Analysis of Rayleigh waves with circular wavefront: a maximum likelihood approach

Geophysical Journal International

Geophys. J. Int. (2017) **210,** 1570–1580
Advance Access publication 2017 May 29
GJI Seismology

doi: 10.1093/gji/ggx225

Analysis of Rayleigh waves with circular wavefront: a maximum likelihood approach

Stefano Maranò, Manuel Hobiger, Paolo Bergamo and Donat Fäh

ETH Zurich, Swiss Seismological Service, 8092 *Zürich, Switzerland. E-mail: stefano.marano@sed.ethz.ch*

Accepted 2017 May 23. Received 2017 May 6; in original form 2016 June 18

SUMMARY

Analysis of Rayleigh waves is an important task in seismology and geotechnical investigations. In fact, properties of Rayleigh waves such as velocity and polarization are important observables that carry information about the structure of the subsoil. Applications analysing Rayleigh waves include active and passive seismic surveys. In active surveys, there is a controlled source of seismic energy and the sensors are typically placed near the source. In passive surveys, there is not a controlled source, rather, seismic waves from ambient vibrations are analysed and the sources are assumed to be far outside the array, simplifying the analysis by the assumption of plane waves. Whenever the source is in the proximity of the array of sensors or even within the array it is necessary to model the wave propagation accounting for the circular wavefront. In addition, it is also necessary to model the amplitude decay due to geometrical spreading. This is the case of active seismic surveys in which sensors are located near the seismic source. In this work, we propose a maximum likelihood (ML) approach for the analysis of Rayleigh waves generated at a near source. Our statistical model accounts for the curvature of the wavefront and amplitude decay due to geometrical spreading. Using our method, we show applications on real data of the retrieval of Rayleigh wave dispersion and ellipticity. We employ arrays with arbitrary geometry. Furthermore, we show how it is possible to combine active and passive surveys. This enables us to enlarge the analysable frequency range and therefore the depths investigated. We retrieve properties of Rayleigh waves from both active and passive surveys and show the excellent agreement of the results from the two surveys. In our approach we use the same array of sensors for both the passive and the active survey. This greatly simplifies the logistics necessary to perform a survey.

Key words: Time-series analysis; Site effects; Surface waves and free oscillations.

1 INTRODUCTION

Analysis of Rayleigh waves is an important task in seismology and geotechnical investigations. It enables us to gather knowledge of geological and geophysical features of the subsoil. In fact, properties of Rayleigh waves such as velocity and polarization are important observables that carry information about the structure of the subsoil. The analysis of these properties allows geophysicists to gain insight into the subsoil avoiding more expensive invasive techniques (e.g. borehole techniques). For example, seismic surveying methods represent a valuable tool in oil and gas prospection (Sheriff & Geldart 1995) and in geophysical investigations (Tokimatsu *et al.* 1992; Okada 1997).

Applications analysing Rayleigh waves include active and passive seismic surveys. In active surveys, there is a controlled source of seismic energy and the sensors are typically placed near the source (Park *et al.* 1999). In passive surveys, there is not a controlled source, rather, seismic waves from ambient vibrations are analysed (Bonnefoy-Claudet *et al.* 2006) and the sources are assumed to be far outside the array. Measurements acquired in active and passive surveys need to be analysed with array processing methods.

A widely popular assumption in array processing is to consider the source of energy to be located in the far field. This assumption enables us to model the wavefront as plane and consider the wave amplitude as constant, greatly simplifying the analysis of the signals. A myriad of applications are well suited to this setting and a huge amount of literature has been produced using this assumption (Krim & Viberg 1996; Van Trees 2002). In particular, a number of authors use this assumption in seismological applications, including the study of earthquake in the far field (Capon 1969) and the analysis of recordings from passive surveys (Ohrnberger *et al.* 2004; Poggi & Fäh 2010; Maranò *et al.* 2012).

However, whenever the source is in the proximity of the array of sensors or even within the array it is necessary to model the wave propagation accounting for the circular wavefront. In addition, it is also necessary to model the amplitude decay due to geometrical

spreading. This is the case of active seismic surveys, in which sensors are located near the seismic source.

This manuscript is primarily concerned with the development of a method for the analysis of Rayleigh waves recorded from active surveys. Passive surveys will be considered to show the viability of performing jointly active and passive surveys with the same array of sensors. In fact, using the proposed technique, we are able to use the same array configuration designed for passive surveying also in the active survey.

In the seminal paper by Park *et al.* (1999), Rayleigh waves generated by a near field source are analysed. The method presented in Park *et al.* (1998) is used for processing the data. In this approach, the amplitude decay due to geometrical spreading is accounted for by multiplying the measured signal by a factor compensating for the known amplitude reduction at the receiver. Then, all the measurements are transformed in the phase velocity-frequency domain by using an integral transform. In this approach, a single sensor component is processed. We emphasize that multiplying the measured signal by a given factor is not a robust approach. In fact, while it addresses the effect of geometrical spreading, it also amplifies any undesired portion of the signal (e.g. noise).

A cross-correlation method accounting for the circular wavefront is presented by Almendros *et al.* (1999). In this work geometrical spreading is accounted for by dividing the signal amplitude by a factor of $1/\sqrt{r}$, where r is the sensor-source distance. Boiero *et al.* (2011) consider a source radiating in the near field. The distances between source and sensors are computed and then a method for frequency-wavenumber analysis is applied as on a line of sensors, ignoring the amplitude decay. Strobbia & Foti (2006) propose a linear regression technique. The phase of the signal is estimated at each sensor, then the phase at each sensor is used in the regression as the dependent variable and the sensor offset is used as the independent variable. A straight line is fitted to the phase-offset pairs and the velocity is inferred from the slope of the line. The authors also suggested to include a weighting in the regression, this indirectly accounts to some extent for the smaller wave energy at far offset sensors.

In this manuscript, we propose a maximum likelihood (ML) method for the analysis of Rayleigh waves generated by a seismic source in the proximity of the sensors. We construct a statistical model accounting for the curved wavefront and the amplitude decay due to geometrical spreading. The proposed method, models exactly the Rayleigh wave equation at arbitrary distance from the source. In our approach, we model all the three sensor components jointly and retrieve Rayleigh wave velocity and Rayleigh wave ellipticity for the fundamental mode and, when possible, for higher modes. We consider arrays with arbitrary geometry.

Park *et al.* (2005) highlight the advantages of combining passive and active surveys. Indeed, it is possible to enlarge the analysable frequency range and therefore the depths investigated with the surveys. Being able to observe properties of surface wave over a broader frequency range, enables us to reduce the uncertainty in the inversion for a structural model. In our work, we show how it is possible to combine active and passive surveys employing the same array of sensors and therefore greatly simplifying the logistic required in performing the survey. In previous works combining active and passive surveys two distinct sets of hardware are needed (Park *et al.* 2005; Moro *et al.* 2015).

To summarize, the contribution of this paper is twofold

(i) We propose an ML approach for the analysis of Rayleigh waves. Our statistical model accounts for the curvature of the wavefront and amplitude decay due to geometrical spreading. Using our method, we show applications on real data of the retrieval of Rayleigh wave dispersion and ellipticity. We employ arrays with arbitrary geometry.

(ii) We show how it is possible to combine active surveys and passive surveys. We retrieve properties of Rayleigh waves from both types of surveys and show the excellent agreement of the results from the two surveys. In our approach we use the same array of sensors for both the passive and the active survey. This greatly simplifies the logistics necessary to perform a survey.

The remainder of this paper is organized as follow. In Section 2, we describe the equations we use to describe the ground motion induced by a single Rayleigh wave with circular wavefront. The proposed approached is presented in Section 3. Applications of our method to real sites are presented in Section 4. The combination of both passive and active surveys at the same site is considered. Concluding remarks are found in Section 5.

2 SYSTEM MODEL

We consider the analysis of Rayleigh waves measured by seismic sensors in the vicinity of the source. We account for the source being near or even within the array of sensors measuring the waves. Therefore, it is clear that the common assumption of plane wavefront widely used in array processing is not valid and it is necessary to model the curvature of the wavefront. In addition, it is also necessary to account for the wave amplitude decay due to geometrical spreading.

