

Shear Wave Velocities in the San Gabriel and San Bernardino Basins, California

Yida Li¹, Valeria Villa¹, Robert W Clayton¹, Patricia Persaud²

¹Seismological Laboratory, California Institute of Technology, Pasadena, CA, USA.

²Department of Geology and Geophysics, Louisiana State University, Baton Rouge, LA, USA

Corresponding author: Yida Li (yidali@caltech.edu)

Key Points:

- We construct a 3D Vs model in San Gabriel and San Bernardino basins using ambient noise correlation between dense array nodal, broadband and accelerometer stations.
- We separated the Rayleigh wave fundamental mode and first higher mode in dispersion analysis based on the Rayleigh wave particle motion.
- Our Vs model predicts deeper and slower sedimentary basins than the SCEC CVMS model, yet is consistent with geological and drilling data in these basins.

Abstract

We construct a new shear velocity model for the San Gabriel, Chino and San Bernardino basins located in the northern Los Angeles area using ambient noise correlation between dense linear nodal arrays, broadband stations, and accelerometers. We observe Rayleigh wave and Love wave in the correlation of vertical (Z) and transverse (T) components, respectively. By combining Hilbert and Wavelet transforms, we obtain the separated fundamental and first higher mode of the Rayleigh wave dispersion curves based on their distinct particle motion polarization. Receiver functions, gravity, and borehole data are incorporated into the prior model to constrain the basin depth. Our 3D shear wave velocity model covers the upper 3 to 5 km of the basin structure in the San Gabriel and San Bernardino basin area. The Vs model is in agreement with the geological and geophysical cross-sections from other studies, but discrepancies exist between our model and a Southern California Earthquake Center (SCEC) community velocity model. Our shear wave velocity model shows good consistency with the CVMS 4.26 in the San Gabriel basin, but predicts a deeper and slower sedimentary basin in the San Bernardino and Chino basins than the community model.

Plain Language Summary

Sedimentary basins northeast of Los Angeles can potentially be a low-velocity channel that focus earthquake energy from the San Andreas fault to the Los Angeles region. To better understand the focusing effect, we build up a new velocity model of this area using a new seismic dataset. With the cross-correlation technique, we extract the travel time information between two stations from the ambient noise, and together with the gravity and receiver functions constraining the depth of the sedimentary basement, we build a 3D shear wave velocity model. Many geological features, like sedimentary basins and faults, are captured in our velocity model. Compared to the community velocity model, our model predicts a deeper sedimentary structure with lower velocity, indicating the focusing effect of the sedimentary basins northeast of Los Angeles might be underestimated.

1 Introduction

The San Gabriel (SG) and San Bernardino (SB) basins are sedimentary basins northeast of the city of Los Angeles (Fig 1). The SG basin consists of two sub-basin structures: the Raymond basin on the west and the San Gabriel basin on the east, separated by the Raymond fault. The SB region, immediately to the east of SG, is composed of three sedimentary basins: Chino basin, Riato-Colton basin, and San Bernardino basin from west to east. Bounded by mountains both to the north and south, the sedimentary structures in SG and SB area were as part of the opening of the Los Angeles basin region in the Miocene.

Understanding the velocity structure of SG and SB area is important for the accurate hazard assessment of the densely populated Los Angeles region because the low-velocity basins in the SG and SB area may function as a waveguide that channels earthquake energy from the San Andreas fault (SAF) into the Los Angeles region (Olsen et al., 2006). Numerical simulations such as the ShakeOut Scenario (Jones et al., 2008) and CyberShake (Graves et al., 2011) show events on the southern SAF may cause large ground motions in downtown Los Angeles. A study using ambient noise correlation estimate (Denolle et al., 2014) found the ground motion could be four times larger than the simulation. This implies the current Southern California Earthquake

Center (SCEC) Community Velocity Model (CVM) used in the ground motion simulations do not adequately account for the channeling effect of the northern sedimentary basins (Clayton et al., 2019). A recent study in the Los Angeles basin constrains the velocity model using dense industry arrays correlated with broadband stations (Jia and Clayton, 2021), and the new fine-scale velocity model's strong motion amplification performs similar to the CVMH model but better than the CVMS model. We attribute the underestimation of ground motion in numerical simulations to the inaccuracy of the community velocity models in the SG and SB area as this area is not as well constrained as the Los Angeles basin where dense industry array data and borehole measurements are more readily available.

The community model (CVMS) in this region has evolved over several generations, with the earliest version of the CVMS model comprised of rule-based basin models constrained by empirical equations and a few well logs (Magistrale et al, 2000). In the subsequent versions, a geotechnical layer was incorporated and full waveform inversion was introduced into the model. However, due to the limited number of broadband stations deployed in the SG and SB region (black triangles in Fig. 1), the modification of the CVMS model through the different versions is small in this area, and the final version of the CVMS model (CVMS 4.26) retains the original CVMS model's primary characteristics from the geology and borehole dataset. In order to better constrain the velocity model in SG and SB area, we deployed a set of linear dense nodal arrays, and combine the ambient noise cross-correlation and receiver function techniques applied to this dataset, along with the Bouguer gravity anomaly and borehole dataset to construct a new shear wave velocity model.

In the past few decades, the ambient noise technique has been widely applied to construct velocity models. With a homogeneous ambient noise source distribution, the cross-correlation of the ambient noise signal from two stations can provide the surface wave Empirical Green's Functions (EGF), in the causal ($t > 0$) and anti-causal ($t < 0$) sense, between the two stations (Snieder, 2004). The correlation of different receiver components generates different surface wave EGF: the Rayleigh wave in the vertical (Z) and radial (R) components and the Love wave in the transverse (T) component (Lin et al., 2008). In this study, we extract Rayleigh wave EGF from ZZ correlation, and Love wave EGF from TT correlation. With the surface wave EGF's, group and phase velocity dispersion curves can be measured (Yao et al., 2006), which allows tomographic phase and group velocity maps to be constructed (Herrmann, 2013). These are then used to invert for shear wave velocity, V_s .

Compared to the crustal-scale survey using the long-period ambient noise correlation between broadband stations, the surface wave EGF in high-frequency ambient noise correlation is less coherent due to the greater structural variations in sedimentary basins. In recent years, the deployment of dense seismic arrays makes it possible to resolve the fine-scale velocity structure of the top 5 km sedimentary layer (Castellanos and Clayton, 2021; Jia and Clayton, 2021; Lin et al., 2013). In addition to the ambient noise correlation, receiver functions are also evaluated from the dense array datasets to constrain the basement depth within the sedimentary basins (Ma and Clayton, 2016; Liu et al, 2018; Wang et al., 2021). Receiver function using our linear dense arrays has shown a coherent converted phase at the basin bottom can be observed in the SG and SB area, which provide an independent constraint on the basin structure in this area (Liu et al., 2018; Wang et al., 2021).

In this study, we construct a shear wave velocity model in the SG and SB area using 10 linear dense array datasets together with broadband stations and accelerometers. We correlate the

vertical (ZZ) and transverse component (TT) ambient noise recordings to obtain Empirical Green's Functions and perform a dispersion analysis to extract the group and phase velocities. We developed a method to separate Rayleigh wave modes in the dispersion analysis based on the Rayleigh wave particle motion. Our Vs model incorporates both group and phase velocity tomography maps and starts with an initial model constrained with receiver functions, Bouguer gravity, and borehole data. We finally compare our Vs model with previous studies and the community velocity models.

2 Data

The dataset is made up of three different types of seismograms: 1) the temporary linear dense Basin Amplification Seismic Investigation (BASIN) nodal arrays, 2) permanent and temporary broadband stations and 3) strong-motion accelerometers. The distribution of the stations is shown in Figure 1. Between 2017 and 2019, 10 linear dense BASIN nodal arrays (SG1 to SG4, and SB1 to SB6) were deployed in the San Gabriel and San Bernardino basins during four deployment periods. The dense arrays consisted of lines with 14 to 260 Fairfield ZLand nodes with a standard 5 Hz 3-component geophone, with the nodes spaced ~250 m apart. Each of the dense arrays was deployed for approximately one month. The broadband stations dataset includes the permanent Southern California Seismic Network (SCSN) stations and 14 temporary broadband stations deployed in 2018, indicated with triangles in Figure 1. In this study, we use the passive ambient noise method on the combined dataset, to extract the EGF and with this construct a three-dimensional Vs model.

