Shear wave velocity model of the Abanico Formation underlying the Santiago City metropolitan area, Chile, using ambient seismic noise tomography

J. Salomón,[1] C. Pastén,[1] S. Ruiz,[2] F. Leyton,[3] M. Sáez[2] and R. Rauld[4]

[1]*Department of Civil Engineering, University of Chile, Av. Blanco Encalada* 2002, *Of.* 431, *Santiago* 8370449, *Chile. E-mail: cpasten@ing.uchile.cl*
[2]*Department of Geophysics, University of Chile, Av. Blanco Encalada* 2002, *Santiago* 8370449, *Chile*
[3]*National Seismological Center, University of Chile, Av. Blanco Encalada* 2002, *Santiago* 8370449, *Chile*
[4]*XTERRAE Geología, Av. Hernando de Aguirre* 194, *Providencia, Santiago, Chile*

SUMMARY

The seismic response of the Santiago City, the capital of Chile with more than 5.5 million inhabitants, is controlled by the properties of the shallower quaternary deposits and the impedance contrast with the underlying Abanico Formation, among other factors. In this study, we process continuous records of ambient seismic noise to perform an ambient seismic noise tomography with the aim of defining the shallower structure of the Abanico Formation underneath the densely populated metropolitan area of Santiago, Chile. The seismic signals were recorded by a network consisting of 29 broad-band seismological stations and 12 accelerograph stations, located in a 35×35 km^2 quadrant. We used the average coherency of the vertical components to calculate dispersion curves from 0.1 to 5 Hz and Bootstrap resampling to estimate the variance of the traveltimes. The reliable frequency band of the dispersion curves was defined by an empirical method based on sign normalization of the coherency real part. The ambient noise tomography was solved on a domain discretized into 256 2×2 km^2 cells. Using a regularized weighted least-squares inversion, we inverted the observed traveltimes between stations, assuming straight ray paths, in order to obtain 2-D phase velocity maps from 0.2 to 1.1 Hz, linearly spaced every 0.05 Hz, in 157 of the 256 square cells of the domain. In each square cell with information, dispersion curves were assembled and used to invert shear wave velocity profiles, which were interpolated using the ordinary Kriging method to obtain a 3-D shear wave velocity model valid from 0.6 to 5 km depth. The 3-D velocity model shows that the Abanico Formation is stiffer in the south of the study area with larger velocity anomalies towards the shallower part of the model. The value of the shear wave velocity narrows with depth, reaching an average value of 3.5 km s^{-1} from 3 to 5 km depth.

Key words: tomography; Seismic noise; basin; surface waves; s waves; traveltime.

INTRODUCTION

2-D and 3-D basin effects, such as lateral and vertical material heterogeneity, variable bedrock depth and basin-edge effects, can aggravate seismic site response and increase ground motion amplification during earthquakes (Kawase 1996; Semblat *et al.* 2005; Makra & Chávez-García 2016; Moczo *et al.* 2018). Understanding the seismic response of sedimentary basins for seismic hazard evaluation of urban areas requires detailed wave velocity models of the shallower quaternary sediments and the deeper geological structures.

In recent decades, several earthquakes have struck the city of Santiago, the economic and political capital of Chile with more than 5.5 million inhabitants (Fig. 1). Significant damage and large MSK intensities in the city were caused by the 1985 M_w 8.0 Valparaiso and the 2010 M_w 8.8 Maule earthquakes (Astroza *et al.* 2012; Ruiz & Madariaga 2018). The Santiago city is located in a basin in the Central Depression of Central Chile at about 500–700 m above the mean sea level between the Coastal Range and the Principal Andean Cordillera (33°30'S, 70°40'W). The sedimentary cover of the basin is relatively shallow and flat with 250 m of average depth and depocentres that reach about 600 m (Yañez *et al.* 2015; González *et al.* 2018). Pilz *et al.* (2010) proposed a shear wave velocity (*Vs*) model for a reduced 26 × 12 km^2 area to a depth of 600 m in the central part of the city, by inverting *H/V* spectral ratios of microtremor recordings, constrained with additional geological and geophysical data.
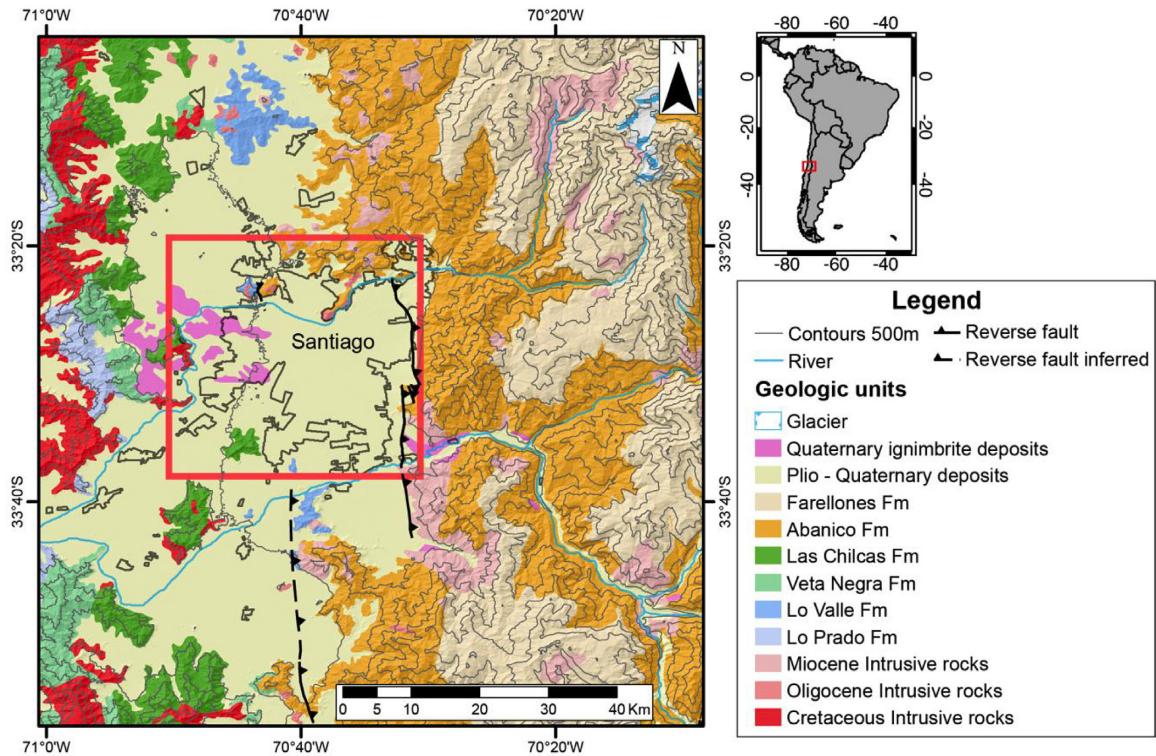
Figure 1. Simplified geological model in the vicinity of the Santiago Metropolitan Area. The red square refers to the study area where the ambient noise tomography applies. The Abanico Formation underlies the plio-quaternary deposits and the quaternary ignimbrite deposits. The reverse fault at the east of the study area is the San Ramón Fault. The upper right-hand side panel shows the location of the Santiago Basin in South America.

On the other hand, the deep structure of the basin consists of the Abanico, Lo Valle/Las Chilcas and the Veta Negra/Lo Prado Formations (Fig. 1). Farías *et al.* (2010) defined a crustal scale model of Central Chile based on seismicity and surface geology and estimated that the shear wave velocity of the upper 10 km varies from 3 to 3.5 km s^{-1}. This model is consistent with 1-D velocity models proposed for part of Central Chile (Godoy *et al.* 1999; Barrientos *et al.* 2004). Nevertheless, the resolution of these models does not allow estimating the shallower structure of the Abanico Formation that is in contact with the quaternary sediments. An accurate seismic response evaluation of the Santiago City Metropolitan Area requires an improved *Vs* model that overcomes the limitations of the existing shallow and deep *Vs* models.

The use of ambient noise cross-correlation has shown to be a promising method to define the deep structure of sedimentary basins at urban scales (e.g. Pastén *et al.* 2016; Inzunza *et al.* 2019; Vassallo *et al.* 2019). The methodology is based on the idea that the cross-correlation of long time-series of seismic noise converges to the Green's function between two sensors, which allows estimating the dispersive characteristics of the medium (i.e. surface wave dispersion curves, e.g. Sabra *et al.* 2005; Shapiro *et al.* 2005; Sánchez-Sesma & Campillo 2006). The cross-correlation of ambient seismic noise has been used to generate velocity maps in different scales and frequency ranges, a technique known as ambient noise tomography (ANT, e.g. Bensen *et al.* 2007; Ward *et al.* 2013).

Pastén *et al.* (2016) estimated shear wave velocity profiles of the Santiago Basin up to 5 km depth using frequency- and time-domain ambient noise cross-correlation methods, modified for urban areas of tens of square kilometres. The frequency-domain method estimated the average coherency, whereas the time-domain method estimated noise correlation functions (NCF), or empirical Green's

functions (EGF), of the vertical components of ambient noise recorded with a temporal network of broad-band seismometers. Combining the data from the temporal network of broad-band seismometers and the methods developed in Pastén *et al.* (2016), as well as ambient noise records from accelerographs and the ambient seismic noise tomography method, the study presented herein develops a 3-D shear wave velocity model of the Abanico Formation, underlying the quaternary deposits of the Santiago City Metropolitan Area.

TECTONIC AND GEOLOGICAL SETTING

The tectonic evolution of Central Chile is controlled by the subduction of the oceanic Nazca Plate beneath the continental South American Plate, at least since the Jurassic (Mpodozis & Ramos 1990). The study area is located west of the Andes Cordillera (Fig. 1), where the foreland consists of: (1) the Chilean Central Depression, whose basement is composed by Eocene to Early Miocene volcano-sedimentary rocks, later filled with Pleistocene–Holocene alluvial, fluvial and pyroclastic deposits (Sellés & Gana 2001) at approximately 500 m above the mean sea level; (2) the Coastal Cordillera, constituted of Paleozoic and Mesozoic rocks with altitudes commonly below 2000 m and (3) the offshore continental margin in front of the Chile trench (Riesner *et al.* 2018).