Our algorithm relies on wave equations describing the Rayleigh wave particle motion at arbitrary distance from the source. Therefore, the model presented in this section is valid both near the source and far away from the source. In addition, the proposed method overcomes the common limitation requiring the sensors to be distanced from the source to allow the Rayleigh wave to fully develop (Stokoe *et al.* 1994). In this work, we restrict our interest on the analysis of Rayleigh waves measured at the surface.

Let $\tilde{u}_x(\omega, \boldsymbol{p})$, $\tilde{u}_y(\omega, \boldsymbol{p})$ and $\tilde{u}_z(\omega, \boldsymbol{p})$ be the spectral displacement along the three Cartesian axes at angular frequency ω and position $\boldsymbol{p} \in \mathbb{R}^2$. We use a right-handed coordinate system with the vertical z-axis pointing upwards.

In this work, we consider a point source with a vertical force. Using such a compressional wave source, more than two thirds of the seismic energy generated is converted to Rayleigh waves (Richart *et al.* 1970). Therefore, due to the nature of the source and the different polarization of Love and Rayleigh waves, we will not consider Love waves in this work.

The radiation pattern of Rayleigh waves for vertically heterogeneous media was derived by Saito (1967). Considering a point source and a single force in the direction of the z-axis, the displacement field at arbitrary position depends on the distance from the source. Let $\boldsymbol{p}_S$ denote the source position, then $r = \|\boldsymbol{p} - \boldsymbol{p}_S\|$ is the distance between an arbitrary position $\boldsymbol{p}$ and the source, with $\|\boldsymbol{x}\|$ being the norm of the vector $\boldsymbol{x}$. The three-components spectral displacement at position $\boldsymbol{p}$, induced by a single Rayleigh wave is

$$\tilde{u}_x(\omega, \boldsymbol{p}) = -\sin\xi \cos\psi \, H_1^{(2)}(\kappa r)\, s(\omega) \tag{1a}$$

$$\tilde{u}_y(\omega, \boldsymbol{p}) = -\sin\xi \sin\psi \, H_1^{(2)}(\kappa r)\, s(\omega) \tag{1b}$$

$$\tilde{u}_z(\omega, \boldsymbol{p}) = \cos\xi \, H_0^{(2)}(\kappa r)\, s(\omega)\,, \tag{1c}$$

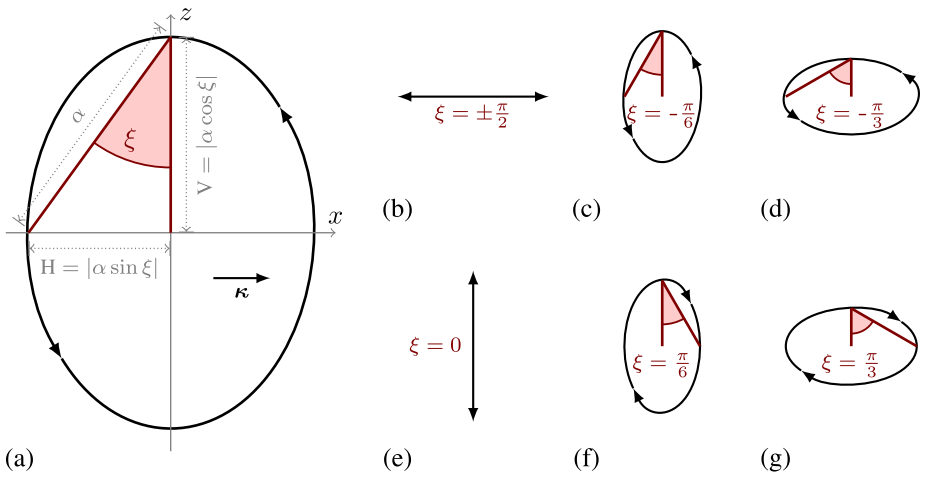


Figure 1. Depiction of the Rayleigh wave particle motion, valid for large κr, for several values of the ellipticity angle ξ. (a) The elliptical particle motion for a retrograde Rayleigh wave. The quantities ξ, α, H and V are shown. (b) The Rayleigh wave is horizontally polarized for $\xi = \pm\pi/2$. (c) Retrograde particle motion for $\xi = -\pi/6$. (d) Retrograde particle motion for $\xi = -\pi/3$. (e) The Rayleigh wave is vertically polarized for $\xi = 0$. (f) Prograde particle motion for $\xi = \pi/6$. (g) Prograde particle motion for $\xi = \pi/3$.

for $\omega > 0$ and where ξ is the ellipticity angle of the Rayleigh wave, ψ is the azimuth angle formed by the vector $\boldsymbol{p} - \boldsymbol{p}_s$ and measured counter-clockwise from the x-axis, $\kappa = 2\pi/\lambda$ is the wavenumber, and λ is the wavelength. With $H_\nu^{(2)}$ we denote Hankel functions of the second kind of order ν (Abramowitz & Stegun 1964). The quantity $s(\omega)$ is the spectrum of the source function.

The angle $\xi \in [-\pi/2; +\pi/2]$ is called ellipticity angle of the Rayleigh wave. For sufficiently large values of κr, it determines the eccentricity and the sense of rotation of the particle motion. More details are provided below.

In eq. (1), the Hankel functions model the dependency of the measured spectrum as a function of the sensor-source distance r. The trigonometric functions of ψ and ξ account for the different amplitudes of the Rayleigh wave on the different sensor components due to the sensor orientation and to the peculiar elliptical particle motion of the Rayleigh wave, respectively.

We now consider the asymptotic expression for eq. (1), valid for large κr. Using the asymptotic expansions for $H_0^{(2)}$ and $H_1^{(2)}$ (Abramowitz & Stegun 1964) we obtain the following expressions

$$\tilde{u}_x(\omega, \boldsymbol{p}) \sim \sqrt{\frac{2}{\pi\kappa r}} \sin\xi \cos\psi \, \exp\left(i\left(-\frac{\pi}{4} - \kappa r\right)\right) s(\omega) \qquad (2a)$$

$$\tilde{u}_y(\omega, \boldsymbol{p}) \sim \sqrt{\frac{2}{\pi\kappa r}} \sin\xi \sin\psi \, \exp\left(i\left(-\frac{\pi}{4} - \kappa r\right)\right) s(\omega) \qquad (2b)$$

$$\tilde{u}_z(\omega, \boldsymbol{p}) \sim \sqrt{\frac{2}{\pi\kappa r}} \cos\xi \, \exp\left(i\left(\frac{\pi}{4} - \kappa r\right)\right) s(\omega)\,, \qquad (2c)$$

which are valid for $\kappa r \to \infty$.

Eqs (2a)–(2c) are analogous to the plane wavefront equations used in many array processing methods. In particular, they are compatible with the equations used in previous work by the authors (Maranò *et al.* 2012; Maranò *et al.* 2017). The amplitude decay proportional to $1/\sqrt{r}$ is typically omitted in plane wavefront models, in fact its effect is negligible for small arrays at large distances from the source. We observe that the $\pi/2$ phase shift between the vertical and the horizontal components is immediately evident in (2).

From eqs (2a)–(2c) we are able to provide a simple interpretation of the Rayleigh wave ellipticity angle ξ and of the particle motion of the wave. Fig. 1(a) depicts the particle motion of a Rayleigh wave and how the ellipticity angle is defined. Rayleigh wave particle motion is depicted with a black ellipse and the sense of rotation with the black arrows on the ellipse. The z axis is vertical to the surface. The x axis lies on the surface and is parallel to the wavevector $\boldsymbol{\kappa}$. The ellipticity angle ξ has the vertex at the topmost intersection between the z axis and the particle motion ellipse. The first side of the ellipticity angle is the z axis. The second side is the line connecting the vertex with the intersection of the x axis and the particle motion ellipse (the intersection is chosen such that the particle motion at the intersection has a $+\pi/2$ radians shift with respect to the vertex). This parametrization to describe Rayleigh wave ellipticity was first introduced in Maranò *et al.* (2012) and further investigated in Maranò *et al.* (2017).

In Fig. 1(a), the quantities H and V are defined. They represent the amplitude of the horizontal and the vertical motion, respectively. The quantity $\mathrm{H/V} = |\tan\xi|$ is known as the ellipticity of the Rayleigh wave. Fig. 1(a) also provides a graphical interpretation of the amplitude α. From eqs (2a)–(2c) we understand that the amplitude depends on the source and on the distance from the source.