3 Method

3.1 Ambient noise correlation

To estimate the shear wave velocity, we first determine the EGF between each station using ambient noise correlations. The noise correlation follows the technique described in Bensen et al. (2007) and Jia and Clayton (2021). To reduce the influence of anthropogenic noise, we correlate only the nighttime (8:00 pm to 8:00 am, local time) ambient noise. We include all the possible ray pairs, including node to node, node to broadband, and node to accelerometer, that have overlapping recording times. The data are correlated in one-hour segments and stacked to get the final correlation. To minimize the effect of earthquakes and broaden the effective period range, we do time domain normalization and spectral whitening prior to the correlation. For the node-to-node correlation, as the stations of every pair have the same instrument response, it cancels out in the spectral whitening, and therefore the removal of instrument response was not required in our case. For the node-to-accelerometer correlation, we will show that a $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ phase shift is introduced because of the difference in the instrument response, and special care should be taken when stacking the causal and anti-causal Green's function (Appendix A). As all of our stations are 3-component, we can extract both Rayleigh and Love waves. We rotate the components from the ZNE into the ZRT coordinate system. The Rayleigh wave particle motion is in the RZ plane and the Love wave particle motion is mainly in the T direction, and hence we correlate the Z component of the virtual source and virtual receiver, called the ZZ correlation, to get the Rayleigh wave EGF, and use the TT correlation to get Love wave EGF. In Figures 2 and

Figure 3, we show examples of the ZZ and TT correlations and dispersion curves for SG1 using station 120 as a virtual source and all stations in the SG1 line as virtual receivers. From both the Love (TT) and Rayleigh (ZZ) waves we can see two consistent dispersive fundamental modes in the $t > 0$ and $t < 0$ domain, as well as some low-frequency first-higher-mode Rayleigh waves. Some high frequency scattered waves are also present in the correlation functions, which interfere with the direct wave EGF signals in some cases.

3.2 Group Velocity Dispersion Picking

Our method for picking the surface wave dispersion curve from the EGF is modified from Yao et al (2006). We firstly fold the EGF at $t=0$. When both the virtual source and receiver are the same type of sensor, the causal ($t > 0$) and anti-causal ($t < 0$) branches are symmetric, and we therefore add the two branches to enhance the signal. For velocity sensors (i.e., nodes) to accelerometer correlations, due to the phase difference in the instrument response, we subtract the causal branch from the folded anti-causal branch. Details on the derivation of this approach are provided in Appendix A.

Next, we apply the Hilbert transform to a set of frequency bands to obtain the signal envelope in terms of period, T . In Figure 3a, we show an example of the group velocity dispersion picking, where the signal envelope function is color-coded in the frequency (period) and group slowness ($u=t/d$) domain. A typical group velocity dispersion curve is picked along the peak of the envelope, which is usually continuous. Solid lines in Figure 3a show the dispersion curve picks for the fundamental model (red) and first higher mode (blue). However, the picking of the group velocity dispersion curve with this method is sometimes ambiguous for two main reasons: 1) When the fundamental mode is close to the higher mode, different modes may interfere with each other and the different modes cannot be separated based on the envelope alone. 2) The envelope pattern is sometimes discontinuous, e.g., the higher mode in Figure 3a at period range between 1.5 s and 4 s. In order to distinguish between the fundamental mode and the first higher mode Rayleigh wave, we developed a new technique based on the polarization of particle motion. For the Love wave, the higher mode is substantially weaker than the fundamental, therefore we only extract its fundamental mode dispersion curves.

3.3 Rayleigh Wave Mode Separation

Our identification of Rayleigh wave modes is based on the particle motion of the waves. For the fundamental mode, the Rayleigh wave particle motion is typically retrograde, while the Rayleigh wave first higher mode is prograde. The retrograde and prograde particle motions reflect the phase lag between the Z and R components. For retrograde motion, the R component is $T/4$ ahead of the Z component, and for prograde, it is $T/4$ behind. For the ambient noise correlation, the phase difference between ZZ and ZR is the same as the phase difference between Z and R (Appendix B), so the relationship between ZZ and ZR reflects the polarization of the Rayleigh wave particle motion in the same way. In a previous study, (Ma et al., 2016) have shown that in the sedimentary basin the ZZ and ZR correlation show consistent retrograde fundamental mode

and prograde first higher mode. Here we present a quantitative way of measuring the particle motion using the Continuous Wavelet Transform (CWT)

$$W_x(s, n) = \left(\frac{\delta t}{s}\right)^{1/2} \sum_{n'=1}^N x_{n'} \Phi_0^* \left[\frac{(n' - n)\delta t}{s} \right]$$

where Φ_0^* is the wavelet function (Torrence and Compo, 1998), s is the wavelet scale, and δt is the time step. As with the Fourier transform, the variation of s gives a spectral pattern in the frequency domain, but the wavelet transform also has an additional dimension, n that reflects the temporal variation. The wavelet transformation has been proven to be a powerful technique to monitor temporal variation in the coda with high precision (Mao et al., 2020). Here, we use it to evaluate the phase difference between the ZZ and ZR correlations, and when combined with the Hilbert transform it produces a clear separation of the fundamental from the first higher mode. We apply the CWT using the Matlab Wavelet Toolbox with the Morse wavelet function Φ_0^* . The phase difference between the ZR and ZZ correlations is $\delta = \arg(W_{ZR}(s, n)) - \arg(W_{ZZ}(s, n))$. For a retrograde fundamental mode, this is $\pi/2$. In contrast, for the prograde first-higher mode $\delta = -\pi/2$. We plot $\sin \delta$ in Figure 3b to quantify the polarization of particle motion in the group slowness and period domain, where red and blue are positive ($\delta = \pi/2$) and negative ($\delta = -\pi/2$) phase shifts, respectively, corresponding to retrograde and prograde particle motion. However, as the $\sin \delta$ pattern only represents the phase difference, noise and signal are not distinguishable in this representation. To combine the amplitude and phase information, we multiply the wave envelope from the Hilbert transformation in Figure 3a and the $\sin \delta$ in Figure 3b to produce the result shown in Figure 3c. In Figure 3c, red representing the retrograde fundamental ($\sin \delta = 1$) mode and blue representing the prograde first higher mode ($\sin \delta = -1$) are clearly separated. Our picking of the Rayleigh wave group velocity dispersion curve is based on this pattern.

3.4 Tomography

With the measured dispersion curves, we applied the straight-ray tomography method to invert the frequency dependent group velocity maps. We discretize the area into a uniform grid with 0.55 km longitudinal spacing and 0.66 km latitudinal spacing. The group velocity tomography is carried out between 0.5 s and 3 s period, using the travel times from the dispersion curves. Figure 4 shows an example of the straight ray coverage of group velocity at period $T=1$ s, where the picked group velocity is color-coded. We evaluate the azimuthal ray coverage of every grid cell following (Ekström, 2006), and the grid cells with low azimuthal ray coverage (i.e., low reliability) are eliminated by replacing the velocities in such grid cells with nan values. We apply damping and smoothing in the inversion through regularization. Our primary Vs model is generated by conducting 1D surface wave inversion on the dispersion curves of every pixel from

group velocity tomography, then evaluating the reference phase velocity from the primary Vs model for the subsequent phase velocity dispersion picking.

3.5 Phase Velocity Picking

We measured the phase velocity as an additional constraint for our Vs model in addition to the group velocity. We followed the method in (Yao et al., 2006) for the single-station-pair phase velocity dispersion measurement. Because of the relatively high-frequencies and large degree of lateral heterogeneity in the basins, an accurate reference phase velocity model is essential for the phase velocity dispersion picking. With a prior reference phase velocity model from the primary Vs model derived using only group velocity dispersion curves (red line in Figure S1), we measure the phase velocity dispersion curves for every available station pair (Figure S1). The final inversion for the Vs model incorporates both phase and group velocity dispersion curves for both Rayleigh and Love waves.