The late Eocene to early Miocene Abanico Formation (Fig. 1) consists of volcaniclastic rocks, tuffs, basic lavas and ignimbrites, as well as interbedded alluvial, fluvial and lacustrine sediments (Charrier *et al.* 2005). Dates of stratified pile of volcanic rocks of the Abanico Formation close to Santiago range from 30.9 to 20.3 Ma (Nyström *et al.* 2003; Vergara *et al.* 2004).
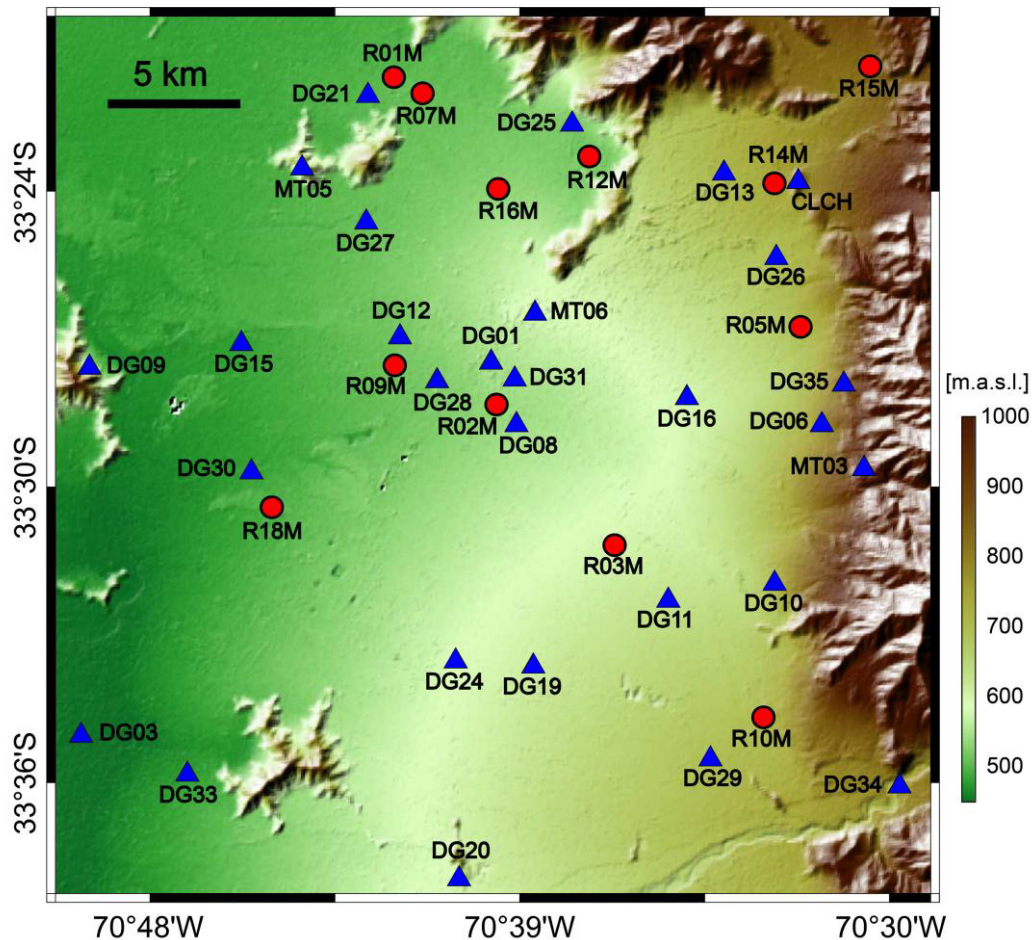
Figure 2. Seismic stations considered in the study area. Blue triangles indicate the location of the broad-band stations whereas red dots represent the location of the accelerographs.

Underlying the Abanico Formation are the Upper Cretaceous volcanic Lo Valle Formation, and the Lower Cretaceous volcanic and sedimentary Las Chilcas, Veta Negra and Lo Prado Formations (Armijo *et al.* 2010). In the western Andean Cordillera front, the Abanico Formation is overlain by the Miocene volcanic Farellones Formation (Charrier *et al.* 2002; Armijo *et al.* 2010).

The kinematics and crustal structure underneath the Santiago Basin have been addressed by numerous studies (Charrier *et al.* 2002; Armijo *et al.* 2010; Farías *et al.* 2010; Giambiagi *et al.* 2014; Muñoz-Saez *et al.* 2014; Riesner *et al.* 2017, 2018). The Central Depression has been interpreted as a basin resulting from the over-thrust of the Principal Cordillera over the relatively rigid western foreland due to the westward propagation of a contractional structural system defined as the West Andean Thrust (WAT). The San Ramon Fault (Fig. 1) is the main WAT structure recognized at the latitude of Santiago and is the limit between those two structural units (Armijo *et al.* 2010; Vargas *et al.* 2014; Ammirati *et al.* 2019).

DATA PROCESSING

Seismic network

Ambient seismic noise was recorded with two seismic networks deployed over the Santiago Basin. The first seismic network, shown

in Fig. 2, consisted of 29 broad-band seismological stations provided by the National Seismological Center of the University of Chile (CSN, Centro Sismológico Nacional). This network recorded months of ambient seismic noise between July 2013 and May 2014 at a sampling rate of 100 Hz. The broad-band stations were equipped with a Nanometrics Trillium Compact 120 s sensor, a 24-bit Quanterra Q330 digitizer, a Marmot data logger and a GPS antenna. The second seismic network consisted of 12 accelerographs, sampling at 200 Hz, that belongs to the Chilean National Office for Emergency (ONEMI, Oficina Nacional de Emergencia del Ministerio del Interior y Seguridad Pública). The accelerograph stations were equipped with a Kinemetrics Basalt sensor, with an EpiSensor Force Balance triaxial accelerometer and a 24-bit digitizer. The location of the seismic stations and the recording periods used in this study are detailed in Table S1.

Phase velocity estimation

In order to extract EGF between receivers and estimate Rayleigh wave dispersion curves, we calculated the coherency of the vertical component of the ambient seismic noise records. The zero-crossings of the real part of the coherency can be used to estimate the phase velocity (Ekström 2014). Based on the spectral method described in Pastén *et al.* (2016), we followed these steps:

(1) The instrumental response was removed from the raw data applying a spectral deconvolution with the corresponding transfer function of the sensors (Jin & Gaherty 2015).

(2) The records were visually inspected and high-pass filtered with a 4th-order Butterworth filter of 0.01 Hz cutoff frequency, and divided into 2-min windows. Longer time windows (e.g. 360, 180, 60 and 10 min) reduce the stacking level required to improve the resolution of the zero crossings in the average coherency, particularly at high frequencies for typical interstation distances. This is relevant in cases of limited available data; typically, pairs of stations with less than 3 weeks of simultaneous noise recording. Since this study focuses on characterizing the transition between the upper continental crust structure and the shallower sediments, 2-min windows provided a balance between the resolution of the average coherency at high frequencies and a reasonable computational cost. Following the methodology reported in Ekström (2014), we did not pre-condition the signals (i.e. filtering and whitening) because the main factor controlling the calculation process is the stacking level. When a time-domain method is preferred over a spectral approach, stacking methods, such as the one proposed by Shirzad & Shomali (2013), can be used to improve the resolution of the extracted EGF.

(3) The coherency of the records at stations a and b was calculated in each of the k 2-min windows as

$$\gamma_{abk}(\omega) = \frac{u_{ak}(\omega) \cdot u_{bk}^*(\omega)}{\sqrt{u_{ak}(\omega) \cdot u_{ak}^*(\omega)} \cdot \sqrt{u_{bk}(\omega) \cdot u_{bk}^*(\omega)}}, \tag{1}$$

where $u_{ik}(\omega)$ is the spectrum of the record at station i in the time-window k and the asterisk represents the complex conjugate.

(4) The coherencies calculated in a day were stacked and the real part was normalized by its maximum absolute value. This normalization intends to create a map (Figs 3a and b) where daily variations of the coherency can be identified.

(5) The average coherency $\bar{\gamma}_{ab}$ was calculated from the daily normalized real part of the coherencies (Figs 3c and d).

(6) The phase velocity $c(f_n)$ was calculated by identifying the zero-crossing of the average coherency (Figs 3c and d). According to Ekström *et al.* (2009), if f_n denotes the frequency of the nth observed zero crossing in the average coherency and z_n refers to the nth zero of the zero-order Bessel function of the first kind J_0, the corresponding phase velocity can be determined as (Aki 1957)

$$c(f_n) = \frac{2\pi f_n \Delta}{z_n}, \tag{2}$$

where Δ is the interstation distance. However, association of a given zero crossing in the average coherency with a particular zero crossing of J_0 may be difficult because noise in the average coherency can cause missed or extra zero crossings. To account for this effect, a set of phase-velocity estimates are calculated as

$$c_m(f_n) = \frac{2\pi f_n \Delta}{z_{n+m}}, \tag{3}$$

where $m = 0, \pm1, \pm2, \pm3, \ldots$ indicates the missing or extra zero crossings of the average coherency. We chose the m-value such that the resulting dispersion curve $c(f_n)$ falls within a realistic range of the expected phase velocities (Pastén *et al.* 2016).

(7) The variance of the traveltime was estimated by applying a Bootstrap resampling method over the calculated average coherency, following Liu *et al.* (2016). For each pair of stations, we generated 1000 new stacks of daily average coherencies with replacement. The maximum and minimum number of simultaneous recorded days between the station pairs were 180 and 15, respectively. Then, the phase velocity was calculated for each resampled

stack and the standard deviation of the phase velocity $\sigma_{cb}(\omega)$ was calculated for each frequency. Using error propagation, the standard deviation of the traveltime can be calculated as a function of the phase velocity obtained from eq. (3) as

$$\sigma_t(\omega) = \frac{\Delta}{[c_m(\omega)]^2} \sigma_{cb}(\omega). \tag{4}$$

An application of this methodology is illustrated in Fig. 3. The processed station pairs correspond to DG19–DG24 (left-hand side panels) and DG09-MT03 (right-hand side panels), with interstation distances of 2.95 and 29.42 km, respectively. Figs 3(a) and (b) show the normalized real part of the coherency for each one of the recorded days. Figs 3(c) and (d) display the average coherency and the identified zero crossings (indicated with yellow dots). Figs 3(e) and (f) show the estimated dispersion curves for different m-values. The chosen dispersion curve (black line) and its standard deviation (red dashed line) are presented in Figs 3(g) and (h).

Estimation of the reliable frequency band

The ANT is a method that enables estimating a 3-D shear wave velocity model by combining the information from a large number of dispersion curves. The resolution of the model strongly depends on the spatial distribution of stations (number and location of stations) and the frequency band in which the dispersion curves between each pair of stations can be resolved. Furthermore, the minimum f_{min} and maximum f_{max} valid observable frequencies between stations are mainly conditioned by the interstation distance and the number of stacked daily coherency functions. At short interstation distances, low frequency Rayleigh waves with large wavelengths cannot be reliably detected due to uncertainty in the phase estimation (Luo *et al.* 2015). On the other hand, geometric attenuation of high frequency waves hinders the detection of short wavelength waves at large interstation distances (Lai *et al.* 2002).

The spectral method proposed by Pastén *et al.* (2016) does not prescribe limits for the maximum and minimum reliable frequencies (f_{min} and f_{max}) in the dispersion curves. The large variability in interstation distances in the seismic network requires establishing an automatic method to determine a reliable frequency band $[f_{min}, f_{max}]$ of the average coherency that reduces the subjectivity of visual inspection. We implemented a systematic procedure to determine the temporal stability of the daily coherency functions at each frequency. This empirical method consists of the following steps:

(1) The real part of the daily coherency functions were *sign* normalized by assigning 1 to positive amplitudes and -1 to the negative amplitudes (Fig. 4a).