In the Supporting Information, we briefly discuss how the distance from the source affects the particle motion.

When $\xi \in (-\pi/2, 0)$, the Rayleigh wave particle motion is said to be *retrograde* (i.e. the oscillation on the vertical component (u_z) is shifted by $+\pi/2$ radians with respect to the oscillation on the direction of propagation). Two possible retrograde particle motions and corresponding ellipticity angles are shown in Figs 1(c) and (d). When $\xi \in (0, \pi/2)$ the particle motion is said to be *prograde*, *cf.* Figs 1(f) and (g). For $\xi = \pm\pi/2$ (*cf.* Fig. 1b) and $\xi = 0$ (*cf.* Fig. 1e) the particle motion polarization is horizontal and vertical, respectively.

The real-valued time domain signals $u_x(t, \boldsymbol{p})$, $u_y(t, \boldsymbol{p})$, $u_z(t, \boldsymbol{p})$ describe the ground motion induced by a Rayleigh wave. They are related to their frequency domain counterparts $\tilde{u}_x(\omega, \boldsymbol{p})$, $\tilde{u}_y(\omega, \boldsymbol{p})$, $\tilde{u}_z(\omega, \boldsymbol{p})$ of eqs (1) and (2) by an inverse Fourier transform.

3 PROPOSED METHOD

Our goal is to perform ML estimation of the wave parameters $\boldsymbol{\theta}$ from noisy observations $\boldsymbol{y}$. Wave parameters include the

frequency-dependent wavenumber κ and ellipticity angle ξ, that is, $\boldsymbol{\theta} = (\kappa, \xi)^\mathsf{T}$. In addition, the source function $s(\omega)$ and the noise level at the different sensors are also estimated. The source position $\boldsymbol{p}_s$ is known.

First, we seek an efficient way to compute the likelihood of the observations $\boldsymbol{y}$ for a wave parameter vector $\boldsymbol{\theta}$. Second, a maximization of the likelihood function enables us to perform ML estimation and to obtain an estimate $\hat{\boldsymbol{\theta}}$ of the parameter vector. We now provide a self-contained and high-level description of the method. We provide some additional technical details in the following section.

We rely on noisy measurements from L channels. In the case of N three-component sensors, we have $L = 3N$. In particular, on the ℓth channel the measurements $Y_k^{(\ell)}$ at discrete instants t_k are

$$Y_k^{(\ell)} = u(t_k, \boldsymbol{p}_\ell) + Z_k^{(\ell)} \quad \text{for } k = 1, \ldots, K\,, \tag{3}$$

where $u(\boldsymbol{p}_\ell, t_k)$ is a deterministic function with unknown wave parameters (as in eq. 1) and $Z_k^{(\ell)}$ is zero-mean additive white Gaussian noise with variance σ_ℓ^2. With this signal model, the probability density function (PDF) of the measurements $\boldsymbol{Y}$ is

$$p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{\ell=1}^{L}\prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma_\ell^2}} \mathrm{e}^{-\left(y_k^{(\ell)} - u_k^{(\ell)}\right)^2/2\sigma_\ell^2}\,, \tag{4}$$

where we grouped all the measurement as $\boldsymbol{y} \triangleq \{y_k^{(\ell)}\}_{l=1,\ldots,L}^{k=1,\ldots,K}$ and defined $u_k^{(\ell)} \triangleq u(t_k, \boldsymbol{p}_\ell)$.

In the proposed technique, instead of computing the likelihood of the observation directly from (4), we model it by means of a factor graph (Loeliger *et al.* 2007). This enables the algorithm a more efficient computation of the likelihood. Details of the factor graph are provided in the following.

Given observations $\boldsymbol{y}$ of $\boldsymbol{Y}$, we call the likelihood of the observation $f(\boldsymbol{\theta}) \triangleq p_{\boldsymbol{Y}}(\boldsymbol{y} = \boldsymbol{Y}|\boldsymbol{\theta})$. An ML estimate $\hat{\boldsymbol{\theta}}$ is found by maximizing the likelihood

$$\hat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} f(\boldsymbol{\theta})\,. \tag{5}$$

This maximization is performed analytically for the parameter $s(\omega)$. It is performed numerically for the remaining parameters using a grid search first and then a numerical optimization routine (Byrd *et al.* 1995).

In an actual survey, multiple Rayleigh waves may be present simultaneously. This happens, for example, in presence of multiple modes of propagation. For model selection, that is, the choice of the number of waves to consider in the model, we employ a variant of the Bayesian information criterion (BIC; Schwarz 1978). We use a penalized version of BIC where a scalar γ allows us to control the complexity of the model. The penalized BIC is defined as

$$\mathrm{BIC}_\gamma = -2f(\hat{\boldsymbol{\theta}}) + \gamma N_p \ln(LK)\,, \tag{6}$$

where N_p denotes the number of parameters of the statistical model and LK is the number of measurements. The parameter $\gamma \geq 0$ enables us to follow a pure BIC model selection strategy for $\gamma = 1$, an ML approach (which will result in overfitting) for $\gamma = 0$, or any intermediate strategy for different values $\gamma \in (0, 1)$. This approach for model selection was first used by the authors in the analysis of ambient vibration recordings Maranò *et al.* (2017).

3.1 Factor graph and likelihood computation

In this section, we detail the machinery behind the computation of the likelihood of the observations $f(\boldsymbol{\theta})$. The derivation of the algorithm relies on a specific type of graphical model called factor graph

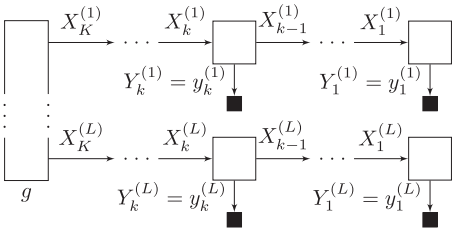


Figure 2. Overall factor graph.

(Loeliger *et al.* 2007). Graphical models provide a framework for modelling complex systems and solve inference problems involving a large number of random variables. Using a factor graph it is possible to represent the factorization of a PDF of random variables and parameters. Relationships among the variables on the graph are made explicit by the structure of the graph itself. Using sum-product message passing (a.k.a. belief propagation) it is possible to efficiently compute the marginal probability of any random variable and to compute the likelihood of the observations (Kschischang *et al.* 2001).

We now describe in detail the construction of the factor graph. The overall view of the factor graph is given in Fig. 2. There are L horizontal state-space models (SSMs) connecting the measurements $\boldsymbol{y}$. On the left, the glue factor g links together the SSMs. We now describe the construction of each SSM and then of the glue factor.

For every channel ℓ, we consider a discrete time SSM with state $X_k^{(\ell)} \in \mathbb{C}$ in order to relate the measurements $\{Y_k^{(\ell)}\}_{k=1,\ldots,K}$ with their Fourier transform V_ℓ. For $k = 1, \ldots, K$

$$X_{k+1}^{(\ell)} = F X_k^{(\ell)} \tag{7a}$$

$$Y_k^{(\ell)} = \mathrm{Re}\left(X_k^{(\ell)}\right) + Z_k^{(\ell)}, \tag{7b}$$

with $F = \exp(i\omega\Delta t)$ and Δt being the uniform temporal sampling step. The quantity $V_\ell \triangleq X_K^{(\ell)}$ can be interpreted as the Fourier transform of the measurements on channel ℓ. Fig. 3 depicts a part of the factor graph of (7) for a single measurement $Y_k^{(\ell)}$.

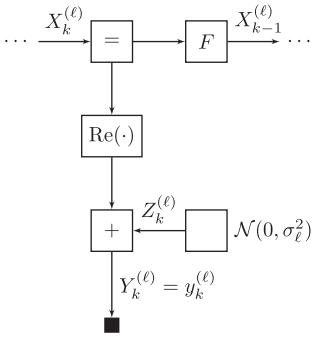


Figure 3. Factor graph of (7).

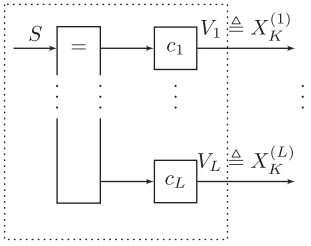


Figure 4. Glue factor g *cf.* (8).

