

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

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

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

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

5 Corresponding author: Yida Li (ydali@caltech.edu)

6 **Key Points:**

- 7 • We construct a 3D Vs model in San Gabriel and San Bernardino basins using ambient
8 noise correlation between dense array nodal, broadband and accelerometer stations.
- 9 • We separated the Rayleigh wave fundamental mode and first higher mode in dispersion
10 analysis based on the Rayleigh wave particle motion.
- 11 • Our Vs model predicts deeper and slower sedimentary basins than the SCEC CVMS
12 model, yet is consistent with geological and drilling data in these basins.
13

14 **Abstract**

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

29 **Plain Language Summary**

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

40

41 **1 Introduction**

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

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

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

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

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

89 Compared to the crustal-scale survey using the long-period ambient noise correlation between
90 broadband stations, the surface wave EGF in high-frequency ambient noise correlation is less
91 coherent due to the greater structural variations in sedimentary basins. In recent years, the
92 deployment of dense seismic arrays makes it possible to resolve the fine-scale velocity structure
93 of the top 5 km sedimentary layer (Castellanos and Clayton, 2021; Jia and Clayton, 2021; Lin et
94 al., 2013). In addition to the ambient noise correlation, receiver functions are also evaluated from
95 the dense array datasets to constrain the basement depth within the sedimentary basins (Ma and
96 Clayton, 2016; Liu et al, 2018; Wang et al., 2021). Receiver function using our linear dense
97 arrays has shown a coherent converted phase at the basin bottom can be observed in the SG and
98 SB area, which provide an independent constraint on the basin structure in this area (Liu et al.,
99 2018; Wang et al., 2021).

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

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

109

110 **2 Data**

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

123 **3 Method**

124 **3.1 Ambient noise correlation**

125 To estimate the shear wave velocity, we first determine the EGF between each station using
126 ambient noise correlations. The noise correlation follows the technique described in Bensen et al.
127 (2007) and Jia and Clayton (2021). To reduce the influence of anthropogenic noise, we
128 correlate only the nighttime (8:00 pm to 8:00 am, local time) ambient noise. We include all the
129 possible ray pairs, including node to node, node to broadband, and node to accelerometer, that
130 have overlapping recording times. The data are correlated in one-hour segments and stacked to
131 get the final correlation. To minimize the effect of earthquakes and broaden the effective period
132 range, we do time domain normalization and spectral whitening prior to the correlation. For the
133 node-to-node correlation, as the stations of every pair have the same instrument response, it
134 cancels out in the spectral whitening, and therefore the removal of instrument response was not
135 required in our case. For the node-to-accelerometer correlation, we will show that a $-\frac{\pi}{2}$ and $\frac{\pi}{2}$
136 phase shift is introduced because of the difference in the instrument response, and special care
137 should be taken when stacking the causal and anti-causal Green's function (Appendix A). As all
138 of our stations are 3-component, we can extract both Rayleigh and Love waves. We rotate the
139 components from the ZNE into the ZRT coordinate system. The Rayleigh wave particle motion
140 is in the RZ plane and the Love wave particle motion is mainly in the T direction, and hence we
141 correlate the Z component of the virtual source and virtual receiver, called the ZZ correlation, to
142 get the Rayleigh wave EGF, and use the TT correlation to get Love wave EGF. In Figures 2 and

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

149 **3.2 Group Velocity Dispersion Picking**

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

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

171 **3.3 Rayleigh Wave Mode Separation**

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

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

$$183 \quad W_x(s, n) = \left(\frac{\delta t}{s}\right)^{1/2} \sum_{n'=1}^N x_{n'} \Phi_0^* \left[\frac{(n' - n)\delta t}{s} \right]$$

184 where Φ_0^* is the wavelet function (Torrence and Compo, 1998), s is the wavelet scale, and δt is
 185 the time step. As with the Fourier transform, the variation of s gives a spectral pattern in the
 186 frequency domain, but the wavelet transform also has an additional dimension, n that reflects the
 187 temporal variation. The wavelet transformation has been proven to be a powerful technique to
 188 monitor temporal variation in the coda with high precision (Mao et al., 2020). Here, we use it to
 189 evaluate the phase difference between the ZZ and ZR correlations, and when combined with the
 190 the Hilbert transform it produces a clear separation of the fundamental from the first higher
 191 mode. We apply the CWT using the Matlab Wavelet Toolbox with the Morse wavelet function
 192 Φ_0^* . The phase difference between the ZR and ZZ correlations is $\delta = \arg(W_{ZR}(s, n)) - \arg(W_{ZZ}$
 193 $(s, n))$. For a retrograde fundamental mode, this is $\pi/2$. In contrast, for the prograde first-higher
 194 mode $\delta = -\pi/2$. We plot $\sin \delta$ in Figure 3b to quantify the polarization of particle motion in the
 195 group slowness and period domain, where red and blue are positive ($\delta = \pi/2$) and negative ($\delta = -$
 196 $\pi/2$) phase shifts, respectively, corresponding to retrograde and prograde particle motion.
 197 However, as the $\sin \delta$ pattern only represents the phase difference, noise and signal are not
 198 distinguishable in this representation. To combine the amplitude and phase information, we
 199 multiply the wave envelope from the Hilbert transformation in Figure 3a and the $\sin \delta$ in Figure
 200 3b to produce the result shown in Figure 3c. In Figure 3c, red representing the retrograde
 201 fundamental ($\sin \delta = 1$) mode and blue representing the prograde first higher mode ($\sin \delta = -1$)
 202 are clearly separated. Our picking of the Rayleigh wave group velocity dispersion curve is based
 203 on this pattern.