3.6 Initial Model

The inversion for the Vs model from dispersion curves is highly dependent on the initial model. We construct our initial model based on the prior basin depth (PBD) model from Villa et al. (2022) shown in Figure 5. The PBD model integrates multiple observations: receiver functions, Bouguer gravity, and borehole data. The receiver functions provide the sediment-basement interface beneath the dense arrays (Liu et al, 2018; Wang et al., 2021; Ghose et al, 2022), and the Bouguer gravity is used to extrapolate the basin depth determined along the seismic profiles into a 3-D model. Data from 17 boreholes are also used to calibrate and validate the 3-D basin depth model. Using the basin depth model, we construct an initial Vs model with $V_s = 0.3$ km/s at the surface and a linear increase with depth to $V_s = 2.3$ km/s at the basin bottom. In addition, the prior model also contains a low-velocity zone. The low-velocity zone is a prominent feature in the San Gabriel basin, associated with the shallow marine Fernando Formation (Brocher et al., 1998; West et al., 1988). The CVM-S 4.26 model (Lee et al., 2014) inherits the low-velocity feature from the CVMS1-3 models (Kohler et al., 2003; Magistrale et al., 1996) in which the SG area is based on borehole data and geological models. We preserve these low-velocity features present in the CVM-S 4.26 model as a prior feature in our initial model.

3.7 Vs Model

Our final Vs model combines the phase and group velocities of Rayleigh (ZZ) and Love (TT) waves. Both the fundamental modes and the 1st higher mode of the Rayleigh wave group velocity are included. The initial model used in the tomography includes the information from gravity, borehole data, receiver functions, and the CVMS 4.26 model. We use the SURF96 software (Herrmann, 2013) to conduct the S wave velocity inversion from the dispersion curves for each grid point. In the prior basin depth model, the conversion from travel time to depth is based on the velocity model, therefore the updated Vs model produces a new initial model with each updated basin depth. We iterate over the initial model and the Vs model until the Vs model converges (shows little change). We then merge our final Vs model on top of the CVMS 4.26 model in the region defined by the PBD model: the Vs above the depth of PBD model is from our Vs model, and deeper than 1 km below the PBD model, the Vs is taken from the CVMS

4.26, within the 1 km zone below the PBD model we use a linear weighting function to smooth the transition from our V_s model to the CVMS 4.26.

4 Results

We show the group velocity maps of the fundamental Rayleigh (Figure 6a, c, e) and Love (Figure 6b,d,f) waves at periods of 1, 2, and 3 s. The group velocity at different periods is sensitive to different depths, and $T = 1, 2, 3$ s have a typical sensitivity kernel covering ~0-1 km, ~0-1.5 km, and ~0-2.5 km depth, respectively. The group velocity dispersion is independent of the PBD model, reflecting features derived purely from the ambient noise data. Despite the large spatial variations, we see several features consistent with the PBD model. In the San Gabriel basin, a prominent low-velocity region from 1 to 3 s indicates a sedimentary basin with over 2.5 km depth. To the northwest of the San Gabriel basin, a sharp increase of group velocity for $T > 1$ s, indicates a transition from the 2.5 km deep San Gabriel basin to the shallower Raymond basin (< 1 km depth) near the Raymond fault. In the San Bernardino area, the group velocity map is more complex, but prominent features are evident, such as the Chino basin in the west and the San Bernardino basin in the east with relatively low group velocities, and the Jurupa Hills in the central south region with high group velocities.

The shear wave velocity (V_s) model is shown in Figure 7 at depths of 0.5, 1, 1.5, and 2 km. The spatial distribution of the low V_s regions (sedimentary basins) is similar to the group velocity maps, and variations of maximum depth within the sedimentary basins can be inferred from the V_s model: the Raymond basin is less than 1 km deep, the San Gabriel basin is deeper than 2 km, the Chino basin is around 1 km deep and the San Bernardino basin is between 1 km and 2 km deep. In addition to the ambient noise data, the V_s model is also dependent on the PBD model. In the following section, we discuss and compare the V_s model with the PBD model, as well as other basin depth models from geological cross-sections, other geophysical constraints, and borehole data.

5 Discussion

In this section, we compare our V_s model to several other independent observations to validate the robustness of the V_s model. The location of four cross-sections (black lines, AA' to EE') and three sonic boreholes well logs (red stars) are shown in Figure 1. The five cross-sections were analyzed in previous studies: AA' through the San Bernardino basin is from Stephenson et al. (2002), BB' is the cross-section in the Raymond basin from Buwalda (1940), CC' and DD' are cross-sections 14 and 15 in the San Gabriel basin (Davis and Namson, 2013, 2017) and EE' is the cross-section in Rialto-Colton basin from (Linda R. Woolfenden, 1997; Paulinski, 2012). The comparison of our V_s model with the PBD model (dashed lines) and models from other references (dotted lines) is shown in Figure 8.

In the San Bernardino basin, the structure between 10 and 20 km distance along the AA' profile was consistently constrained by seismic reflection data and gravity-aeromagnetic modeling (Stephenson et al., 2002). From south to north along AA', the sedimentary basin depth increases sharply to 1.7 km near the San Jacinto fault (~12 km from A) and slowly decreases after passing

the Loma Linda fault (~14 km from A). In our Vs model, a low-velocity structure shows a good correlation with the basin model from Stephenson et al. (2002), both laterally and in-depth. The Raymond basin, bounded by the Raymond fault on the southeast, is a relatively shallow basin compared to the adjacent San Gabriel basin. Based on gravity and borehole data, the BB' cross-section (Buwalda, 1940) constrains the central Raymond basin depth to ~1.5 km, slightly deeper than the low-velocity structure (~1 km deep) from our Vs model. Across the Raymond fault, the PBD model reveals a sharp transition from the ~1 km deep Raymond basin to the ~3 km deep San Gabriel basin, consistent with the conspicuous reduction of group velocity at the Raymond fault, which creates a sudden deepening of the low-velocity layer in the Vs profile at a distance of 16 km in BB' profile. The CC' and DD' profiles (Davis and Namson, 2013, 2017) constrain the depths and shapes of the western and eastern San Gabriel basin. In the CC' profile, the low-velocity layer shows a sharp decrease at 8 km from the start of the profile (C), coincident with the Whittier fault that offsets the sedimentary layer and basement rock in the geologic cross-section. DD' is a cross-section in the eastern part of the San Gabriel basin. The profile is bounded by the Whittier fault to the south and the Sierra Madre fault to the north. The Vs model only captures the Sierra Madre fault at the distance of ~22 km from the start of the profile (D), while in the south, the Whittier fault is located outside the Vs model coverage. The EE' profile cuts through the Rialto-Colton basin located northwest of the San Bernardino basin. In Figure 8e, the dotted line represents the base of the water-bearing layer (Linda R. Woolfenden, 1997; Paulinski, 2012) from resistivity logs. Due to the limited borehole depth, the base of the water-bearing layer is not necessarily equivalent to the sedimentary basin depth. Our velocity model overall predicts a low-velocity layer comparable to the water-bearing layer, but with a much larger variation in depth. However, the location of the Barrier J and (unnamed) fault Q (Anderson et al., 2004; Lu and Danskin, 2001) coincides with the boundary of the graben-like structure in our model. In the five cross-sections, AA' to EE', our Vs model agrees with the basin depth from other references, and the fault structures inferred from sharp lateral Vs gradients agree with the fault locations that offset the sedimentary layers.

The sonic velocity from well logs provides a ground truth of the velocity structure of the sedimentary layers. We compare our Vs model to three available sonic well logs (Fig. 9). One prominent feature in the sonic velocity from well logs is the low-velocity zones in the Ferris borehole at 1800 m and in Live Oak Park (LOP) borehole at 1200 m depth (locations shown on Figure 1). The low-velocity layer is associated with the Fernando formation, a ubiquitous marine layer in San Gabriel and Los Angeles basins that underlies the non-marine Duarte Conglomerate (Yeats, 2004). The low-velocity zone is also present in the CVMS 4.26 Vs model (Lee et al., 2014), as it was inherited from the prior CVMS model (Magistrale et al., 1996) and is based on the borehole data (Magistrale, 2000). In our Vs model, the prior model is a linear model based on the PBD model, and a low-velocity feature is incorporated in the prior model if it exists in the CVMS 4.26. Preservation of the low-velocity zone makes it consistent between the borehole data, CVMS 4.26 and our Vs model.