The effect of the SSM (7) is akin to the effect of a discrete Fourier transform. More details are available in Reller *et al.* (2011).

As the next step, we wish to couple the state vector of all the channels $\{V_\ell\}_{\ell=1,\dots,L}$ with

$$V_\ell = c_\ell S\,, \tag{8}$$

for $\ell = 1, \dots, L$ where $c_\ell \in \mathbb{C}$ reflects the amplitude decays and phase delays prescribed by eq. (1) and S is representative of the complex amplitude of the source signal $s(\omega)$ at frequency ω.

As a concrete example, let us consider the channels ℓ, $\ell+1$, $\ell+2$ referring to measurements taken at distance r from the source on the x, y, z sensor components, respectively. The corresponding values for c_ℓ are

$$c_\ell = -\sin\xi\cos\psi\, H_1^{(2)}(\kappa r) \tag{9a}$$

$$c_{\ell+1} = -\sin\xi\sin\psi\, H_1^{(2)}(\kappa r) \tag{9b}$$

$$c_{\ell+2} = \cos\xi\, H_0^{(2)}(\kappa r)\,, \tag{9c}$$

cf. eq. (1).

Eqs (7) and (8) precisely describe the relationships between all the quantities involved, from the measurements $\{Y_k^{(\ell)}\}_{l=1,\dots,L}^{k=1,\dots,K}$ to the Fourier transform of the measurements $\{V_\ell\}_{\ell=1,\dots,L}$, and to the wave at the source signal S. We emphasize that the presented statistical model is linear. We model such relationships using a factor graph (Loeliger *et al.* 2007) and use the sum-product algorithm to compute the likelihood of the observations. The sum-product algorithm is a message passing algorithm that enables us to compute, in this instance, the likelihood of the observations.

We consider message passing on the factor graph using scaled Gaussian messages. The message on the edge X, in the direction of the arrow, has the form of a scaled Gaussian PDF

$$\overrightarrow{\mu}_X(x) = \overrightarrow{\gamma}_X e^{-x^\mathsf{T}\overrightarrow{\mathbf{W}}_X x/2 + x^\mathsf{T}\overrightarrow{\mathbf{W}}_X\overrightarrow{\boldsymbol{m}}_X}\,, \tag{10}$$

and is parametrized by scale factor $\overrightarrow{\gamma}_X$, mean vector $\overrightarrow{\boldsymbol{m}}_X$, and inverse covariance matrix $\overrightarrow{\mathbf{W}}_X$. For the sake of exposition, so far we considered complex vectors and matrices. For the actual message passing implementation we consider their real equivalent and use the message passing rules for Gaussian messages given in Loeliger *et al.* (2007).

First, we apply message passing rules to L separate SSMs described by (7) and obtain the messages $\overleftarrow{\mu}_{X_K^{(\ell)}}(x)$ for $\ell = 1, \dots, L$. Second, we apply message passing to the glue factor described by (8) and obtain the message $\overleftarrow{\mu}_S(s)$.

The likelihood is computed by evaluating the backward message on the edge $\boldsymbol{S}$ at the value of the mean $\overleftarrow{\boldsymbol{m}}_S$. It can be shown that the log-likelihood is

$$\ln f(\boldsymbol{\theta}) = \overleftarrow{\mu}_S(\overleftarrow{\boldsymbol{m}}_S) = \frac{1}{2}\overleftarrow{\boldsymbol{m}}_S^\mathsf{T}\overleftarrow{W}_S\overleftarrow{\boldsymbol{m}}_S + \ln\overleftarrow{\gamma}_S\,. \tag{11}$$

We observe that the messages pertaining the SSM in (7) do not depend on the wave parameters $\boldsymbol{\theta}$, that is, are sufficient statistics, and can be computed only once. In contrast, the message passing for (8) needs to be repeated for different values of $\boldsymbol{\theta}$.

Likelihood computation in presence of multiple waves (e.g. multiple modes of propagation of the Rayleigh wave) may require an iterative approach such as what is used by the authors in Maranò *et al.* (2012).

In an alternative development, would be possible to model directly the partial differential equations the wave equation in circular coordinates, following the approach suggested in Maranò *et al.* (2015).

4 APPLICATIONS

In order to validate the effectiveness of the proposed approach we performed measurements at three different sites and we compare the Rayleigh wave dispersion curve and Rayleigh wave ellipticity curves retrieved with different methods and instrumentation. Specifically, at each site we perform the following analyses:

(i) Using the approach proposed in this manuscript, we processed measurements from active surveys using both a two-dimensional array of triaxial velocimeters and a line of triaxial geophones. In this section, these two settings are referred to as *Active 2D* and *Active 1D*. Rayleigh wave dispersion and ellipticity angle curves are retrieved.

(ii) We analyse ambient vibrations recordings using the method proposed in Maranò *et al.* (2012). Ambient vibrations are recorded using the same array deployment and instrumentation used for the Active 2D setting. We refer to this setting as *Passive*. Rayleigh wave dispersion and ellipticity angle curves are retrieved.

(iii) The geophone line is also processed with FK analysis, a standard and widely used processing technique for active surface wave data (Socco *et al.* 2010). This offers a benchmark of the proposed method with existing techniques. This setting is referred to as FK *Vertical* and/or FK *Longitudinal*, depending on the processed component. Rayleigh wave dispersion is retrieved.

The aforementioned analyses enable us to validate the proposed technique in two ways. First, we compare Rayleigh wave dispersion and ellipticity angle curves obtained with the same 2-D array geometry analysing passive and active recordings. Second, we compare the Rayleigh wave dispersion curves retrieved with the technique presented in this paper with the curve retrieved using a comparison approach, that is, FK analysis.

Fig. 5(a) shows a map of Switzerland with the locations of the sites considered in this manuscript. At each site both a passive and an active survey are performed. Figs 5(b)–(d) show the detail of the surveys at each site. Yellow triangles show the locations of the sensors of the 2-D array used in the Active 2D and Passive processing. Each sensor is a triaxial Lennartz 5s velocimeter and the recording is digitized with a Quanterra Q330 digitizer with a sampling rate of 200 Hz. The line of triaxial geophones is depicted with a solid green line. Geophones have a corner frequency of 4.5 Hz and are digitized with a sampling rate of around 16 kHz. Geophone recordings are used for the Active 1D and FK processing. Shooting locations, that is, the different locations where the hammering was performed, are pinpointed with red stars.

At the site of SLUB, *cf.* Fig. 5(d), the two-dimensional array was not recording during the active experiment, therefore active measurements are analysed with a line of vertical-component geophones, depicted with a green line, and with a line of Lennartz 5s sensors, depicted with blue circles.

Concerning the Active 2D analysis, the instruments are continuously recording and no triggering instrumentation is used during the active measurements. At each shooting location multiple shots are performed. The timing of each shot is identified with human intervention by inspecting the recordings visually. For the Passive experiment, the same array is recording the ambient vibrations wavefield for around two hours.

Recordings from sensors too close to the active source are discarded whenever the waveform exhibits clipping or long-period disturbances induced by the hammering and tilt. Few shooting positions which did not deliver usable results were also discarded during the processing.

