813

[12] Pang J, Cheng F, Shen C, Dai T, Ning L and Zhang K 2019 Automatic passive data selection in time domain for imaging near-surface surface waves *Journal of Applied Geophysics* **162** 108–117