Besides the incorporation of the low-velocity zone from the CVMS 4.26 model as a prior feature in the initial model used in our inversion, the construction of the Vs model is mainly based on the PBD model and the dispersion curves, both of which are independent of the CVMS model. We compare a set of the group velocity dispersion curves in the San Gabriel basin predicted from our Vs model to the CVMS model prediction (Fig. 10). In the San Gabriel basin, both models predict slower group velocities in the south compared to the north, but overall, the dispersion

curve generated from CVMS 4.26 is faster than our measured dispersion curves. A direct comparison of our Vs model and CVMS 4.26 in different basin areas (Fig. 11) illustrates the difference between the models. In our model, we see the sedimentary basins are in general deeper with lower seismic velocities, and the variation of velocity with depth is always smoother than in the CVMS model.

6 Conclusion

We cross-correlate the ambient noise between 10 linear nodal arrays, SCSN broadband stations, 18 temporary broadband stations, and strong motion accelerometers. We obtain the Rayleigh wave and Love wave EGF from the ZZ and TT component ambient noise cross-correlation. In the dispersion analysis, the Rayleigh wave fundamental mode and first higher mode were separated using Rayleigh wave particle motion polarization. We constructed the Vs model by incorporating group and phase velocity tomography, and constraints from receiver functions and Bouguer gravity datasets. Our Vs model is consistent with geological and geophysical cross-sections from independent studies and the sonic borehole dataset in terms of basement depth and fault locations. Compared to the SCEC CVMS community model, our Vs model generally contains deeper and slower basin structures, especially in the San Bernardino area. This discrepancy might resolve the underestimation of ground motion predicted in future seismic wavefield simulations.

7. Model Product

The results of this study are designed to seamlessly fit into the CVMS4.26 model. They are available as a rectilinear block of shear wave velocities between longitude 116.90° W and 118.37° W, and latitude between 33.90° N and 34.25° N. Since the CVMS4.26 was used as the starting model, this block can be used as a direct replacement for the corresponding block in the CVMS4.26 model. This will increase the resolution and details in the San Gabriel, Chino, and San Bernardino basins without disturbing the CVMS4.26 model outside of these basins.

Appendix A: Instrumental response for seismogram to accelerometer correlation.

In the ambient noise correlation, the removal of instrumental response is unnecessary when the two stations have the same instrumental response. In the frequency domain, the correlation function $C_{XY(\omega)} = \frac{X(\omega)I(\omega)}{|X(\omega)I(\omega)|} \frac{\bar{Y}(\omega)\bar{I}(\omega)}{|\bar{Y}(\omega)\bar{I}(\omega)|}$ where $X(\omega)$, $Y(\omega)$ are the Fourier transformation of ambient noise waveform, $I(\omega)$ is the Fourier transformation of instrumental response, bar for conjugate, and the modulus in the denominator is due to spectrum whitening. As $\frac{I(\omega)}{|I(\omega)|} \frac{\bar{I}(\omega)}{|\bar{I}(\omega)|} = 1$, $C_{XY(\omega)} = \frac{X(\omega)}{|X(\omega)|} \frac{\bar{Y}(\omega)}{|\bar{Y}(\omega)|}$, so that the instrumental response has no effect on the dense array- dense array correlation.

For the correlation between dense array and accelerometer, however, the instrumental response causes a non-trivial phase lag. Assuming the station x is a seismogram, which records the velocity $x(t)$, and station y is an accelerometer recording the acceleration $dy(t)/dt$. The

correlation $C_{XY(\omega)} = \frac{X(\omega)}{|X(\omega)|} \frac{i\omega \bar{Y}(\omega)}{|i\omega \bar{Y}(\omega)|} = \frac{X(\omega)}{|X(\omega)|} \frac{i\bar{Y}(\omega)}{|\bar{Y}(\omega)|}$, where the $i\omega$ comes from the time-derivative operator. Assuming the causal ($t>0$) and anti-causal ($t<0$) branches of EGF are symmetric in the waveform, we fold the waveform at $t=0$ and stack the causal and anti-causal parts before dispersion analysis (Supplementary Fig S2.a). However, the correlation between seismogram and accelerometer has a $\pi/2$ phase shift due to the i in the frequency domain. The $\pi/2$ shift leads to a π (half period) shift when we fold the waveform at $t=0$, meaning a flip of sign between causal and anti-causal branches (Supplementary Fig S2.b). Therefore, we subtract the causal by the anti-causal branch to account for instrumental response when stacking the correlation function from seismogram-accelerometer cross-correlation.

Appendix B: ZZ and ZR phase difference is the same as Z and R phase difference.

In the dispersion analysis, we use the phase lag between the Z and R components in the path of Rayleigh wave propagation to quantify the Rayleigh wave particle motion. In this section, we show the phase difference between ZZ and ZR in ambient noise correlation is equivalent to the Z and R phase difference.

With a given noise source, signals in different components can be written as $s_1(t)=\cos(wt + \phi_1)$ and $s_2(t)=\cos(wt + \phi_2)$. For receivers, the recorded waveform from s_1 at station x is $x_1(t)=\cos(w(t - \frac{r_1}{c}) + \phi_1)$, and the recorded waveform from s_2 at station y is $y_1(t)=\cos(w(t - \frac{r_2}{c}) + \phi_2)$, where r_1 and r_2 are the distances from source to the two receivers and c is the velocity. The correlation between the two receivers is

$$C_{xy} = \frac{1}{2T} \int_{-T}^T \cos(w(\tau - \frac{r_1}{c}) + \phi_1) \cos(w(t + \tau - \frac{r_2}{c}) + \phi_2) d\tau$$

$$= \frac{1}{2} \cos(w(t - \frac{r_2-r_1}{c}) + \phi_2 - \phi_1) \quad (T \gg 1)$$

For ZZ correlation, $\phi_1 = \phi_2$; for ZR correlation, $\phi_1 = \phi_Z$, $\phi_2 = \phi_R$. $C_{ZR} = \frac{1}{2} \cos(w(t - \frac{r_2-r_1}{c}) + \phi_R - \phi_Z)$, $C_{ZZ} = \frac{1}{2} \cos(w(t - \frac{r_2-r_1}{c}))$. Therefore, we proved the phase difference between ZZ and ZR is $\phi_Z - \phi_R$, equal to the phase difference between the Z and R components of the source.

Acknowledgments

The authors thank Zhe Jia for help with the ambient noise correction code and helpful discussions. The authors thank all the volunteers who helped with deploying the dense nodal arrays. This research was supported by the National Science Foundation awards 2105358 and 210532. The BASIN project was partly supported by U.S. Geological Survey awards GS17AP00002 and G19AP00015, and Southern California Earthquake Center awards 18029 and 19033. Nodal instruments were provided by Incorporated Research Institutions for Seismology (IRIS), Portable Array Seismic Studies of the Continental Lithosphere (PASSCAL), University of Utah, Louisiana State University, and the University of Oklahoma.

Data Availability

The final Vs model can be downloaded from <http://doi.org/10.22002/D1.20248>. All the node and temporary broadband data used in the study are scheduled to be available at the IRIS DMC by the end of 2022. The permanent strong motion and broad data are available from the Southern California Earthquake Data Center (SCEDC).

References

- Anderson, M., Matti, J., Jachens, R., 2004. Structural model of the San Bernardino basin, California, from analysis of gravity, aeromagnetic, and seismicity data: STRUCTURE OF THE SAN BERNARDINO BASIN. *J. Geophys. Res. Solid Earth* 109. <https://doi.org/10.1029/2003JB002544>
- 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 broadband surface wave dispersion measurements. *Geophys. J. Int.* 169, 1239–1260. <https://doi.org/10.1111/j.1365-246X.2007.03374.x>
- Brocher, T.M., Ruebel, A.L., Wright, T.L., Okaya, D.A., 1998. Compilation of 20 sonic and density logs from 12 oil test wells along LARSE Lines 1 and 2, Los Angeles region, California. *US Geol Surv Open-File Rept* 98 366, 53.
- Buwalda, J.P., 1940. *Geology of the Raymond Basin*.
- Castellanos, J.C., Clayton, R.W., 2021. The Fine-Scale Structure of Long Beach, California, and Its Impact on Ground Motion Acceleration. *J. Geophys. Res. Solid Earth* 126. <https://doi.org/10.1029/2021JB022462>
- Clayton, R., Persaud, P., Denolle, M., Polet, J., 2019. Exposing Los Angeles's Shaky Geologic Underbelly. *Eos* 100. <https://doi.org/10.1029/2019EO135099>
- Davis, T., Namson, J., 2013. Southern California Cross Section Study Map Showing 2012 AAPG Annual Mtg Field Trip Stops.
- Davis, T.L., Namson, J.S., 2017. Field excursion: Petroleum traps and structures along the San Andreas convergent strike-slip plate boundary, California. *AAPG Bull.* 101, 607–615. <https://doi.org/10.1306/011817DIG17040>
- Denolle, M.A., Dunham, E.M., Prieto, G.A., Beroza, G.C., 2014. Strong Ground Motion Prediction Using Virtual Earthquakes. *Science* 343, 399–403. <https://doi.org/10.1126/science.1245678>
- Ekström, G., 2006. Global detection and location of seismic sources by using surface waves. *Bull. Seismol. Soc. Am.* 96, 1201–1212.
- Graves, R., Jordan, T.H., Callaghan, S., Deelman, E., Field, E., Juve, G., Kesselman, C., Maechling, P., Mehta, G., Milner, K., Okaya, D., Small, P., Vahi, K., 2011. CyberShake: A Physics-Based Seismic Hazard Model for Southern California. *Pure Appl. Geophys.* 168, 367–381. <https://doi.org/10.1007/s00024-010-0161-6>