(2) The normalized daily coherency functions were averaged to obtain an average normalized coherency. If the normalized daily coherency functions were stable over time, the average normalized coherency would have amplitudes close to ±1.

(3) The standard deviations of the normalized daily coherency functions $\sigma_{snorm}(\omega)$ were calculated for each frequency (Fig. 4b). If the normalized daily coherency functions were stable over time, the standard deviations would be close to zero.

(4) The standard deviations were plotted as a function of frequency and the resulting curve was smoothed with a 0.2 Hz bandwidth moving average (Fig. 4b). We found that frequencies with standard deviations lower than $\sigma_{snorm}^{th} = 0.75$ were associated to unequivocal zero-crossings in the average coherency. Based on this
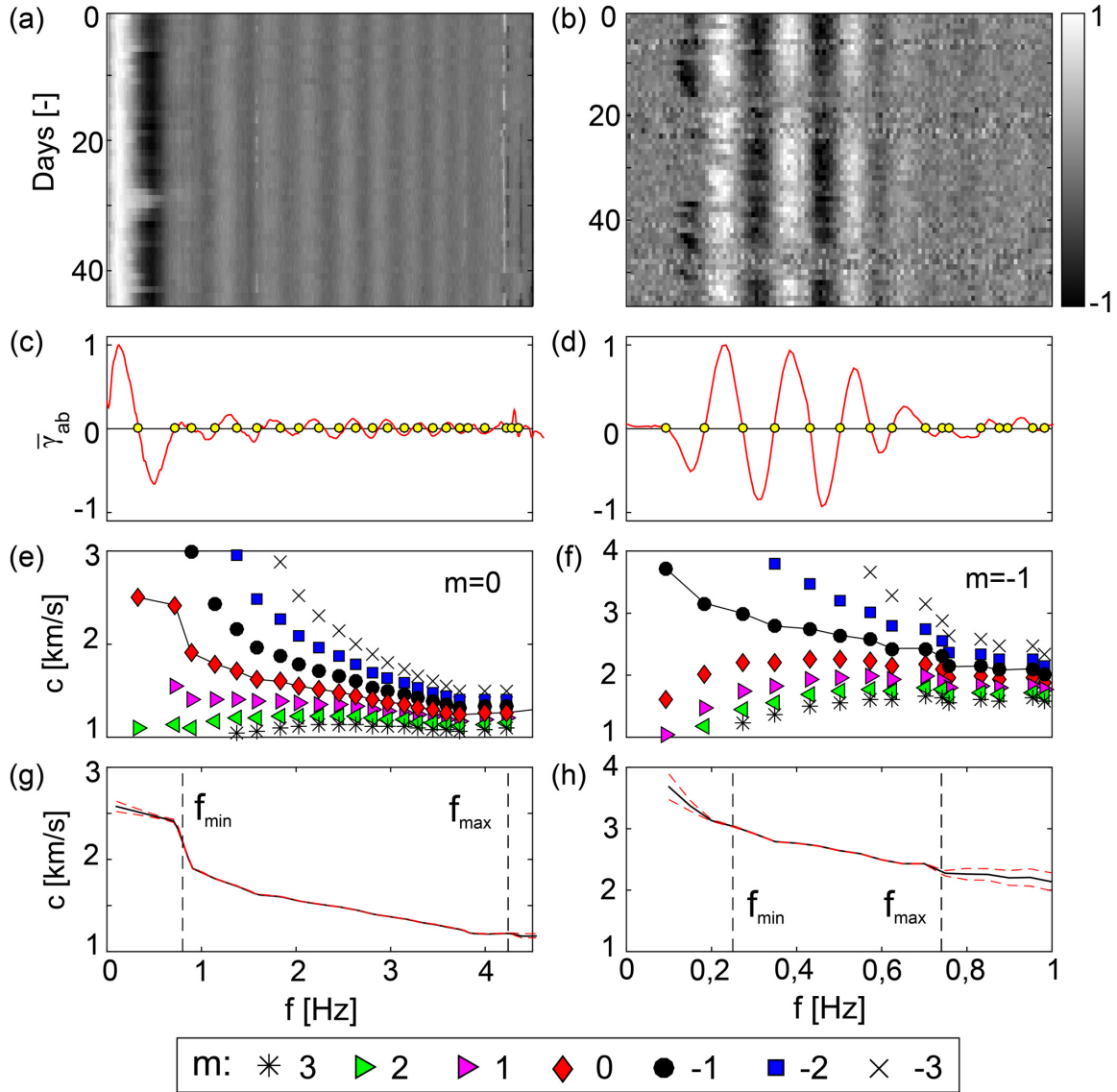
Figure 3. Examples of average coherency, dispersion curves and estimated reliable frequency band for the station pairs DG19-DG24 (plots in left-hand column, interstation distance of 2.95 km) and DG09-MT03 (plots in right-hand column, interstation distance of 29.42 km). Panels (a) and (b) show the real part of the coherency for every recorded day, (c) and (d) display the average coherency and the identified zero crossings (yellow dots), (e) and (f) show the estimated dispersion curves for different *m*-values (–3, –2, –1, 0, 1, 2, 3) and (g) and (h) the chosen dispersion curve (black line) and its standard deviation (red dashed line) calculated from Bootstrapping method.

threshold standard deviation $\sigma_{\text{snorm}}^{\text{th}}$, we defined the minimum f_σ^{min} and the maximum f_σ^{max} cut-off frequencies (Figs 4b and c).

Then, the minimum observable frequency was determined as:

$$f_{\text{min}} = \max\left(f_{\omega_0}, f_\lambda, f_\sigma^{\text{min}}\right) \qquad (5)$$

Where f_{ω_0} is the frequency of the first zero-crossing in the average coherency and f_λ is the frequency associated to a wavelength equal to the interstation distance (Luo *et al.* 2015), according to

$$f_\lambda = \frac{c\left(f_{\lambda=\Delta}\right)}{\Delta}. \qquad (6)$$

On the other hand, the maximum observable frequency was defined from the threshold standard deviation criterion $f_{\text{max}} = f_\sigma^{\text{max}}$ (Figs 4b and c).

Using all the available station pairs, we calculated more than 500 average coherencies and applied the reliable frequency band

method in order to identify valid range of frequencies for each dispersion curve. Fig. 5(a) shows the calculated dispersion curves after applying the reliable frequency band method and Figs 5(b) shows a histogram with the number of station pairs that have dispersive information as a function of frequency. Fig. 5(a) shows that the majority of dispersion curves were successfully calculated between 0.1 and 5 Hz, but a closer look at Fig. 5(b) indicates that most of the station pairs have dispersion curves with information below 1.1 Hz.

TRAVELTIME INVERSION

Inversion scheme and data validation

The traveltime of a wave passing between two stations can be expressed as the integral of the slowness field (inverse of velocity)
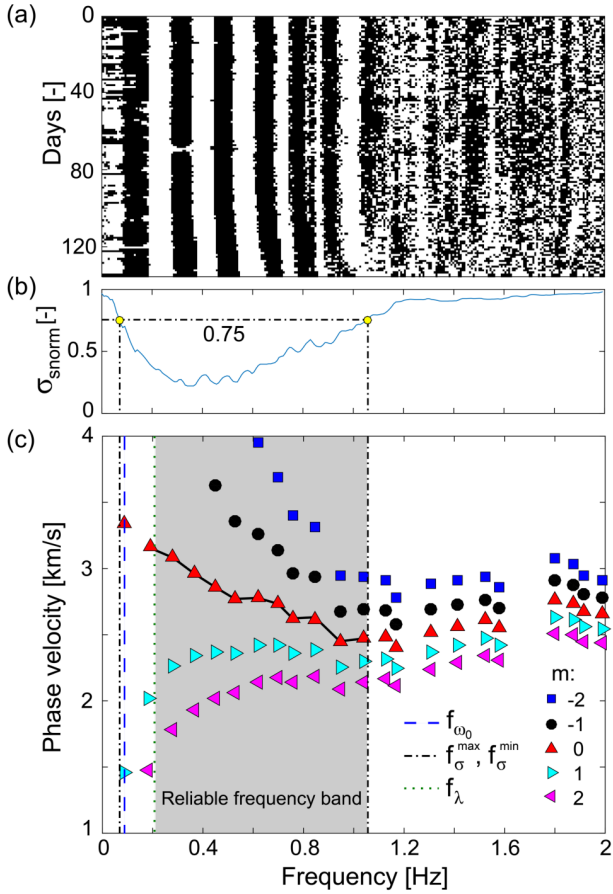
Figure 4. Implementation of the reliable frequency band method for the station pair DG03–DG20. (a) Sign-normalized daily coherency functions, where black is one and white is zero, (b) standard deviation as a function of frequency and (c) selected dispersion curve (red triangles, $m = 0$) defined in the reliable frequency band in gray.

throughout the ray path at a given frequency. The problem of deriving the velocity field from surface records is highly nonlinear due to the relationship between the ray path and the velocity structure of the medium (Nolet 2012). However, given the error levels associated to the data from shallow surveys, a linear approximation based on straight ray paths is sufficient to estimate a reliable slowness model in the presence of slight velocity anomalies (Kugler *et al.* 2007; Picozzi *et al.* 2009). Then, assuming straight rays, the observed traveltimes between receivers can be directly obtained by dividing the interstation distance in the phase velocity for a given frequency f, that is $t^{obs} = \Delta / c(f)$.

In order to obtain 2-D phase velocity maps, we inverted the M observed traveltimes $\underline{t}^{obs}$ [$M \times 1$] between stations. For discrete values of frequency between 0.1 and 5.0 Hz, we performed a regularized weighted least squares inversion (Menke 2018), solving

$$\begin{bmatrix} \underline{C}_\chi^{-1/2} \underline{G} \\ \varepsilon \underline{\nabla}^2 \end{bmatrix} [\underline{s}] = \begin{bmatrix} \underline{C}_\chi^{-1/2} \underline{t}^{obs} \\ \varepsilon \underline{\nabla}^2 \underline{s}_0 \end{bmatrix}, \tag{7}$$

where $\underline{s}$ [$N \times 1$] is the slowness vector of the N cells in which the domain is discretized, $\underline{\nabla}^2$ [$M \times N$] is a discrete Laplacian operator, ε [-] is the regularization parameter, $\underline{G}$ [$M \times N$] is a design matrix, $\underline{s}_0$ [$N \times 1$] is an *a priori* slowness model and $\underline{C}_\chi^{-1/2}$ [$M \times M$] is the Cholesky decomposition of the observational variance,

$$C_{\chi(i,i)}^{-1/2} = \frac{1}{\sigma_{ti}}, \tag{8}$$

where σ_{ti} (eq. 4) is the standard deviation of the ith traveltime observation ($i = 1, 2, \ldots, M$).