204 **3.4 Tomography**

205 With the measured dispersion curves, we applied the straight-ray tomography method to invert
 206 the frequency dependent group velocity maps. We discretize the area into a uniform grid with
 207 0.55 km longitudinal spacing and 0.66 km latitudinal spacing. The group velocity tomography is
 208 carried out between 0.5 s and 3 s period, using the travel times from the dispersion curves. Figure
 209 4 shows an example of the straight ray coverage of group velocity at period $T=1$ s, where the
 210 picked group velocity is color-coded. We evaluate the azimuthal ray coverage of every grid cell
 211 following (Ekström, 2006), and the grid cells with low azimuthal ray coverage (i.e., low
 212 reliability) are eliminated by replacing the velocities in such grid cells with nan values. We apply
 213 damping and smoothing in the inversion through regularization. Our primary Vs model is
 214 generated by conducting 1D surface wave inversion on the dispersion curves of every pixel from

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

217 **3.5 Phase Velocity Picking**

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

227 **3.6 Initial Model**

228 The inversion for the V_s model from dispersion curves is highly dependent on the initial model.
229 We construct our initial model based on the prior basin depth (PBD) model from [Villa et al.](#)
230 (2022) shown in Figure 5. The PBD model integrates multiple observations: receiver functions,
231 Bouguer gravity, and borehole data. The receiver functions provide the sediment-basement
232 interface beneath the dense arrays (Liu et al, 2018; Wang et al., 2021; [Ghose et al, 2022](#)), and the
233 Bouguer gravity is used to extrapolate the basin depth determined along the seismic profiles into
234 a 3-D model. Data from 17 boreholes are also used to calibrate and validate the 3-D basin depth
235 model. Using the basin depth model, we construct an initial V_s model with $V_s = 0.3$ km/s at the
236 surface and a linear increase with depth to $V_s = 2.3$ km/s at the basin bottom. In addition, the
237 prior model also contains a low-velocity zone. The low-velocity zone is a prominent feature in
238 the San Gabriel basin, associated with the shallow marine Fernando Formation (Brocher et al.,
239 1998; West et al., 1988). The CVM-S 4.26 model (Lee et al., 2014) inherits the low-velocity
240 feature from the CVMS1-3 models (Kohler et al., 2003; Magistrale et al., 1996) in which the SG
241 area is based on borehole data and geological models. We preserve these low-velocity features
242 present in the CVM-S 4.26 model as a prior feature in our initial model.

243 **3.7 V_s Model**

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

255 4.26, within the 1 km zone below the PBD model we use a linear weighting function to smooth
 256 the transition from our V_S model to the CVMS 4.26.

257 **4 Results**

258 We show the group velocity maps of the fundamental Rayleigh (Figure 6a, c, e) and Love
 259 (Figure 6b,d,f) waves at periods of 1, 2, and 3 s. The group velocity at different periods is
 260 sensitive to different depths, and $T = 1, 2, 3$ s have a typical sensitivity kernel covering ~0-1 km,
 261 ~0-1.5 km, and ~0-2.5 km depth, respectively. The group velocity dispersion is independent of
 262 the PBD model, reflecting features derived purely from the ambient noise data. Despite the large
 263 spatial variations, we see several features consistent with the PBD model. In the San Gabriel
 264 basin, a prominent low-velocity region from 1 to 3 s indicates a sedimentary basin with over 2.5
 265 km depth. To the northwest of the San Gabriel basin, a sharp increase of group velocity for $T >$
 266 1s, indicates a transition from the 2.5km deep San Gabriel basin to the shallower Raymond basin
 267 (< 1 km depth) near the Raymond fault. In the San Bernardino area, the group velocity map is
 268 more complex, but prominent features are evident, such as the Chino basin in the west and the
 269 San Bernardino basin in the east with relatively low group velocities, and the Jurupa Hills in the
 270 central south region with high group velocities.

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

280

281 **5 Discussion**

282 In this section, we compare our V_S model to several other independent observations to validate
 283 the robustness of the V_S model. The location of four cross-sections (black lines, AA' to EE') and
 284 three sonic boreholes well logs (red stars) are shown in Figure 1. The five cross-sections were
 285 analyzed in previous studies: AA' through the San Bernardino basin is from Stephenson et al.
 286 (2002), BB' is the cross-section in the Raymond basin from Buwalda (1940), CC' and DD' are
 287 cross-sections 14 and 15 in the San Gabriel basin (Davis and Namson, 2013, 2017) and EE' is
 288 the cross-section in Rialto-Colton basin from (Linda R. Woolfenden, 1997; Paulinski, 2012) The
 289 comparison of our V_S model with the PBD model (dashed lines) and models from other
 290 references (dotted lines) is shown in Figure 8.