- Herrmann, R.B., 2013. Computer programs in seismology: An evolving tool for instruction and research. *Seismol. Res. Lett.* 84, 1081–1088.
- Jia, Z., Clayton, R.W., 2021. Determination of Near Surface Shear-Wave Velocities in the Central Los Angeles Basin With Dense Arrays. *J. Geophys. Res. Solid Earth* 126, e2020JB021369.
- Jones, L.M., Bernknopf, R., Cox, D., Goltz, J., Hudnut, K., Mileti, D., Perry, S., Ponti, D., Porter, K., Reichle, M., Seligson, H., Shoaf, K., Treiman, J., Wein, A., 2008. The ShakeOut Scenario (Open-File Report No. 2008–1150), USGS Numbered Series. U.S. Geological Survey, Reston, VA.
- Kohler, M.D., Magistrale, H., Clayton, R.W., 2003. Mantle heterogeneities and the SCEC reference three-dimensional seismic velocity model version 3. *Bull. Seismol. Soc. Am.* 93, 757–774.
- Lee, E.-J., Chen, P., Jordan, T.H., Maechling, P.B., Denolle, M.A., Beroza, G.C., 2014. Full-3-D tomography for crustal structure in southern California based on the scattering-integral and the adjoint-wavefield methods. *J. Geophys. Res. Solid Earth* 119, 6421–6451.
- Lin, F.-C., Li, D., Clayton, R.W., Hollis, D., 2013. High-resolution 3D shallow crustal structure in Long Beach, California: Application of ambient noise tomography on a dense seismic array. *GEOPHYSICS* 78, Q45–Q56. <https://doi.org/10.1190/geo2012-0453.1>
- Lin, F.-C., Moschetti, M.P., Ritzwoller, M.H., 2008. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps. *Geophys. J. Int.* 173, 281–298. <https://doi.org/10.1111/j.1365-246X.2008.03720.x>
- Linda R. Woolfenden, D.K., 1997. Geohydrology and water chemistry in the Rialto-Colton Basin, San Bernardino County, California (USGS Numbered Series), Water-Resources Investigations Report. <https://doi.org/10.3133/wri974012>
- Liu, G., Persaud, P., Clayton, R.W., 2018. Structure of the Northern Los Angeles Basins Revealed in Teleseismic Receiver Functions from Short-Term Nodal Seismic Arrays. *Seismol. Res. Lett.* 89, 1680–1689. <https://doi.org/10.1785/0220180071>
- Lu, Z., Danskin, W.R., 2001. InSAR analysis of natural recharge to define structure of a ground-water basin, San Bernardino, California. *Geophys. Res. Lett.* 28, 2661–2664. <https://doi.org/10.1029/2000GL012753>
- Ma, Y., Clayton, R.W., 2016. Structure of the Los Angeles Basin from ambient noise and receiver functions. *Geophys. J. Int.* 206, 1645–1651. <https://doi.org/10.1093/gji/ggw236>
- Ma, Y., Clayton, R.W., Li, D., 2016. Higher-mode ambient-noise Rayleigh waves in sedimentary basins. *Geophys. J. Int.* 206, 1634–1644.
- Magistrale, H., 2000. The SCEC Southern California Reference Three-Dimensional Seismic Velocity Model Version 2. *Bull. Seismol. Soc. Am.* 90, S65–S76. <https://doi.org/10.1785/0120000510>
- Magistrale, H., McLaughlin, K., Day, S., 1996. A geology-based 3D velocity model of the Los Angeles basin sediments. *Bull. Seismol. Soc. Am.* 86, 1161–1166.
- Mao, S., Mordret, A., Campillo, M., Fang, H., van der Hilst, R.D., 2020. On the measurement of seismic traveltimes changes in the time–frequency domain with wavelet cross-spectrum analysis. *Geophys. J. Int.* 221, 550–568.
- Olsen, K.B., Day, S.M., Minster, J.B., Cui, Y., Chourasia, A., Faerman, M., Moore, R., Maechling, P., Jordan, T., 2006. Strong shaking in Los Angeles expected from southern San Andreas earthquake. *Geophys. Res. Lett.* 33, L07305. <https://doi.org/10.1029/2005GL025472>

- Paulinski, S., 2012. Structural, hydrogeologic framework, and textural model of the Rialto-Colton Basin and the Chino and North Riverside area. Dr. Diss. Calif. State Univ. Sacram.
- Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. *Phys. Rev. E* 69, 046610. <https://doi.org/10.1103/PhysRevE.69.046610>
- Stephenson, W.J., Odum, J.K., Williams, R.A., Anderson, M.L., 2002. Delineation of faulting and basin geometry along a seismic reflection transect in urbanized San Bernardino Valley, California. *Bull. Seismol. Soc. Am.* 92, 2504–2520.
- Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 79, 61–78.
- Wang, X., Zhan, Z., Zhong, M., Persaud, P., Clayton, R.W., 2021. Urban basin structure imaging based on dense arrays and bayesian array-based coherent receiver functions. *J. Geophys. Res. Solid Earth* 126, e2021JB022279.
- West, J.C., Redin, T.W., Manings, G.C., Bartling, W.A., Green, K.E., 1988. Correlation section across Los Angeles basin from Palos Verdes Hills to San Gabriel Mountains. *Correl. Sect. CS* 3R.
- Yao, H., van Der Hilst, R.D., De Hoop, M.V., 2006. Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis—I. Phase velocity maps. *Geophys. J. Int.* 166, 732–744.
- Yeats, R.S., 2004. Tectonics of the San Gabriel Basin and surroundings, southern California. *Geol. Soc. Am. Bull.* 116, 1158. <https://doi.org/10.1130/B25346.1>