The *a priori* slowness model was obtained from the average phase velocity value for each frequency. The optimum regularization parameter ε was chosen using the generalized cross-validation method (Craven & Wahba 1978) and the L-curve criterion (Aster *et al.* 2018). The use of both methods balances the data fitting and the influence of *a priori* information on the estimated slowness model (Hansen 1992).

The grid size is usually defined as a function of the average station spacing (e.g. the grid size is about a third of the average station spacing). However, we adopted an alternative criterion based on the ray's wavelengths, defined by Pilz *et al.* (2010). The criterion considers that a straight ray path is deviated in less than a quarter of its wavelength in the presence of slight velocity anomalies. Thus, if the grid size was selected according to this criterion, any ray deviation would remain within the limits of the cell. This criterion attempts to minimize the errors propagated to the phase velocity maps due to the straight ray assumption. Considering this criterion, the inversion was performed over a grid of 256 2×2 km^2 cells, and each frequency was inverted separately. For the mean phase velocities obtained in our study, the wave paths would exhibit deviations ranging between $\sim$0.4 and 1.7 km. Cells that were not crossed by any ray were discarded in the inversion. Although, a path density higher than three is usually adopted in other studies, we relaxed this criterion due to our limited ray coverage toward the boundaries of the study area.

Since the inversion considers straight ray paths, the slowness model uncertainties σ_s can be directly obtained from the covariance matrix $\underline{C}_s$, defined as (Menke 2018)

$$\underline{C}_s = \left[\underline{G}^T \underline{C}_\chi^{-1} \underline{G} + \underline{\nabla}^T \varepsilon^2 \underline{I} \underline{\nabla} \right]^{-1}, \tag{9}$$

where $\underline{I}$ [$M \times M$] is the identity matrix. Then, the slowness model uncertainty of the jth cell is defined as ($j = 1, 2, \ldots, N$)

$$\sigma_{sj}^2 = C_{s(j,j)}. \tag{10}$$

Finally, the standard deviation of the phase velocity of the jth cell was estimated using error propagation as

$$\sigma_{cj} = \frac{\sigma_{sj}}{s_j^2}, \tag{11}$$

where s_j is the slowness of the jth cell.

Resolution test

The spatial resolution of a velocity model is limited by lack of information and errors in the data (Yanovskaya 1997). The effective resolution limit and the capability of the ray-path coverage to recover the real velocity model can be estimated by performing a checkerboard test (An 2012); here, the average resolution limit corresponds to the dimension of the smallest anomaly solved by the proposed velocity model (Lebedev & Nolet 2003). Considering the entire range of phase velocities observed in the dispersion curves (Fig. 5a), we constructed synthetic velocity models alternating velocity anomalies of different sizes with a 30% of perturbation with respect to the average estimated phase velocity for each frequency (see Fig. 5a). The synthetic data derived from the checkerboard was perturbed adding a Gaussian noise with a coefficient of variation CoV = 2%.

Fig. 6 shows the results of the checkerboard tests for 0.4 and 0.7 Hz. The average phase velocities of the synthetic phase velocity models were 2.76 and 2.33 km s^{-1}, respectively. Figs 6(a) and
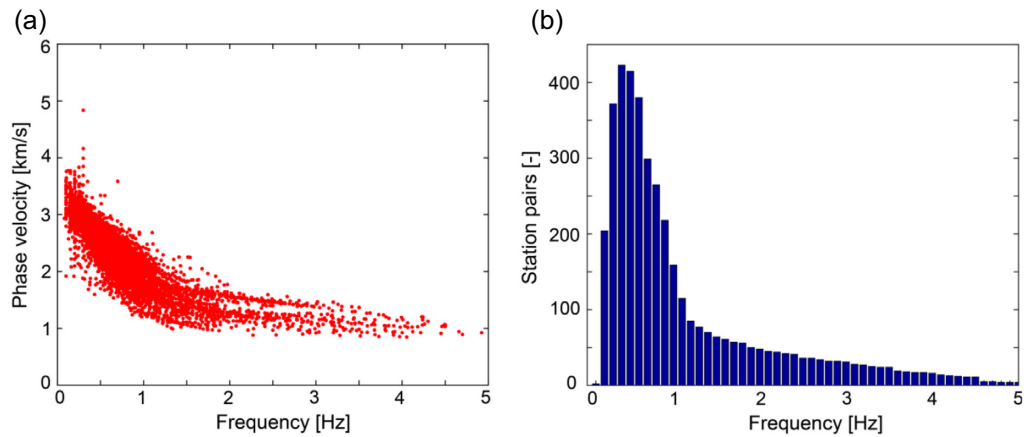
Figure 5. Summary of all the calculated dispersion curves in the study area. (a) Individual phase velocities as a function of frequency and (b) histogram of the number of points measured in every frequency bin.
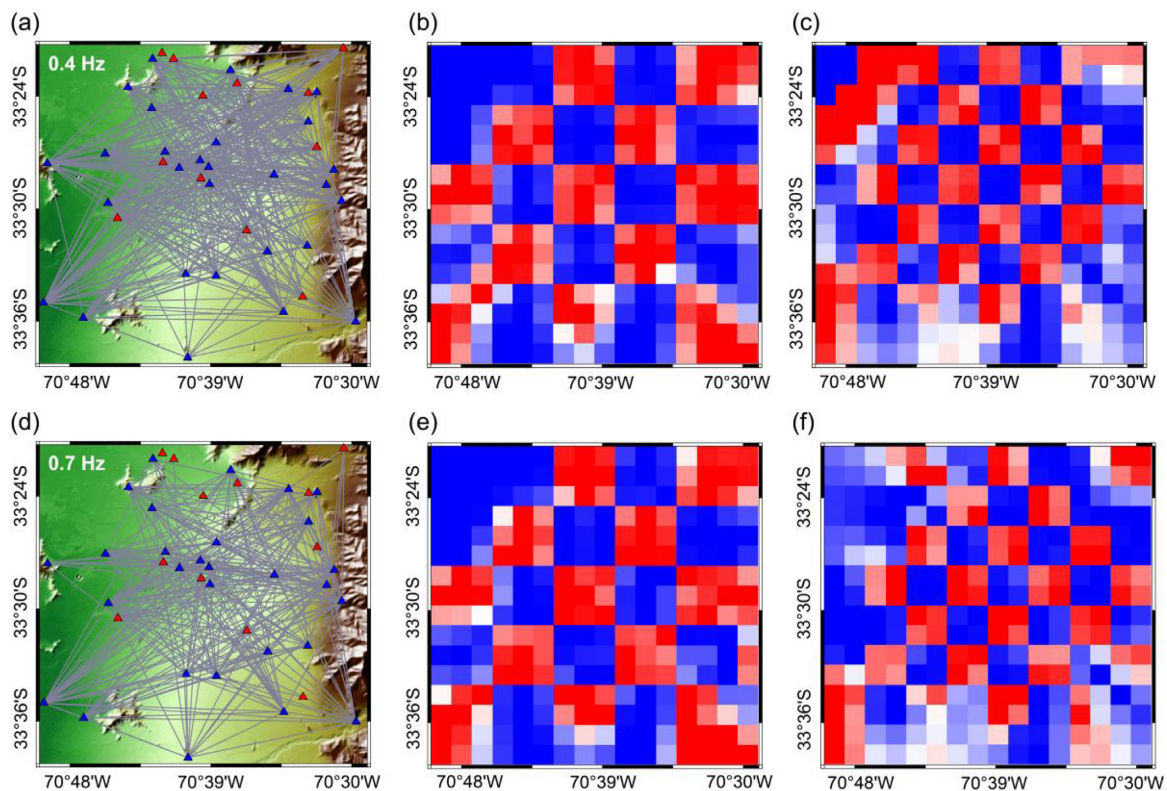

Figure 6. Results of the checkerboard tests for 0.4 Hz (upper panels) and 0.7 Hz (lower panels). The average phase velocities of the synthetic phase velocity models were 2.76 and 2.33 km s⁻¹, respectively. (a) and (d) straight-ray path coverage for both frequencies, (b) and (e) recovered models considering initial anomalies of 6 × 6 km² and (c) and (f) recovered models considering initial anomalies of 4 × 4 km².

(d) illustrate the straight-ray paths for both frequencies, whereas Figs 6(b) and (e) show the recovered models considering initial anomalies of 6 × 6 km² and Figs 6(c) and (f) the recovered models considering initial anomalies of 4 × 4 km².

The ray path coverage allows a proper recovery of the initial models in the study area, except in the northwest and the boundaries of the model, where fewer ray paths cross the cells. Similar results were obtained for various frequencies, allowing to conclude that the ray coverage is enough to resolve structures extended up to 4 × 4 km².

Phase velocity maps

Once the valid frequency band of the model was established, we performed a regularized inversion using the traveltime data derived from the dispersion curves. In this stage, 19 phase velocity maps were independently calculated for discrete frequencies ranging from 0.2 to 1.1 Hz, linearly spaced every 0.05 Hz. Subsequently, dispersion curves were reconstructed for every cell crossed by at least one ray, combining the phase velocities from the maps. The regularized weighted least squares inversion considered a value of the regularization parameter ε that depends on the ray coverage and the
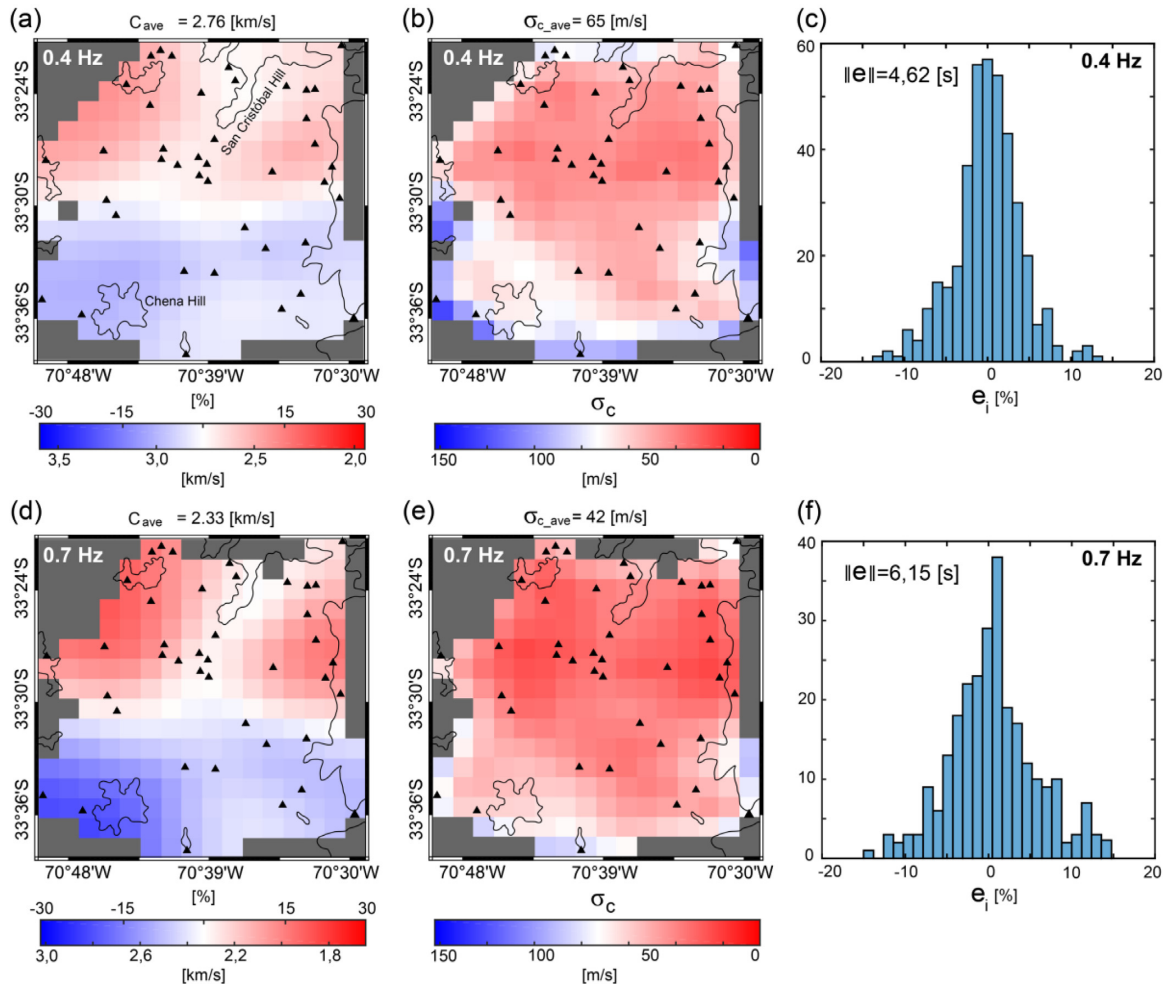
Figure 7. Phase velocity maps (a and d), standard deviations (b and e), and estimated errors (c and f) for 0.4 Hz (upper panels) and 0.7 Hz (lower panels). The triangles indicate the location of the seismic stations.