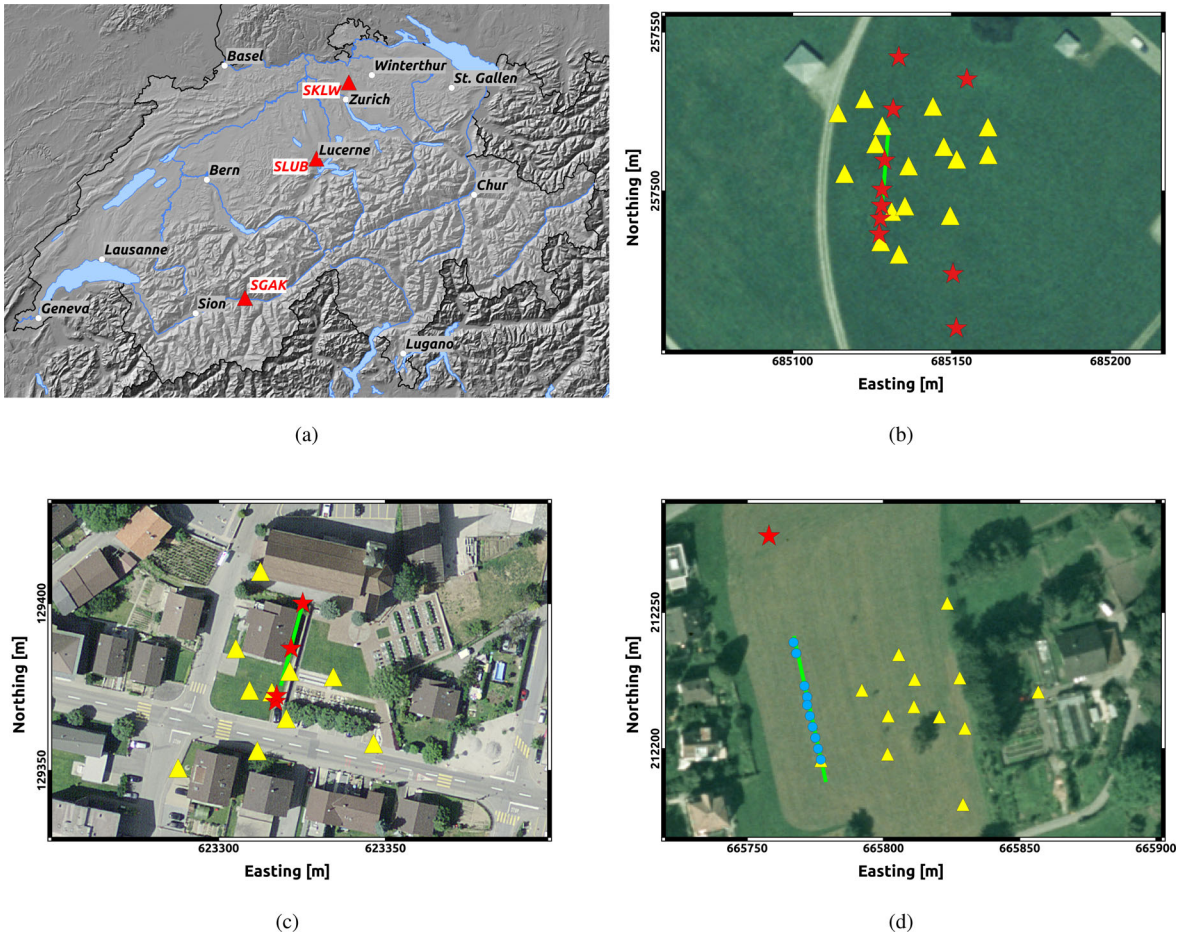


Figure 5. (a) Location of the sites considered within Switzerland. Red triangles mark the sites presented in this manuscript. (b) Array layout of Zurich Airport (SKLW). Yellow triangles show the locations of the sensors of the 2-D array. The green line depicts the line of geophones. The red stars depict the shooting locations. (c) Array layout of Gampel (SGAK). (d) Array layout of Bramberg (SLUB). Coordinates use the Swiss coordinate system CH1903. Reproduced by permission of swisstopo (JA100120).

The H/V ratios at each site are shown in Fig. 6(d). Each light red line depicts the H/V ratio computed at each station of the array. The darker red line depicts the H/V ratio averaged over all stations. These plots enable us to evaluate qualitatively the variability of the subsurface at each site. H/V peaks and troughs are also closely related to the ellipticity of the Rayleigh wave (Maranò *et al.* 2017).

In the following sections we show the results for the three sites considered. In the pictures, we depict with blue dots the estimates obtained from the passive survey, in red the estimates obtained from the Active 2D survey, in magenta the Active 1D, in black and green the estimates from FK analysis. Curves of the retrieved Rayleigh wave wavenumber and Rayleigh wave ellipticity angle are depicted. The Rayleigh wave wavenumber is expressed in $1/m$ and therefore is scaled of a 2π factor with respect to the wavenumber κ appearing in eq. (1). The Rayleigh wave ellipticity angle is the parameter ξ appearing in eq. (1).

Wavenumber and ellipticity angle curves depict the median of the estimated parameters. Uncertainty of the retrieved curves is described by the standard deviation of the estimates and it is depicted with error bars. For passive surveys, the standard deviation is computed over the entire recording. For active surveys, several curves are present. Each curve is representative of a distinct shooting location. The standard deviation is computed for the multiple shots at each location.

4.1 Zurich airport (SKLW)

The site SKLW in Kloten, Canton of Zurich is located on a large glacial gravel terrace in a wide, but shallow basin. The array is located in close vicinity of the airport of Zurich.

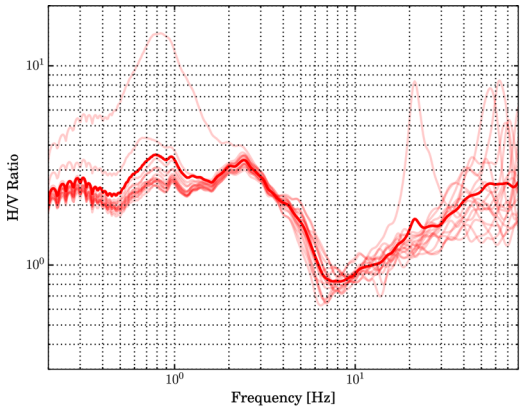
The array is composed of 16 sensors and has an aperture of 50 m, the array layout is shown in Fig. 5(b). The array layout is designed to reduce the occurrence of outliers following the approach presented in Maranò *et al.* (2014).

For the passive experiment, the duration of the recording is 135 min. More details on the analysis of this site are provided in Maranò *et al.* (2017).

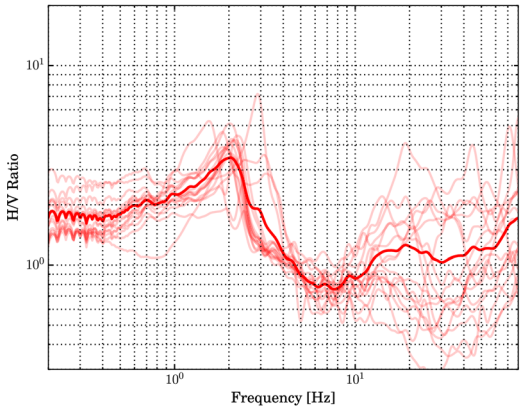
For the Active 2D experiment, sensors are left in their original position, thus the same 2-D array is used. Ten shooting locations are chosen within and around the array (Fig. 5b). Seismic energy is generated by hitting a metallic plate with a sledgehammer. Hammering is repeated about 10 times at each shooting location. For the Active 1D experiment and the FK comparison the geophone line depicted in Fig. 5(b) with the green solid line is used.

The H/V curve of Fig. 6(a) shows a peak at 2.45 Hz due to the sediment-rock interface.

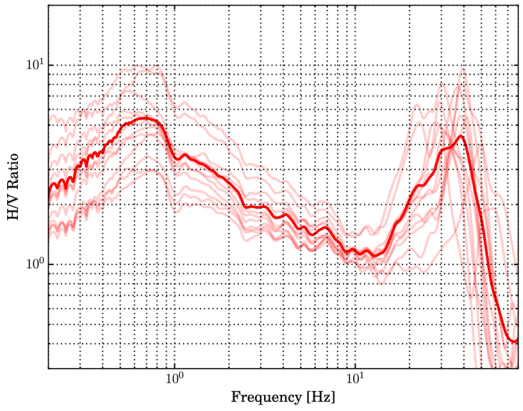
In Fig. 7(a) the retrieved Rayleigh wave wavenumber is shown, Fig. 7(c) depicts the estimated Rayleigh wave velocity. Multiples modes of propagation are retrieved. The results of the passive analysis refer to the whole recording. Concerning active measurements,



(a)



(b)



(c)

Figure 6. H/V ratios computed at different sites. (a) Zurich Airport (SKLW). (b) Gampel (SGAK). (c) Bramberg (SLUB).

the many estimates refer to several shooting locations and processing techniques.

In Fig. 7(c), we observe that the fundamental mode extends from below 5 Hz until 90 Hz and is retrieved by several processing methods. At low frequency, the fundamental mode is retrieved solely with the passive analysis (blue). From around 10 Hz and higher frequencies all the methods considered retrieve the fundamental mode and the estimated values are in good agreement among the different methods and shooting locations. Above 50 Hz only the Active 1D and the FK methods are able to retrieve the fundamental mode, possibly due to the higher sampling frequency.