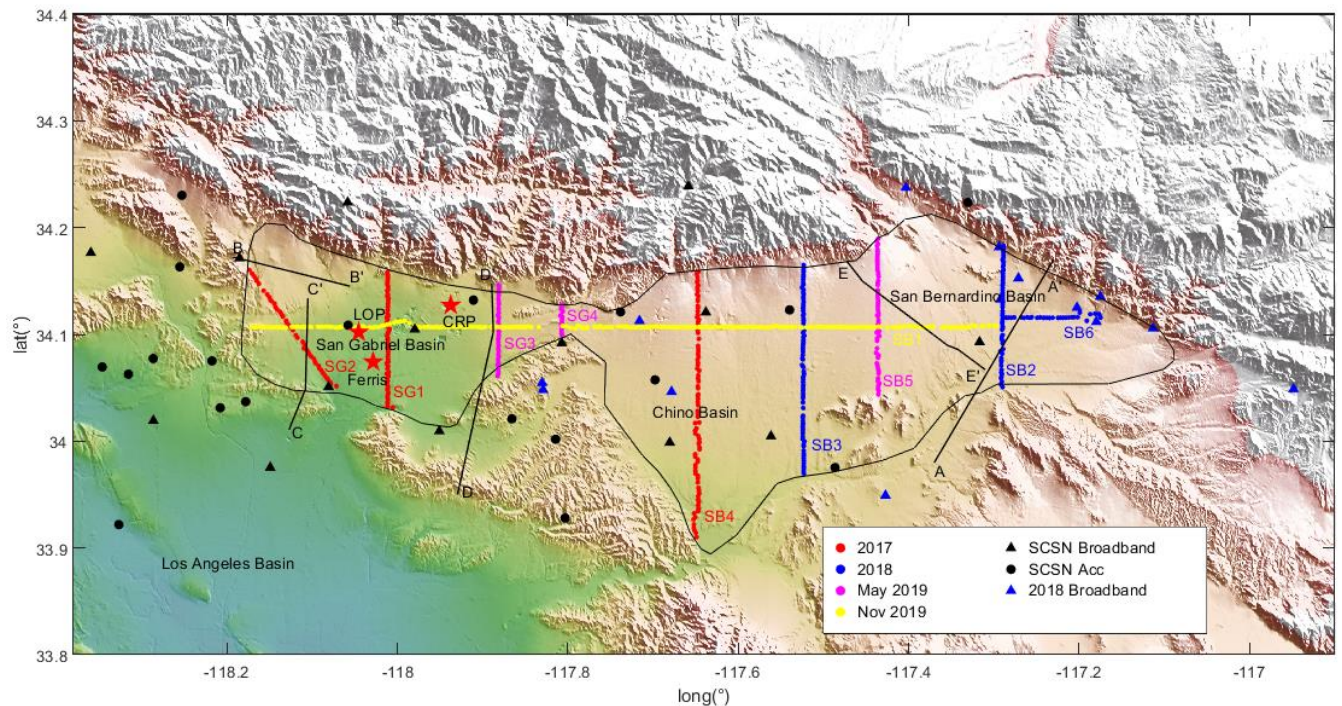


Figure 1. Distribution of BASIN nodal arrays (colored dots), broadband stations (black and blue triangles), and SCSN accelerometers (black dots). Color represents the deployment time for the temporary node stations. Black triangles are the permanent Southern California Seismic Network (SCSN) stations. Black lines (AA' to EE') are geological cross-sections, and red stars in the San Gabriel basin are borehole well logs.

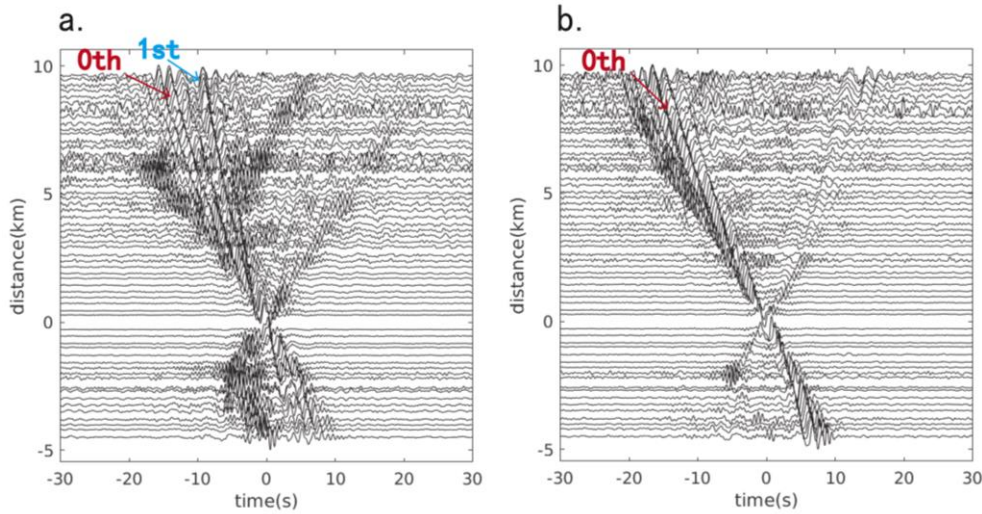


Figure 2. Intra-array correlation function from the SG1 dense linear array. (a) The ZZ component depicts Rayleigh waves. (b) The TT component with virtual source SG120. Correlation functions are filtered between 0.2 and 2 Hz.

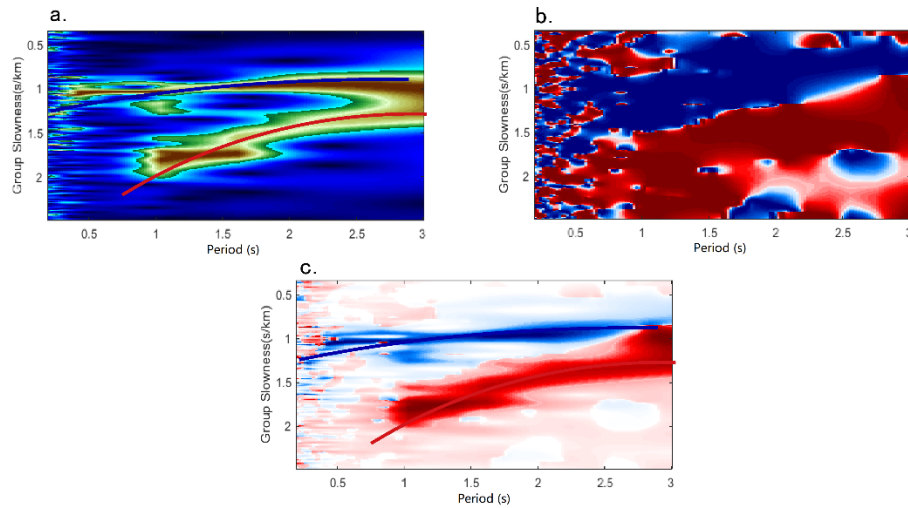


Figure 3. An example of Rayleigh wave group velocity dispersion analysis in the frequency-time domain. a) Hilbert transform of the ZZ correlation function. b) Phase difference δ between ZR and ZZ from the Wavelet transform. Red for δ between $[0, \pi]$, retrograde particle motion. Blue for δ between $[-\pi, 0]$, prograde particle motion. c) Combination of a) and b). Red and blue lines are inferred retrograde fundamental mode and prograde first higher mode dispersion curves. The correlation is from station pair SG102-SG160.

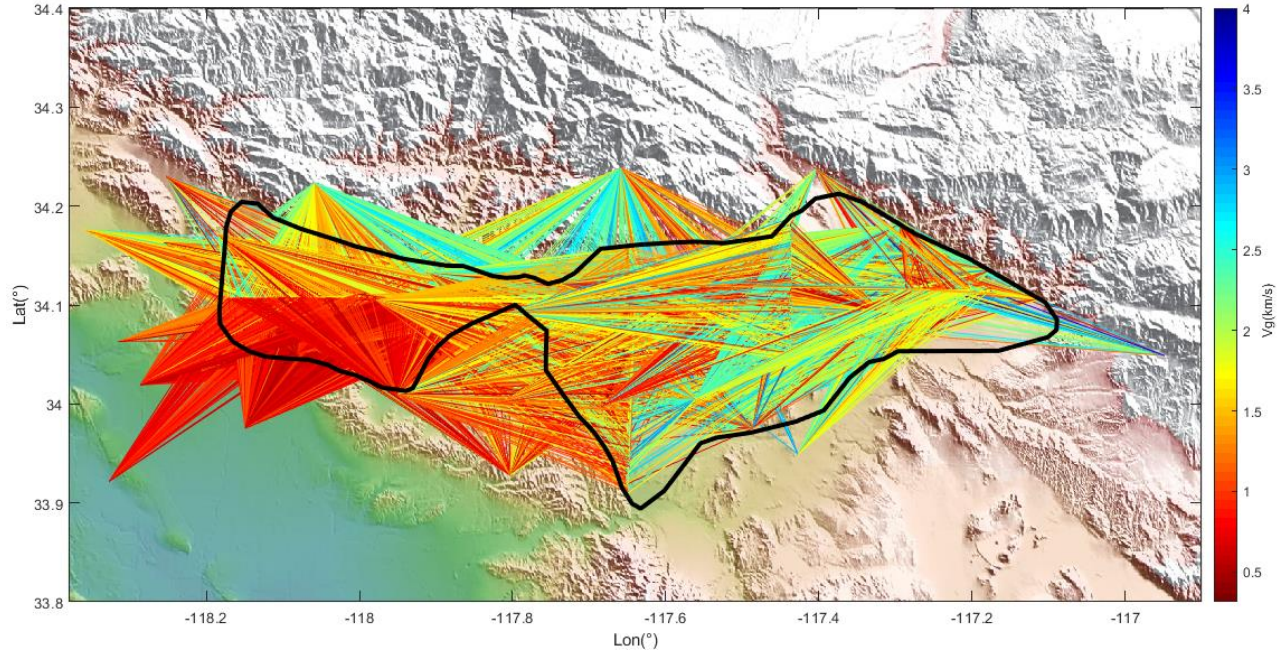


Figure 4. Ray coverage of Rayleigh wave fundamental mode group velocity at $T=1s$.