uncertainties of the observations, estimated from the bootstrapping method. Thus, for every discrete value of frequency (0.2, 0.25, ..., 1.1 Hz), the optimum ε was estimated using the L-curve criterion and the generalized cross-validation method. To avoid numerical artifacts in the subsequent *Vs* model, a constant value of $\varepsilon = 40$ was selected (calculated ε values ranged between 15 and 100) as it balances the smoothness of the phase velocity maps, the fitting errors, and the stability of the inversion process. The tomography was performed by implementing the method described in this section in a Matlab code.

Fig. 7 shows examples of the phase velocity maps for 0.4 Hz (Fig. 7a) and 0.7 Hz (Fig. 7d). The calculated maps display a strong velocity contrast between the northern (low velocities) and southern zones of the basin (high velocities) at both frequencies. This north–south contrast results in maximum perturbation values of 15 and 20% at 0.4 and 0.7 Hz, respectively. Furthermore, the standard deviations do not exceed 0.15 km s^{-1} (Figs 7b and e), which represents less than 5% of the estimated cell velocity. These values are consistent with the standard deviation values obtained for the dispersion curves using the Bootstrapping resampling. The estimated errors are less than 10% for all the observed traveltimes (Figs 7c and f), which indicates an acceptable data fitting.

The estimated phase velocity models for 0.3 Hz (Fig. 8a), 0.5 Hz (Fig. 8b), 0.9 Hz (Fig. 8c) and 1.1 Hz (Fig. 8d) show that the

north–south velocity contrast is larger at higher frequencies, which is likely related to shallower structure of the Abanico Formation.

SHEAR WAVE VELOCITY MODEL

After the phase velocity maps were generated, we reconstructed individual dispersion curves for 157 cells, which were logarithmically resampled in frequency for the subsequent shear wave velocity inversion. This dispersion curves were used to obtain 1-D shear wave velocity profiles for each cell containing information. Similar to the inversion scheme performed by Pastén *et al.* (2016), the 157 dispersion curves were inverted using a Monte Carlo simulation, generating 900 000 velocity models for every dispersion curve. For this purpose, we used the Dinver package, from the Geopsy software (Wathelet 2005). This software allows reducing the searching parameter domain by defining the number of layers and the range of shear wave velocity values.

A test inversion was performed over a limited set of representative dispersion curves of the entire set of 157 curves in order to automate the inversion process and to find the optimal domain of parameters that ensures a balance between computation time and data fitting. Three representative dispersion curves were determined using the self-organizing map (SOM) clustering method (Kohonen 2001; de Matos *et al.* 2006), a neural network method that maps a data set
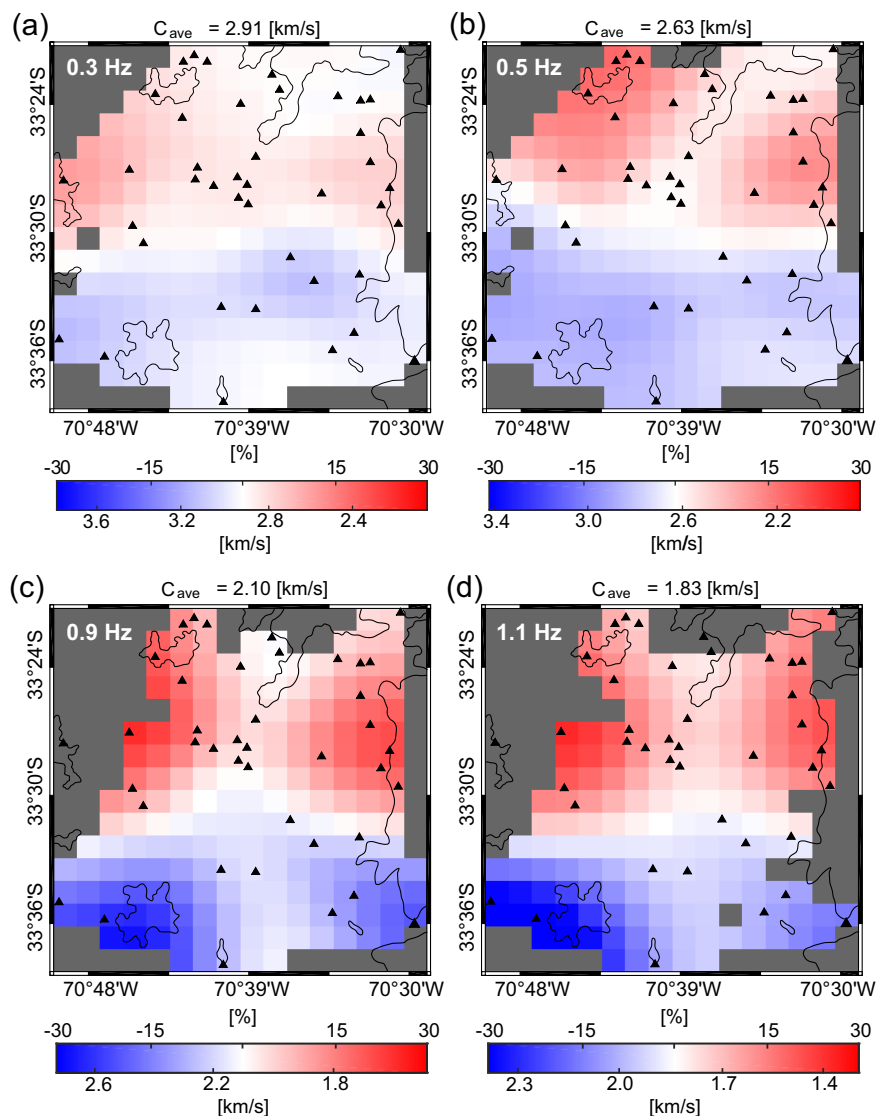
Figure 8. Phase velocity models for (a) 0.3 Hz, (b) 0.5 Hz, (c) 0.9 Hz and (d) 1.1 Hz. The estimated mean phase velocities are shown in the upper part of each map.

into a 2-D representation. This method selects the most representative dispersion curves after a learning process where each curve is compared to those in the map: in each iteration, the most similar curve is averaged with the observation, counting a hit; hence, the most representative curve will become even more similar to the real ones. After the learning process, we selected the most representative curves as those with the largest number of hits. Fig. S1 presents the three most representative dispersion curves with the largest amount of hits of all the observed curves.

The simplest shear wave velocity (Vs) model that accounts for the features of the three representative dispersion curves consists of two layers over a half-space, located between 2 and 5 km depth. The shallower layer increases the Vs following a power law with depth in three sublayers, and the second layer has a constant Vs.

The searching parameter domain considered in the Dinver package is described next. The shallower layer considers shear wave velocities in the range 0.4–2 km s^{-1} at the upper end and 1.5–2 km s^{-1} at the lower edge. The contact between the first and the second layer

is restrained to 0.3–4 km depth. The second layer is considered uniform, with Vs ranging between 1.5 and 4 km s^{-1}. The shear wave velocity of the half-space was allowed to change between 2 and 5 km s^{-1} and its compressional wave velocity varied between 3 and 10 km s^{-1}. Although we acknowledge the importance of estimating a first order approximation of the compressional wave velocity (Vp) and density (e.g. Xing *et al.* 2016), the inversion process considered a uniform density of 2600 kg m^{-3}, a Poisson's ratio between 0.2 and 0.5 and Vp between 0.2 and 1 km s^{-1} for the layers above the half-space since the resulting models were more sensitive to values of Vs than the values of these parameters (Wathelet 2005).