A higher mode, extending from 25 to 100 Hz, is also found using different techniques. The estimated wavenumbers estimated with the different techniques are mostly consistent. Additional higher modes are found with the Active 2D analysis between 40 and 60 Hz.

In Fig. 7(c), we observe that the FK Vertical and the FK Longitudinal techniques over and underestimate the velocity of the Rayleigh wave at certain frequency ranges. This is observed for the fundamental mode around 20 Hz and for the higher mode just above 60 Hz. The Active 1D method, which is relying on exactly the same data, appears to follow more reliably the velocity of the mode. The robustness of the proposed method can be explained by the joint use of the three components.

Fig. 7(b) depicts the retrieved Rayleigh wave ellipticity angle of the fundamental mode. There is a very good agreement between the ellipticity angle curve retrieved in the Passive analysis (blue), the Active 2D (red) and Active 1D (magenta) analyses. Between 2 and 3 Hz, there is a singularity where the particle motion is horizontally polarized, this is visible from the ellipticity curve (i.e. $\xi = \pi/2$) and the peak of the H/V ratio, *cf.* Fig. 6(a). The particle motion of the fundamental mode is prograde until 6 Hz (i.e. $\xi > 0$) and it is retrograde at higher frequencies (i.e. $\xi < 0$). The retrieved ellipticity angle shows a retrograde particle motion at high frequencies, with the ellipticity angle becoming smaller and reflecting a polarization more horizontal.

Fig. 7(b) depicts the ellipticity angle of the higher mode, retrieved with the Active 2D and Active 1D analyses. The particle motion is prograde over all the observed frequency range.

4.2 Gampel (SGAK)

The site SGAK in Gampel, Canton of Valais is located on creek sediments on the northern edge of the deep sedimentary plain of the Rhône river.

Concerning the passive survey, ambient vibrations are recorded by a 2-D array of 16 sensors with an aperture of 300 m. Fig. 5(c) depicts the inner part of the array. Four shooting locations are depicted with a red star and the line of geophones with the solid green line. For the passive experiment, the array records ambient vibrations for 185 min.

The H/V curve of Fig. 6(b) shows a peak at 2 Hz due to the sediment-rock interface. We observe the great variability of the curves above 10 Hz, reflecting shallow heterogeneities at the site.

Fig. 8(a) depicts the wavenumber estimated using different methods. Fig. 8(b) depicts the corresponding velocity estimates. The fundamental mode is retrieved correctly by all methods until 25 Hz. In particular, there is a good agreement between the passive and the active results.

The ellipticity angle of the fundamental mode is shown in Fig. 8(c). In the frequency range accessible by the active measurements, the particle motion is retrograde as indicated by the negative ellipticity angle. The passive method shows a possible change to prograde particle motion below 4 Hz, but this result does not seem reliable with respect to the H/V measurement (*cf.* Fig. 6b). First, the H/V peak is around 2 Hz, that is, at a significantly lower frequency. Second, if there is a real change of the sense of particle motion, there has to be a singular ellipticity peak (with ellipticity angle $\pm\pi/2$) at a lower frequency and a singular ellipticity trough (with ellipticity angle 0) at a higher frequency. This is the opposite of what the passive array measurement seems to indicate in this

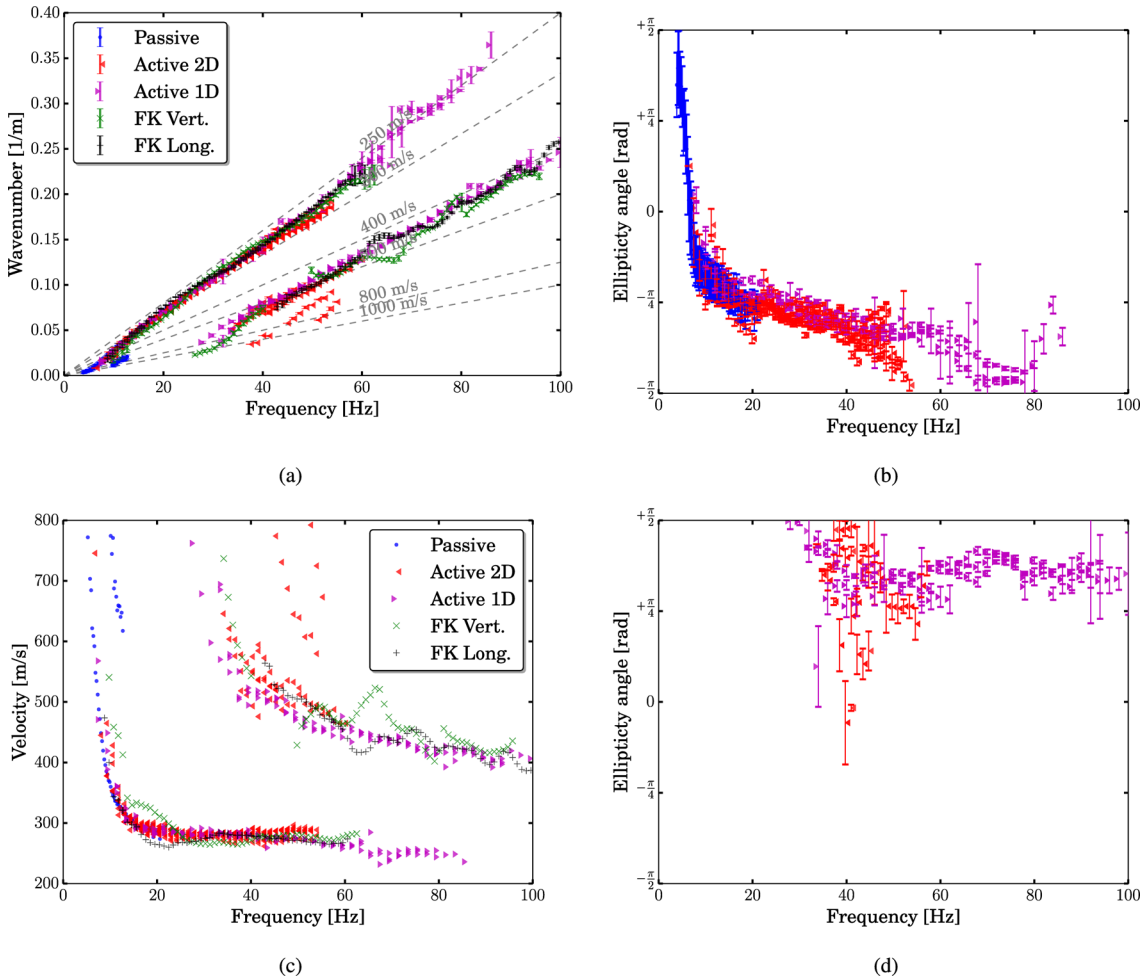


Figure 7. Zurich Airport (SKLW). (a) Rayleigh wave wavenumber. (b) Rayleigh wave ellipticity angle of the fundamental mode. (c) Rayleigh wave velocity. (d) Rayleigh wave ellipticity angle of the higher mode.

case. Third, further investigations show that the H/V peak is likely to be a non-singular one, that is, that the velocity contrast is not high enough to produce a singularity, even if the H/V peak is well pronounced. Therefore, we conclude that the apparent change of particle motion of the passive measurement is a misestimation at the border of the array resolution.

Fig. 8(a) also shows wavenumber estimates belonging to higher modes of propagation. There is some degree of agreement among the active modes used, but an exact mode addressing is not straightforward. Such difficulty at high frequency may be due to the heterogeneities of the site, also visible from the H/V curve.

4.3 Bramberg (SLUB)

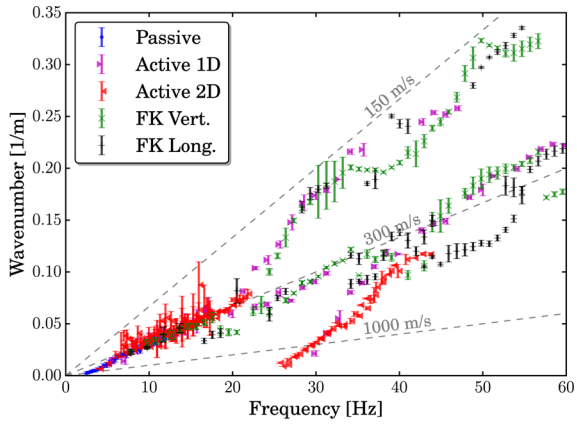
The site SLUB in Bramberg, Canton of Lucerne is located on a molassic hill. Previous site characterization found that the site is located on a layer of sandstone over a rock layer. The estimated V_{s30} is close to 1100 m s^{-1} (Michel *et al.* 2014).