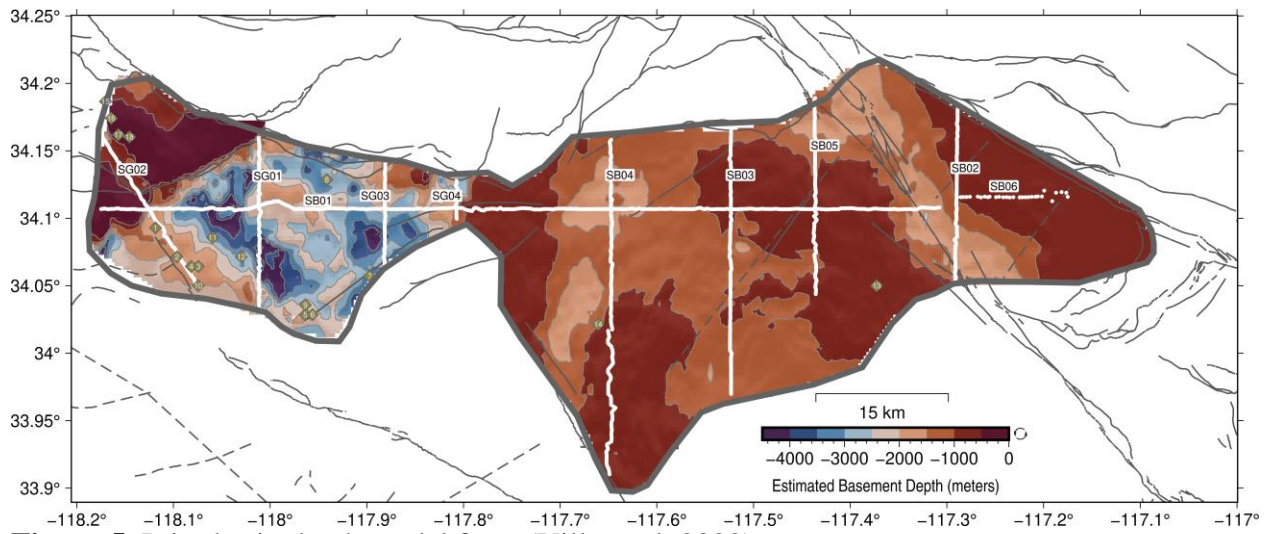


Figure 5. Prior basin depth model from (Villa et al, 2022)

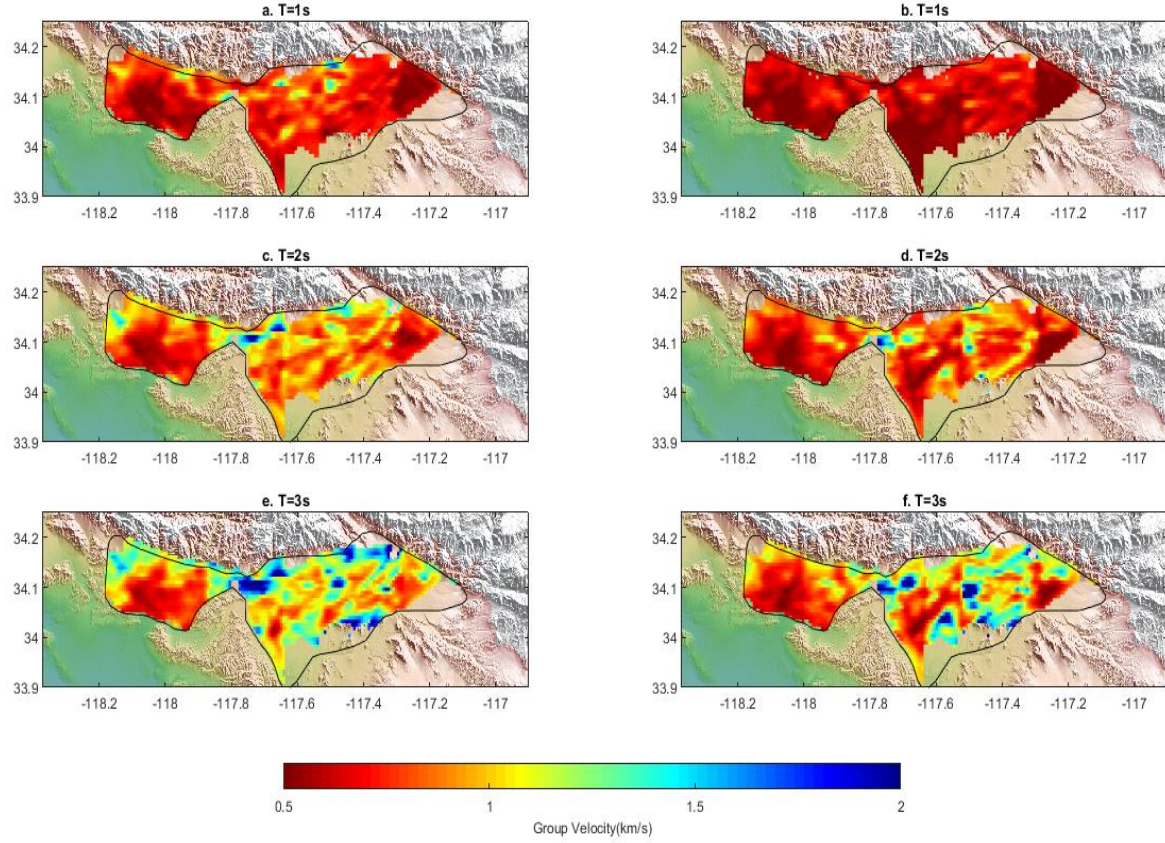


Figure 6. Group velocity maps for Rayleigh wave (a, c, e) and Love wave (b, d, f) group velocity models at $T=1$, 2, 3s.