The full inversion of the 157 recovered dispersion curves was performed between 0.2 and 1.1 Hz. Fig. 9(a) shows the best-fitted shear wave velocity profiles for the 157 inverted dispersion curves, along with the average model plus/minus the standard deviation. Note that the contact between the first and second layer is between 0.5 and 1 km depth and the contact between the second layer and the half space is between 2 and 3 km depth. The range of half-space
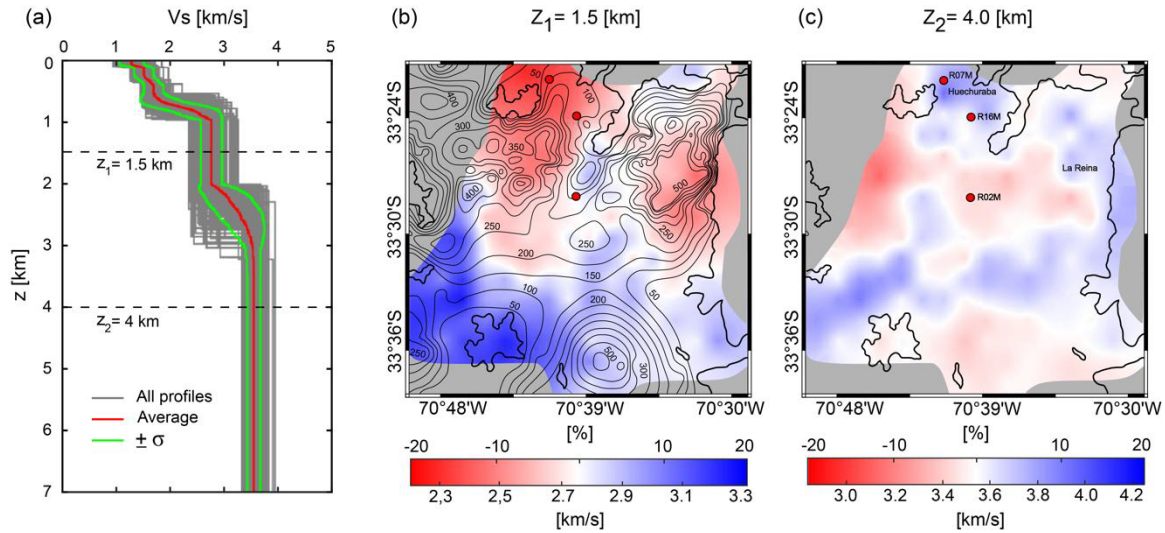
Figure 9. Shear wave velocity model. (a) Best fitted shear wave velocity profiles for the 157 inverted dispersion curves, along with the average model plus/minus the standard deviation and 2-D longitudinal slices at (b) 1.5 km and (c) 4 km depth. The contours shown in (b) is the model for the thickness of the quaternary sediments proposed by González *et al.* (2018).
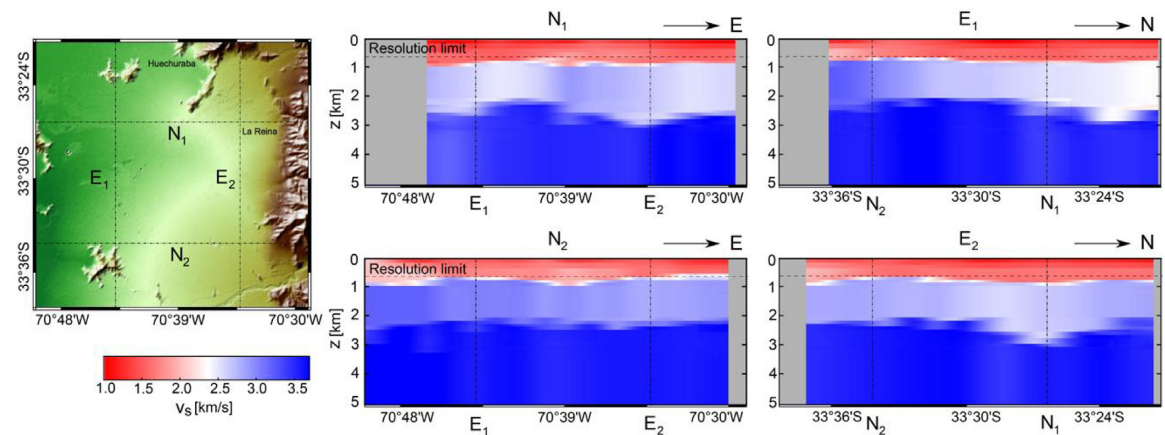

Figure 10. North–south and east–west 2-D profiles from the 3-D velocity model. The resolution limit of 0.6 km depth is explained in the text.
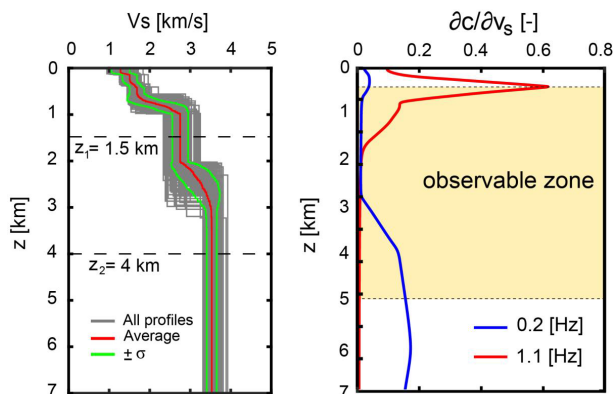

Figure 11. Sensitivity kernels of the average shear wave velocity model at 0.2 and 1.1 Hz.

velocities (3.3–3.9 km s^{-1}) is narrower than the velocities of the two shallower layers at depths larger than 3 km.

Finally, the 3-D shear wave velocity model was obtained by interpolating the 1-D profiles using the ordinary Kriging method (Chiles & Delfiner 2009). Figs 9(b) and (c) show 2-D longitudinal slices at 1.5 and 4 km depth, respectively. The thickness of the quaternary sediments overlying the bedrock proposed by González *et al.* (2018) complements the results shown in Fig. 9(b). Fig. 10 shows two north–south and two east–west 2-D profiles produced from the 3-D velocity model.

The upper and lower resolution limits of our model were estimated at $z_{min} = 0.6$ km and $z_{max} = 5$ km. The depth limit of 5 km is related to the depth where previous studies have already covered, while the upper limit of 0.6 km is due to the depth discretization. These limits are related to the sensitivity kernels for the fundamental mode of surface waves, calculated for the average *Vs* profile shown in Fig. 9(a), using the software srfker96 (Herrmann 2013). A sensitivity kernel is estimated as the partial derivative of the phase velocity with respect to the shear wave velocity $\partial c / \partial V_s$, meaning the sensitivity of the phase velocity to perturbations in the *Vs* model. Two kernels for the lowest (0.2 Hz) and highest (1.1 Hz) frequencies that the model resolves are shown in Fig. 11. The 1.1 Hz kernel shows a maximum sensitivity at 0.5 km depth, whereas the 0.2 Hz kernel exhibits a higher sensitivity deeper than 4 km depth.
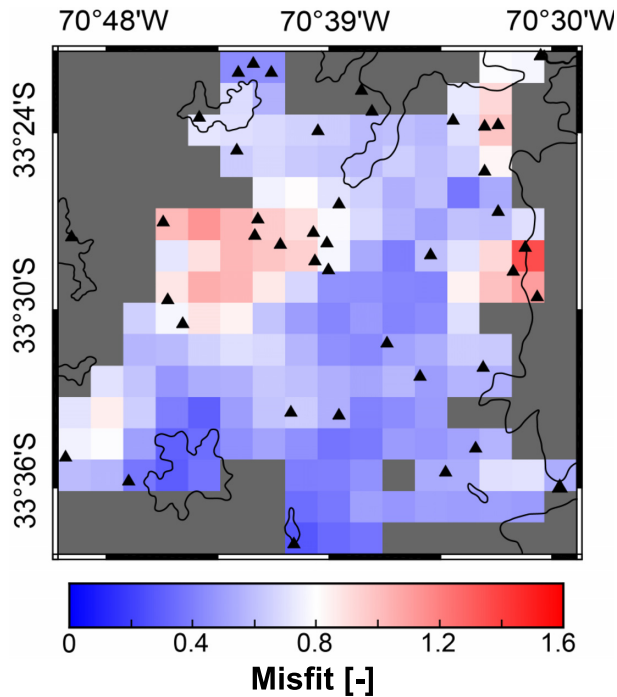
Figure 12. Misfit of the dispersion curves in the velocity model.

The spatial distribution of the data fitting for the 1-D shear wave velocity inversion can be displayed using the misfit parameter defined as

$$misfit = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{\left(c_{pi}-c_{oi}\right)^2}{\sigma_i^2}}, \qquad (12)$$

where c_{pi} is the predicted and c_{oi} is the observed phase velocity at frequency f_i, σ_i is the velocity uncertainty at frequency f_i and n is the number of considered frequencies in the dispersion curves.

The misfit is an overall measure of the difference between the observed dispersion curves (the data) and the fitted curve. The spatial distribution of the misfit in our velocity model is shown in Fig. 12. The larger misfits are identified in the east boundary of the model and the central part of the model. The lower misfits are strongly correlated with the Santiago gravels, the stiffest and strongest sedimentary soil in the basin.

DISCUSSION

The distance between stations, their orientation, and the number of coincident recorded days control the frequency range and the spatial coverage that can be solved in the ANT. The minimum frequency resolved in the dispersion curves is, in most cases, controlled by the wavelength calculated from the interstation distance (eq. 6). Furthermore, the minimum and maximum frequencies resolved in the dispersion curves decay rapidly with the interstation distance. For instance, in average, the minimum and maximum frequencies at an interstation distance of 1 km are 1.6 and 5.3 Hz, whereas these frequencies decrease to 0.3 and 1.1 Hz at an interstation distance of 10 km.

The larger amount of broad-band stations, their longer recorded time (Table S1), and their better spatial distribution in the study area, compared to the accelerographs (Fig. 2), result in dispersion curves from pairs of broad-band stations resolved in a broader

frequency band, between 0.1 and 5.0 Hz. Phase velocity dispersion curves were recovered from 73% of broad-band station pairs. The contribution of the accelerographs is more decisive when they are paired with broad-band stations in the same frequency band, but less significant at frequencies higher than 0.6 Hz when they are paired among themselves (only 41% of the accelerograph station pairs allow recovering dispersion curves).

Regardless of the type of sensors, there is a stronger coherency in the station pairs oriented in the east-west direction and almost lack of coherency in the station pairs oriented with an azimuth ±20°, which may be attributed to the distribution of the energy sources. One consequence of this directionality is that the imaginary part of the coherency is not always zero because the ambient noise field is not diffuse (Prieto *et al.* 2009). Examples of the real and imaginary parts of the average coherency for two station pairs are shown in Fig. S2. The directionality can be identified as antisymmetry in the time-domain EGF, particularly at low frequencies where the ocean–continent interaction towards the west of Santiago may induce long-period surface waves. Examples of time-domain EGFs for different centre frequencies are shown in Fig. S3. The calculation methodology is detailed in Pastén *et al.* (2016). This directional effect at low frequencies cannot be removed from the EGF even if longer recording times are considered in the processing and may affect the deeper shear wave velocities of the proposed model. The directionality is less pronounced at higher frequencies presumably due to the more even distribution of energy sources.

The longest time lags we solved in this study (using the time-domain cross-correlation method described in Pastén *et al.* 2016) were about 15 s between the stations that were separated the farthest apart (distances of about 25–30 km). We could not solve longer time lags because of the low coherence of the signals recorded at those distant stations at high frequencies. The loss of coherence may be due to the anthropic activity in the Santiago City generating high frequency noise that hinders the diffuse background wave field. For the typical interstation distances in this study, time lags were lower than 10 s (examples are shown in Fig. S3). The chosen 2-min windows in the calculation of the coherency are sufficient to capture the typical time lags.

The phase velocity contrast between the northern (low velocities) and southern zones of the basin (high velocities), observed in the phase velocity maps in Figs 7 and 8, results in a marked shear wave velocity contrast between 1 and 2 km depth (Fig. 9b). This contrast attenuates at depths larger than 3 km (Fig. 9c). In view of these velocity contrasts, the ray paths may bend significantly off the great-circle path and our assumption of straight rays may not be completely valid. As a consequence, the use of a uniform cell size of 2×2 km² may not keep ray deviations within the limits of the cells and an adaptive grid size as a function of the frequency could be preferred. In addition, estimation of the Rayleigh wave phase traveltimes using fast-marching method that incorporates ray bending could improve the estimation of the velocity model in future works.