Concerning the passive survey, ambient vibrations are recorded by a 2-D array of 13 sensors with an aperture of 84 m. The layout of the array is depicted with yellow triangles in Fig. 5(d). The duration of the recording is 98 min.

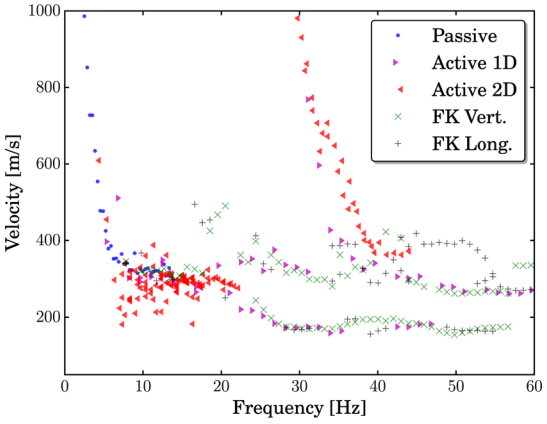
For the active survey, 10 triaxial velocimeters are arranged on a straight line. Sensors are spaced 4 m apart, with a wider gap in correspondence of two malfunctioning sensors. The layout is shown in Fig. 5(d) with blue circles. A line of vertical-component geophones is also placed next to the velocimeters and is depicted with a solid green line. Seismic energy is generated by dropping a 120 kg mass from a height of 1.2 m. The mass is dropped at a location 42 m away from the first sensor, aligned with the line of sensors. The mass is dropped 20 times and each shot is analysed independently.

The H/V curve of Fig. 6(c) shows a first peak around 0.6 Hz due to the rock-rock interface. This singularity is not resolved neither with the passive or the active survey. Around 40 Hz a peak due to the sediment-rock interface is visible.

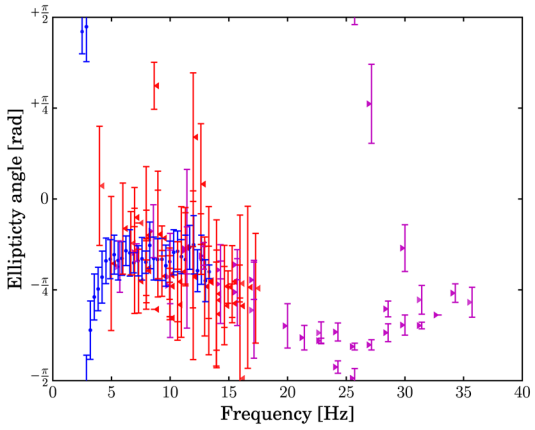
In Fig. 9(a), the retrieved Rayleigh wave wavenumber is shown. Fig. 9(b) depicts the corresponding velocity estimates. In blue, the estimated wavenumber obtained from the passive survey. The blue dots are representative of the results of the processing of the entire recording. In magenta, the estimated wavenumber obtained from the active survey using the method proposed in this article and the line of triaxial sensors, that is, Active 1D. The magenta dots are representative of the results of processing the 20 shots. We observe that there is a good agreement in the wavenumber estimated in the passive and in the active surveys for the fundamental mode.



(a)

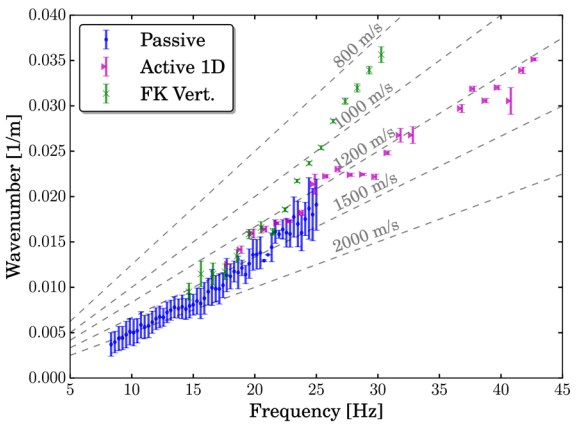


(b)

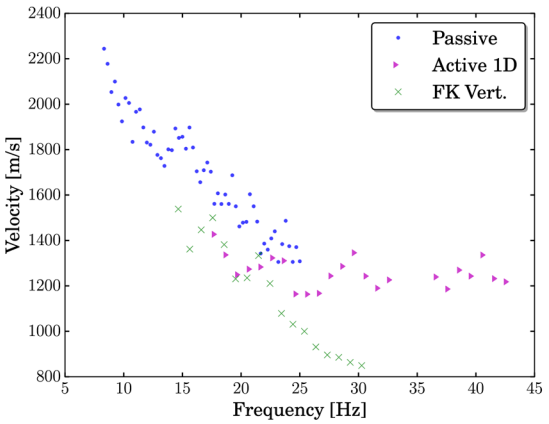


(c)

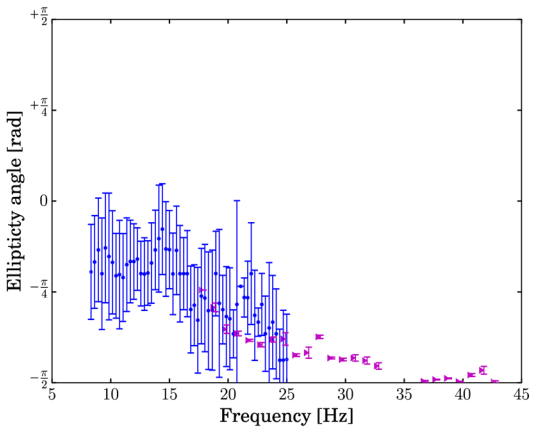
Figure 8. Gampel (SGAK). (a) Rayleigh wave wavenumber. (b) Rayleigh wave velocity. (c) Rayleigh wave ellipticity angle.



(a)



(b)



(c)

Figure 9. Bramberg (SLUB). (a) Rayleigh wave wavenumber. (a) Rayleigh wave velocity. (c) Rayleigh wave ellipticity angle.

In the active survey, it is not possible to identify at frequencies above 31 Hz the fundamental mode. The H/V peak around 30 Hz, see Fig. 6(c), shows significant variability. This variability reflects the heterogeneities of the most shallow layer and explains why the fundamental mode does not propagate at these frequencies.

Fig. 9(c) depicts the retrieved Rayleigh wave ellipticity angle. Again we observe that there is a good agreement of the results of the passive and active survey. Particle motion of the Rayleigh wave is retrograde (i.e. $\xi < 0$) for frequencies below 33 Hz, which has to be expected for rock sites. Between 37 and 42 Hz the particle

motion is horizontally polarized. This reflects the presence of a shallow layer of soft material and is also visible on the H/V peak of Fig. 6(c).

In Fig. 9(a) is also visible the wavenumber estimated by processing the geophone line with FK analysis. At 25 Hz the retrieved wavenumber begins to significantly deviate from the curve retrieved using the Active 1D analysis. We observe that at these frequencies, the Rayleigh wave fundamental mode is almost horizontally polarized, *cf.* Fig. 9(c). Therefore little energy is measured with vertical-component geophones and the analysis produces an incorrect result while the three-components analysis is able to capture all the energy of the Rayleigh wave.

5 DISCUSSION AND FINAL REMARKS

In this paper, we proposed a method for the analysis of Rayleigh waves generated at a nearby source relying on measurements from an array with arbitrary geometry. In such a scenario it is necessary to model the circular wavefront and consider the amplitude decay due to geometrical spreading. In the proposed approach, we use the model derived in Saito (1967) and account exactly for the geometrical spreading as described by the Hankel functions. Our approach follows an ML criterion and models all the measurements and all the wave parameters jointly.