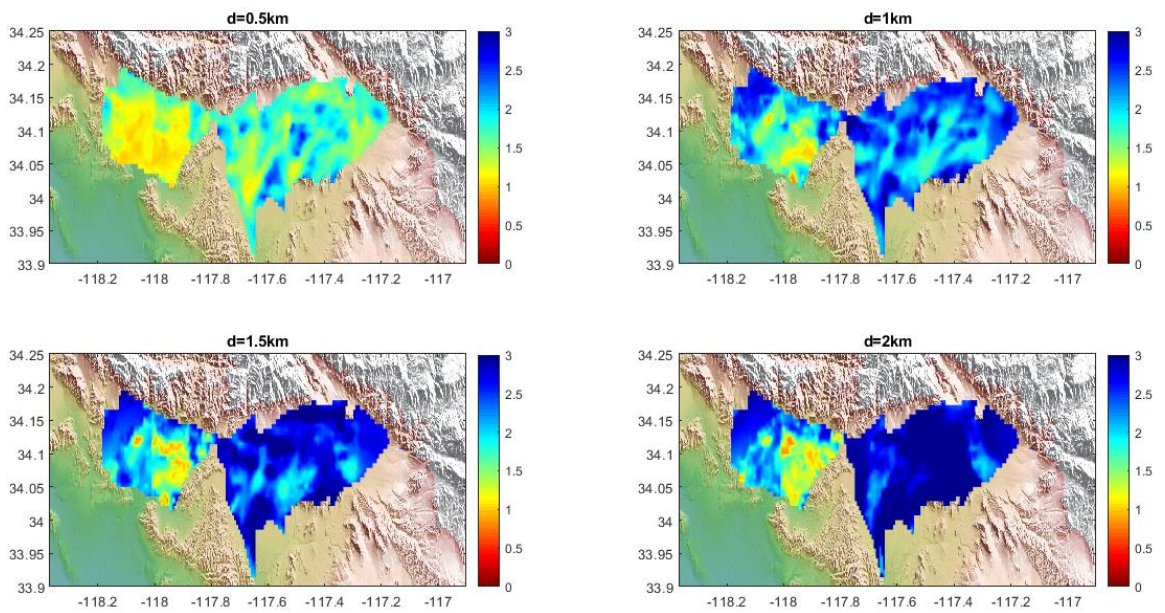
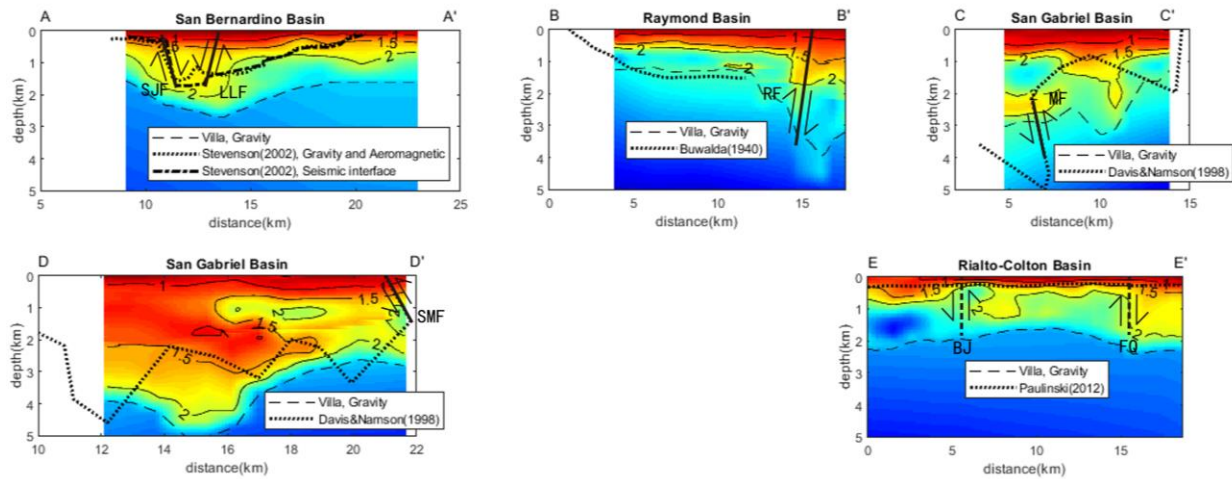
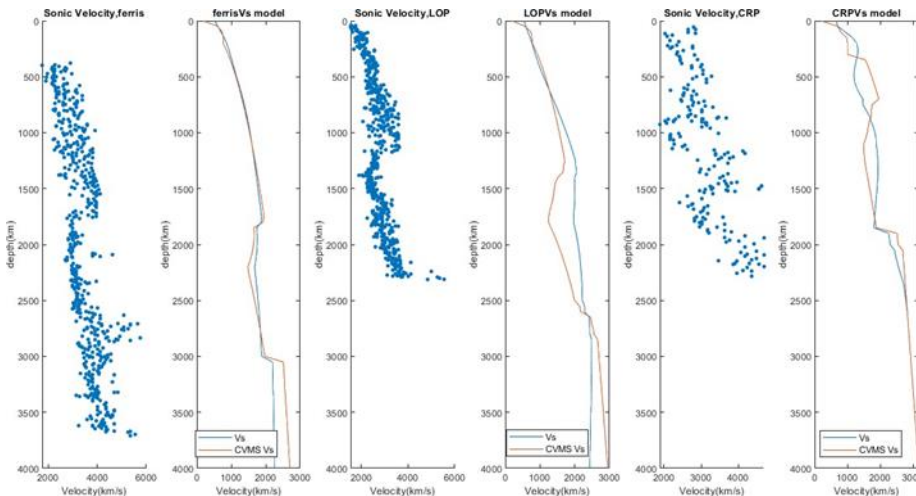


Figure 7. Vs model at the depths of 0.5, 1, 1.5, and 2 km.**Figure 8.** Cross-sections of Vs model compared against prior basin model (dashed line) and basin model constraint from other references (dotted line). Locations of the cross sections are shown in Figure 1 with black lines. Abbreviations for faults: SJF-San Jacinto fault; LLF-Loma Linda fault; RF-Raymond fault; SMF-Sierra Madre fault; BJ-Barrier J; FQ- fault Q.**Figure 9.** Sonic well logs from the Ferris, LOP, and CRP boreholes compared with our Vs model (blue) and CVMS 4.26 (orange). Locations of the boreholes are shown with red stars in Figure 1.

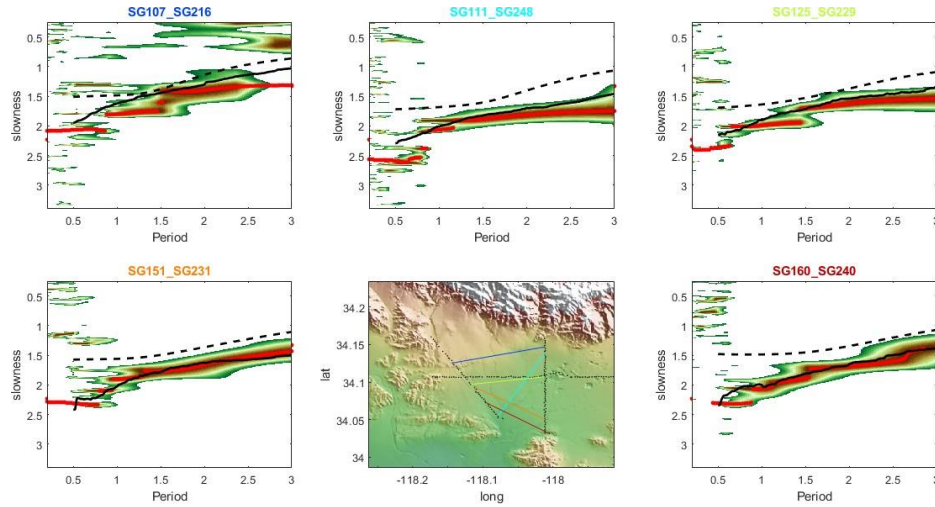


Figure 10. Love wave group velocity dispersion curves in the San Gabriel basin predicted by our model (black solid line) and CVMS model (black dashed line). The background is the envelope from the correlation function, and the red curves are the actual picks.

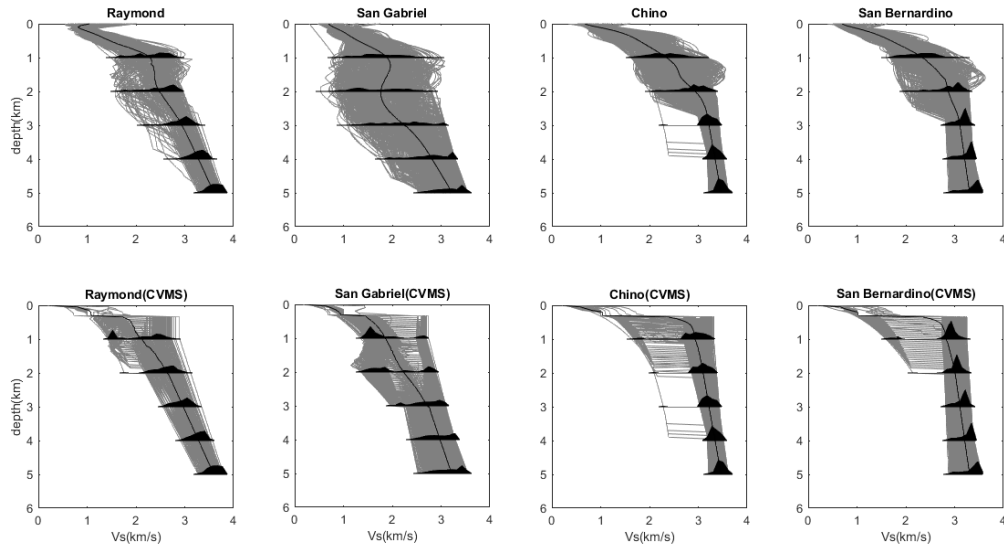


Figure 11. Compilation and distribution of Vs with depth in the Raymond, San Gabriel, Chino, and San Bernardino basins (gray lines) from our Vs model (upper panels) and CVMS 4.26 (lower panels). The black shaded regions show the distribution of Vs values at different depths.