The resolution depths of our velocity model (from 0.5 to 6 km depth) can be estimated more precisely by using the methodology developed in Qiu *et al.* (2019) that uses an improved Neighborhood Algorithm (Wathelet 2008) to generate 40 200 *Vs* models from which they select the ones with the lowest misfits to assess the uncertainty of the 1-D *Vs* inversion as a function of depth.

Our velocity model may be biased by considering cells with only one ray crossing them. The model can be improved by either removing those cells from the domain or increasing the grid size. Fig. 13 shows a histogram with the number of cells as a function of the number of rays crossing them. Cells crossed by one ray
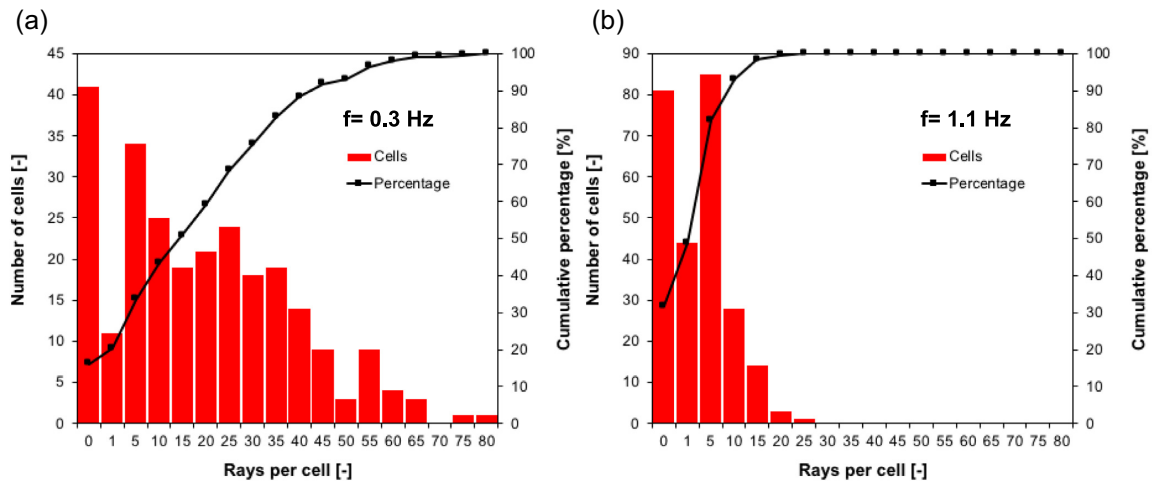
Figure 13. Histogram of number of cells as a function of the number of rays crossing them for (a) $f = 0.3$ Hz and (b) $f = 1.1$ Hz.
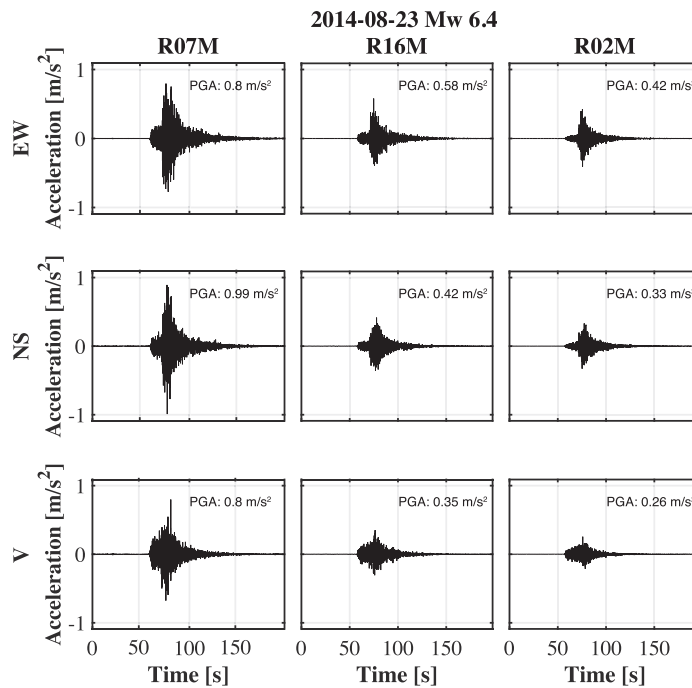

Figure 14. Seismic records of the 2014-08-23 M_w 6.4 earthquake, recorded in stations R07M, R16M and R02M in the east–west (EW), north–south (NS) and vertical (V) directions.

represent 5% of the 215 cells solved in the velocity model at a frequency $f = 0.3$ Hz (Fig. 13a), whereas cells crossed by one ray account for 25% of the 175 cells that defined the velocity model at $f = 1.1$ Hz (Fig. 13b). Restricting the criterion of the number of rays crossing cells can impact the velocity model at higher frequencies, which can be complemented with local geophysical campaigns in future works.

Albeit the limitations, the obtained velocity model of the Abanico Formation is a first order approximation that improves the resolution of available large scale velocity models for Central Chile at shallow depths (Ward *et al.* 2013; Marot *et al.* 2014) and complements the velocity models of the shallower quaternary sediments (Pilz *et al.* 2010).

Our *Vs* model identifies a high velocity area in the southwest and at least two zones with high impedance contrast between the

half-space and the shallower structure in the north area of Huechuraba and the east area of La Reina (Figs 9 and 10). Shallower wave velocities in the Huechuraba area are also lower than the rest of the study area, with a shear wave velocity of the upper 30 m near to $Vs_{30} = 280$ m s^{-1}. The heterogeneous velocity distribution in the Abanico Formation may further aggravate the site effects induced by the shallower quaternary deposits and increase surface ground motions during earthquakes. Fig. 14 shows the three-components seismic records of an earthquake that occurred on 23 October 2014, at 22:32:23 (UTC), moment magnitude $M_w = 6.4$, epicentre –32.74, –71.50 and 40 km depth. The records in the east–west (EW), north–south (NS) and vertical (V) directions correspond to stations R07M, R16M and R02M (see the location in Figs 2 and 9). The station closest to the Huechuraba area is R07M (epicentral distance of 103 km) which shows the largest accelerations compared to stations R16M

over a transition soil type and R02M over stiff gravel. The large relative accelerations of station R07M exemplifies the amplification pattern in the norther area of our velocity model.

CONCLUSIONS

This study attempts to define the first shear wave velocity model of the upper part of the Abanico Formation underneath the Metropolitan Area of the Santiago City, Chile, using ANT.

The distance between stations, their orientation, and the number of coincident recorded days limits the frequency range and the spatial coverage that can be solved with the ANT. In fact, 34% of the station pairs were discarded in the tomography.

The implemented Bootstrap method allows estimating the variance associated of the dispersion curves and the traveltimes, as well as assisting the error analysis of complex formulations without closed-form expressions. In this work, the technique defines an average standard deviation of less than 2.5% for the calculated traveltimes.

The method developed to automatically determine the reliable frequency band of the dispersion curves allows solving phase velocities between 0.1 and 5.0 Hz, but the ray path coverage limits the ANT to frequencies between 0.2 and 1.1 Hz. These frequencies are associated to wavelengths from 1.7 to 16 km, approximately, which help defining the shear wave velocity model from 0.6 to 5 km depth.

Traveltime tomography allows estimating 2-D phase velocity maps from 0.2 to 1.1 Hz, using the dispersive information of all the available ray paths. The distribution and distance between stations are the critical variables that determine the inversion stability, the optimal level of regularization, and the spatial model resolution. The studied domain was discretized into 2×2 km^2 cells, following the criterion of the quarter wavelength and ensuring the stability of the traveltime inversion at frequencies above 0.7 Hz.

The velocity anomalies in the phase velocity maps increase up to 30% with respect to the average value in the E–N plane, as the frequency increases. The average maximum resolution of the phase velocity model is 4 km. However, the phase velocity model does not resolve areas with scarce ray path coverage in the edges of the domain and the northwest of the study area.

The lateral variations or anomalies in the 3-D shear wave velocity model increase up to 20% of the average value in the E–N plane towards the surface, but the variations decrease with depth. The variations detected in the model show that the southern sector is stiffer than the northern one. The sensitivity of the model is limited to 0.6–5.0 km depth, lacking the resolution required to elucidate the transition between the Abanico Formation and the shallower quaternary sediments. The *Vs* model developed in this study will help advancing the understanding of the site effects in highly populated megacities that are prone to seismic risk.

ACKNOWLEDGEMENTS

REFERENCES

Aki, K., 1957. Space and time spectra of stationary waves, with special reference to microtremors, *Bull. Earthq. Res. Inst. Tokyo Univ.,* **35,** 415–456.

Ammirati, J.-B., Vargas, G., Rebolledo, S., Abrahami, R., Potin, B., Leyton, F. & Ruiz, S., 2019. The crustal seismicity of the Western Andean Thrust (Central Chile, 33°–34° S): implications for regional tectonics and seismic hazard in the Santiago Area, *Bull. seism. Soc. Am.,* **109,** 1985–1999.

An, M., 2012. A simple method for determining the spatial resolution of a general inverse problem, *Geophys. J. Int.,* **191,** 849–864.

Armijo, R., Rauld, R., Thiele, R., Vargas, G., Campos, J., Lacassin, R. & Kausel, E., 2010. The West Andean Thrust, the San Ramón Fault, and the seismic hazard for Santiago, Chile", *Tectonics,* **29,** 1–34.

Aster, R.C., Borchers, B. & Thurber, C.H., 2018. *Parameter Estimation and Inverse Problems,* Elsevier.

Astroza, M., Ruiz, S. & Astroza, R., 2012. Damage assessment and seismic intensity analysis of the 2010 (Mw 8.8) Maule earthquake, *Earthq. Spectra,* **28,** 145–164.

Barrientos, S., Vera, E., Alvarado, P. & Monfret, T., 2004. Crustal seismicity in central Chile, *J. South Am. Earth Sci.,* **16,** 759–768.

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.,* **169,** 1239–1260.

Charrier, R. *et al.* 2002. Evidence for Cenozoic extensional basin development and tectonic inversion south of the flat-slab segment, southern Central Andes, Chile (33°-36° S.L.), *J. South Am. Earth Sci.,* **15,** 117–139.

Charrier, R. *et al.* 2005. The Abanico extensional basin: regional extension, chronology of tectonic inversion and relation to shallow seismic activity and Andean uplift, *Neues Jahrb. für Geol. und Paläontologie - Abhandlungen,* **236,** 43–77.

Chiles, J.-P. & Delfiner, P., 2009. *Geostatistics: Modeling Spatial Uncertainty,* John Wiley & Sons.