The proposed method provides ML estimates of wave parameters including Rayleigh wave wavenumber and Rayleigh wave ellipticity angle. It is therefore possible to retrieve wavenumber and ellipticity angle curves. To the best of our knowledge, there are no methods for the retrieval of Rayleigh wave ellipticity from active surveys available in literature.

Using the proposed method, it is possible to conveniently combine passive and active surveys. This enables us to enlarge the analysable frequency range and the depths investigated to better constrain the inverted model near the surface. We emphasize that our method enables us to perform extremely time efficient surveys. In fact, we suggest to use the same hardware and array layout for both the passive and the active survey. This approach greatly reduces the logistic effort needed in the organization of the survey.

We have demonstrated the applicability of our method at three different sites in Switzerland. At each site we performed a passive survey measuring the ambient vibration wavefield and an active survey using a controlled source at known locations. We successfully retrieved Rayleigh wave wavenumber and Rayleigh wave ellipticity angle curves for both the fundamental and higher modes using the measurements from both types of survey. We showed that the joint analysis of passive and active surveys yields the parameters of interest over a broad range of frequencies. As expected, passive measurements allow us to investigate lower frequencies and active measurements to investigate higher frequencies. At the considered sites we always found a good agreement between the results from the two surveys.

We compared the dispersion curve retrieved with the proposed method and with a common processing method for active data, FK analysis. We observed that, by modelling the three sensor components jointly, the proposed method is more robust than the single-component FK method.

Future work will include the extension of our methodology to the analysis of Love waves. The dispersion curve of Love waves is an additional observable that can carry important information on the structure of the subsoil. In an active survey, in order to generate Love sources, it is necessary to employ an appropriate source able to excite *SH* waves. Concerning the processing it is necessary to adapt the model proposed in this paper to account for the polarization of Love waves and include a model for the radiation pattern.

The software developed in this work is available from the authors upon request.

ACKNOWLEDGEMENTS

This work was supported by the Swiss National Science Foundation project 'Advanced Single Station and Array Methods for the Analysis of Ambient Vibrations and Earthquake Recording' (200021_153633) and the Swiss strong motion renewal project funded by the Federal Office for the Environment (FOEN). The authors thank Dario Chieppa, Marthe Faber, David Farsky, Lea Kiefer, Clotaire Michel and Valerio Poggi for their help during the measurements.

REFERENCES

Abramowitz, M. & Stegun, I., 1964. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables,* Dover Publications.

Almendros, J., Ibáñez, J.M., Alguacil, G. & Pezzo, E.D., 1999. Array analysis using circular-wave-front geometry: an application to locate the nearby seismo-volcanic source, *Geophys. J. Int.,* **136**(1), 159–170.

Boiero, D., Bergamo, P., Bruno Rege, R. & Socco, L.V., 2011. Estimating surface-wave dispersion curves from 3D seismic acquisition schemes: Part 1 - 1D models, *Geophysics,* **76**(6), G85–G93.

Bonnefoy-Claudet, S., Cotton, F. & Bard, P.-Y., 2006. The nature of noise wavefield and its applications for site effects studies: a literature review, *Earth-Sci. Rev.,* **79**(3–4), 205–227.

Byrd, R.H., Lu, P., Nocedal, J. & Zhu, C., 1995. A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.,* **16**(5), 1190–1208.

Capon, J., 1969. High-resolution frequency-wavenumber spectrum analysis, *Proc. IEEE,* **57**(8), 1408–1418.

Krim, H. & Viberg, M., 1996. Two decades of array signal processing research: the parametric approach, *IEEE Signal Process. Mag.,* **13**(4), 67–94.

Kschischang, F.R., Frey, B.J. & Loeliger, H.-A., 2001. Factor graphs and the sum-product algorithm, *IEEE Trans. Inf. Theory,* **47**(2), 498–519.

Loeliger, H.-A., Dauwels, J., Hu, J., Korl, S., Ping, L. & Kschischang, F.R., 2007. The factor graph approach to model-based signal processing, *Proc. IEEE,* **95**(6), 1295–1322.

Maranò, S., Reller, C., Loeliger, H.-A. & Fäh, D., 2012. Seismic waves estimation and wavefield decomposition: application to ambient vibrations, *Geophys. J. Int.,* **191**(1), 175–188.

Maranò, S., Fäh, D. & Lu, Y.M., 2014. Sensor placement for the analysis of seismic surface waves: sources of error, design criterion and array design algorithms, *Geophys. J. Int.,* **197**(3), 1566–1581.

Maranò, S., Fäh, D. & Loeliger, H.-A., 2015. A state-space approach for the analysis of wave and diffusion fields, in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp. 2564–2568, Brisbane, Australia.

Maranò, S., Hobiger, M. & Fäh, D., 2017. Retrieval of Rayleigh wave ellipticity from ambient vibration recordings, *Geophys. J. Int.,* **209**(1), 334–352.

Michel, C., Edwards, B., Poggi, V., Burjànek, J., Roten, D., Cauzzi, C. & Fäh, D., 2014. Assessment of site effects in Alpine regions through systematic site characterization of seismic stations, *Bull. seism. Soc. Am.,* **104**(6), 2809–2826.

Moro, G.D., Keller, L. & Poggi, V., 2015. A comprehensive seismic characterisation via multi-component analysis of active and passive data, *First Break,* **33**(9), 45–53.

Ohrnberger, M., Schissele, E., Cornou, C., Bonnefoy-Claudet, S., Wathelet, M., Savvaidis, A., Scherbaum, F. & Jongmans, D., 2004. Frequency wavenumber and spatial autocorrelation methods for dispersion curve determination from ambient vibration recordings, in *13th World Conference on Earthquake Engineering,* Vancouver, BC, Canada.

Okada, H., 1997. The *Microtremor Survey Method,* Society of Exploration Geophysicists.

Park, C., Miller, R., Ryden, N., Xia, J. & Ivanov, J., 2005. Combined use of active and passive surface waves, *J. Environ. Eng. Geophys.,* **10**(3), 323–334.

Park, C.B., Miller, R.D. & Xia, J., 1998. Imaging dispersion curves of surface waves on multi-channel record, in *SEG Expanded Abstracts,* vol. 17, pp. 1377–1380.

Park, C.B., Miller, R.D. & Xia, J., 1999. Multichannel analysis of surface waves, *Geophysics,* **64**(3), 800–808.

Poggi, V. & Fäh, D., 2010. Estimating Rayleigh wave particle motion from three-component array analysis of ambient vibrations, *Geophys. J. Int.,* **180**(1), 251–267.

Reller, C., Loeliger, H.-A. & Maranò, S., 2011. Multi-sensor estimation and detection of phase-locked sinusoids, in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp. 3872–3875, Prague, Czech Republic.

Richart, F.E., Hall, J.R. & Woods, R.D., 1970. *Vibrations of Soils and Foundations,* Prentice-Hall.

Saito, M., 1967. Excitation of free oscillations and surface waves by a point source in a vertically heterogeneous Earth, *J. geophys. Res.,* **72**(14), 3689–3699.

Schwarz, G.E., 1978. Estimating the dimension of a model, *Ann. Stat.,* **6**(2), 461–464.

Sheriff, R.E. & Geldart, L.P., 1995. *Exploration Seismology,* Cambridge Univ. Press.

Socco, L.V., Foti, S. & Boiero, D., 2010. Surface-wave analysis for building near-surface velocity models – established approaches and new perspectives, *Geophysics,* **75**(5), 75 A83–75 A102.

Stokoe, K.H., Wright, S.G., Bay, J.A. & Roesset, J.M., 1994. Characterization of geotechnical sites by SASW method, in *Geophysical Characterization of Sites,* pp. 15–25, ed. Woods, R.D., Oxford Publ.

Strobbia, C. & Foti, S., 2006. Multi-offset phase analysis of surface wave data (MOPA), *J. Appl. Geophys.,* **59**(4), 300–313.

Tokimatsu, K., Shinzawa, K. & Kuwayama, S., 1992. Use of short-period microtremors for v_s profiling, *J. Geotech. Eng.,* **118**(10), 1544–1558.

Van Trees, H.L., 2002. *Optimum Array Processing,* John Wiley & Sons, Inc.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. A numerical comparison between the radial and the vertical components of a Rayleigh wave. (a) Amplitude ratio. (b) Phase difference.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.