Craven, P. & Wahba, G., 1978. Smoothing noisy data with spline functions, *Numer. Math.,* **31,** 377–403.

de Matos, M.C., Osorio, P.L. & Johann, P.R., 2006. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps, *Geophysics,* **72,** P9–P21.

Ekström, G., 2014. Love and Rayleigh phase-velocity maps, 5–40 s, of the western and central USA from USArray data, *Earth Planet. Sci. Lett.,* **402,** 42–49.

Ekström, G., Abers, G.A. & Webb, S.C., 2009. Determination of surface-wave phase velocities across USArray from noise and Aki's spectral formulation, *Geophys. Res. Lett.,* **36,** 5–9.

Farías, M., Comte, D., Charrier, R., Martinod, J., David, C., Tassara, A., Tapia, F. & Fock, A., 2010. Crustal-scale structural architecture in central Chile based on seismicity and surface geology: implications for Andean mountain building, *Tectonics,* **29**(3),.

Giambiagi, L. *et al.* 2014. Evolution of shallow and deep structures along the Maipo-Tunuyan transect (33 40'S): from the Pacific coast to the Andean foreland, *Geol. Soc. Lond., Spec. Publ.,* **399,** 63–82.

Godoy, E., Yañez, G. & Vera, E., 1999. Inversion of an Oligocene volcano-tectonic basin and uplifting of its superimposed Miocene magmatic arc in the Chilean Central Andes: first seismic and gravity evidences, *Tectonophysics,* **306,** 217–236.

González, F.A. *et al.* 2018. Characterization of the depocenters and the basement structure, below the central Chile Andean Forearc: a 3D geophysical modelling in Santiago Basin area, *Basin Res.,* **30,** 799–815.

Hansen, P.C., 1992. Analysis of discrete ill-posed problems by means of the L-curve, *SIAM Rev.,* **34,** 561–580.

Herrmann, R.B., 2013. Computer programs in seismology: an evolving tool for instruction and research, *Seismol. Res. Lett.,* **84,** 1081–1088.

Inzunza, D.A., Montalva, G.A., Leyton, F., Prieto, G. & Ruiz, S., 2019. Shallow ambient-noise 3D tomography in the concepción basin, Chile: implications for low-frequency ground motions, *Bull. seism. Soc. Am.,* **109,** 75–86.

Jin, G. & Gaherty, J.B., 2015. Surface wave phase-velocity tomography based on multichannel cross-correlation, *Geophys. J. Int.,* **201,** 1383–1398.

Kawase, H., 1996. The cause of the damage belt in Kobe: "The Basin-Edge Effect," constructive interference of the direct S-wave with the basin-induced diffracted/Rayleigh waves, *Seismol. Res. Lett.,* **67,** 25–34.

Kohonen, T., 2001. *Self-Organizing Maps,* Springer-Verlag.

Kugler, S., Bohlen, T., Forbriger, T., Bussat, S. & Klein, G., 2007. Scholte-wave tomography for shallow-water marine sediments, *Geophys. J. Int.,* **168,** 551–570.

Lai, C.G., Rix, G.J., Foti, S. & Roma, V., 2002. Simultaneous measurement and inversion of surface wave dispersion and attenuation curves, *Soil Dyn. Earthq. Eng.,* **22,** 923–930.

Lebedev, S. & Nolet, G., 2003. Upper mantle beneath Southeast Asia from S velocity tomography, *J. geophys. Res.,* 108,.

Liu, X., Ben-Zion, Y. & Zigone, D., 2016. Frequency domain analysis of errors in cross-correlations of ambient seismic noise, *Mon. Not. R. Astr. Soc.,* **207,** 1630–1652.

Luo, Y., Yang, Y., Xu, Y., Xu, H., Zhao, K. & Wang, K., 2015. On the limitations of interstation distances in ambient noise tomography, *Geophys. J. Int.,* **201,** 652–661.

Makra, K. & Chávez-García, F.J., 2016. Site effects in 3D basin using 1D and 2D models: an evaluation of the differences based on simulations of the seismic response of Euroseistest, *Bull. Earthq. Eng,* **14,** 1177–1194.

Marot, M., Monfret, T., Gerbault, M., Nolet, G., Ranalli, G. & Pardo, M., 2014. Flat versus normal subduction zones: a comparison based on 3-D regional traveltime tomography and petrological modelling of central Chile and western Argentina (29°-35° S), *Geophys. J. Int.,* **199,** 1633–1654.

Menke, W., 2018. *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press.

Moczo, P., Kristek, J., Bard, P.-Y., Stripajová, S., Hollender, F., Chovanová, Z., Kristeková, M. & Sicilia, D., 2018. Key structural parameters affecting earthquake ground motion in 2D and 3D sedimentary structures, *Bull. Earthq. Eng.,* **16,** 2421–2450.

Mpodozis, C. & Ramos, V., 1990. *The Andes of Chile and Argentina,*

Muñoz-Saez, C., Pinto, L., Charrier, R. & Nalpas, T., 2014. Influencia de la carga de los depositos en la generacion de fallas de atajo durante la inversion de una cuenca extensional: El caso de la cuenca de Abanico (Eoceno-Oligoceno), Andes de Chile central (33º-35ºS), *Andean Geol.,* **41,** 1–28.

Nolet, G., 2012. *Seismic Tomography: With Applications in Global Seismology and Exploration Geophysics,* Springer Science & Business Media.

Nyström, J.O., Vergara, M., Morata, D. & Levi, B., 2003. Tertiary volcanism during extension in the Andean foothills of central Chile (33 15′–33 45′ S), *Geol. Soc. Am. Bull.,* **115,** 1523–1537.

Pastén, C., Sáez, M., Ruiz, S., Leyton, F., Salomón, J. & Poli, P., 2016. Deep characterization of the Santiago Basin using HVSR and cross-correlation of ambient seismic noise, *Eng. Geol.,* **201,** 57–66.

Picozzi, M., Parolai, S., Bindi, D. & Strollo, a., 2009. Characterization of shallow geology by high-frequency seismic noise tomography, *Geophys. J. Int.,* **176,** 164–174.

Pilz, M., Parolai, S., Picozzi, M., Wang, R., Leyton, F., Campos, J. & Zschau, J., 2010. Shear wave velocity model of the Santiago de Chile basin derived from ambient noise measurements: a comparison of proxies for seismic site conditions and amplification, *Geophys. J. Int.,* **182,** 355–367.

Prieto, G.A., Lawrence, J.F. & Beroza, G.C., 2009. Anelastic Earth structure from the coherency of the ambient seismic field, *J. geophys. Res.,* **114,** 1–15.

Qiu, H., Lin, F.C. & Ben-Zion, Y., 2019. Eikonal tomography of the Southern California Plate Boundary Region, *J. geophys. Res.,* **124,** 9755–9779.

Riesner, M., Lacassin, R., Simoes, M., Armijo, R., Rauld, R. & Vargas, G., 2017. kinematics of the active west andean fold-and-thrust belt (Central Chile): structure and long-term shortening rate, *Tectonics,* **36,** 287–303.

Riesner, M., Lacassin, R., Simoes, M., Carrizo, D. & Armijo, R., 2018. Revisiting the crustal structure and kinematics of the Central Andes at

33.5° S: implications for the mechanics of Andean mountain building, *Tectonics,* **37,** 1347–1375.

Ruiz, S. & Madariaga, R., 2018. Historical and recent large megathrust earthquakes in Chile, *Tectonophysics,* **733,** 37–56.

Sabra, K.G., Gerstoft, P., Roux, P. & Kuperman, W.A., 2005. Extracting time-domain Green's function estimates from ambient seismic noise, *Geophys. Res. Lett.,* **32,** L03310, doi:10.1029/2004GL021862.

Sánchez-Sesma, F.J. & Campillo, M., 2006. Retrieval of the Green's function from cross correlation: the canonical elastic problem, *Bull. seism. Soc. Am.,* **96,** 1182–1191.

Sellés, D. & Gana, P., 2001. *Geología del área Talagante-San Francisco de Mostazal: Región Metropolitana de Santiago y del Libertador General Bernardo O'Higgins,* Servicio Nacional de Geología y Minería.

Semblat, J.F., Kham, M., Parara, E., Bard, P.Y., Pitilakis, K., Makra, K. & Raptakis, D., 2005. Seismic wave amplification: basin geometry vs soil layering, *Soil Dyn. Earthq. Eng.,* **25,** 529–538.

Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science,* **307,** 1615–1618.

Shirzad, T. & Shomali, Z.H., 2013. Shallow crustal structures of the Tehran basin in Iran resolved by ambient noise tomography, *Geophys. J. Int.,* **196,** 1162–1176.

Vargas, G., Klinger, Y., Rockwell, T.K., Forman, S.L., Rebolledo, S., Baize, S., Lacassin, R. & Armijo, R., 2014. Probing large intraplate earthquakes at the west flank of the Andes, *Geology,* **42,** 1083–1086.

Vassallo, M. *et al.* 2019. Seismic noise cross-correlation in the urban area of Benevento city (Southern Italy), *Geophys. J. Int.,* **217,** 1524–1542.

Vergara, M., López-Escobar, L., Palma, J.L., Hickey-Vargas, R. & Roeschmann, C., 2004. Late Tertiary volcanic episodes in the area of the city of Santiago de Chile: new geochronological and geochemical data, *J. South Am. Earth Sci.,* **17,** 227–238.

Ward, K.M., Porter, R.C., Zandt, G., Beck, S.L., Wagner, L.S., Minaya, E. & Tavera, H., 2013. Ambient noise tomography across the Central Andes, *Geophys. J. Int.,* **194,** 1559–1573.

Wathelet, M., 2008. An improved neighborhood algorithm: parameter conditions and dynamic scaling, *Geophys. Res. Lett.,* **35,** 1–5.

Wathelet, M., 2005. *Geopsy Geophysical Signal Database for Noise Array Processing,* Software, LGIT, Grenoble.

Xing, G., Niu, F., Chen, M. & Yang, Y., 2016. Effects of shallow density structure on the inversion for crustal shear wave speeds in surface wave tomography, *Geophys. J. Int.,* **205,** 1144–1152.

Yañez, G., Muñoz, M., Flores-Aqueveque, V. & Bosch, A., 2015. Gravity derived depth to basement in Santiago Basin, Chile: implications for its geological evolution, hydrogeology, low enthalpy geothermal, soil characterization and geo-hazards, *Andean Geol.,* **42,** 147–172.

Yanovskaya, T.B., 1997. Resolution estimation in the problems of seismic ray tomography, *Izv. Phys. Solid Earth,* **33,** 762–765.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Representative dispersion curves obtained using the self-organizing map (SOM) clustering method.
Figure S2. Examples of the real and imaginary parts of the average coherency for station pairs (a) DG03–DG20 and (b) DG19–DG24.
Figure S3. Time-domain empirical Green's functions (EGF) with typical time-lags obtained from the time-domain cross-correlation method described in Pastén *et al.* (2016).
Table S1. Name, location and recording period of seismic stations used in this study.

Please note: Oxford University Press are 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.