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

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

322 The sonic velocity from well logs provides a ground truth of the velocity structure of the
323 sedimentary layers. We compare our Vs model to three available sonic well logs (Fig. 9). One
324 prominent feature in the sonic velocity from well logs is the low-velocity zones in the Ferris
325 borehole at 1800 m and in Live Oak Park (LOP) borehole at 1200 m depth (locations shown on
326 Figure 1). The low-velocity layer is associated with the Fernando formation, a ubiquitous marine
327 layer in San Gabriel and Los Angeles basins that underlies the non-marine Duarte Conglomerate
328 (Yeats, 2004). The low-velocity zone is also present in the CVMS 4.26 Vs model (Lee et al.,
329 2014), as it was inherited from the prior CVMS model (Magistrale et al., 1996) and is based on
330 the borehole data (Magistrale, 2000). In our Vs model, the prior model is a linear model based on
331 the PBD model, and a low-velocity feature is incorporated in the prior model if it exists in the
332 CVMS 4.26. Preservation of the low-velocity zone makes it consistent between the borehole
333 data, CVMS 4.26 and our Vs model.

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

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

345 **6 Conclusion**

346 We cross-correlate the ambient noise between 10 linear nodal arrays, SCSN broadband stations,
 347 18 temporary broadband stations, and strong motion accelerometers. We obtain the Rayleigh
 348 wave and Love wave EGF from the ZZ and TT component ambient noise cross-correlation. In
 349 the dispersion analysis, the Rayleigh wave fundamental mode and first higher mode were
 350 separated using Rayleigh wave particle motion polarization. We constructed the Vs model by
 351 incorporating group and phase velocity tomography, and constraints from receiver functions and
 352 Bouguer gravity datasets. Our Vs model is consistent with geological and geophysical cross-
 353 sections from independent studies and the sonic borehole dataset in terms of basement depth and
 354 fault locations. Compared to the SCEC CVMS community model, our Vs model generally
 355 contains deeper and slower basin structures, especially in the San Bernardino area. This
 356 discrepancy might resolve the underestimation of ground motion predicted in future seismic
 357 wavefield simulations.

358 **7. Model Product**

359 The results of this study are designed to seamlessly fit into the CVMS4.26 model. They are
 360 available as a rectilinear block of shear wave velocities between longitude 116.90° W and
 361 118.37° W, and latitude between 33.90° N and 34.25° N. Since the CVMS4.26 was used as
 362 the starting model, this block can be used as a direct replacement for the corresponding block in
 363 the CVMS4.26 model. This will increase the resolution and details in the San Gabriel, Chino,
 364 and San Bernardino basins without disturbing the CVMS4.26 model outside of these basins.

365 **Appendix A: Instrumental response for seismogram to accelerometer correlation.**

366 In the ambient noise correlation, the removal of instrumental response is unnecessary when the
 367 two stations have the same instrumental response. In the frequency domain, the correlation
 368 function $C_{XY(\omega)} = \frac{X(\omega)I(\omega) \bar{Y}(\omega)\bar{I}(\omega)}{|X(\omega)I(\omega)| |Y(\omega)\bar{I}(\omega)|}$ where $X(\omega)$, $Y(\omega)$ are the Fourier transformation of ambient
 369 noise waveform, $I(\omega)$ is the Fourier transformation of instrumental response, bar for conjugate,
 370 and the modulus in the denominator is due to spectrum whitening. As $\frac{I(\omega) \bar{I}(\omega)}{|I(\omega)| |I(\omega)|} = \mathbf{1}$, $C_{XY(\omega)} =$
 371 $\frac{X(\omega) \bar{Y}(\omega)}{|X(\omega)| |Y(\omega)|}$, so that the instrumental response has no effect on the dense array- dense array
 372 correlation.

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

376 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
 377 operator. Assuming the causal ($t>0$) and anti-causal ($t<0$) branches of EGF are symmetric in the
 378 waveform, we fold the waveform at $t=0$ and stack the causal and anti-causal parts before
 379 dispersion analysis (Supplementary Fig S2.a). However, the correlation between seismogram and
 380 accelerometer has a $\pi/2$ phase shift due to the i in the frequency domain. The $\pi/2$ shift leads to
 381 a π (half period) shift when we fold the waveform at $t=0$, meaning a flip of sign between causal
 382 and anti-causal branches (Supplementary Fig S2.b). Therefore, we subtract the causal by the anti-
 383 causal branch to account for instrumental response when stacking the correlation function from
 384 seismogram-accelerometer cross-correlation.

385

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

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

391 With a given noise source, signals in different components can be written as $s_1(t)=\cos (wt +$
 392 $\phi_1)$ and $s_2(t)=\cos (wt + \phi_2)$. For receivers, the recorded waveform from s_1 at station x is $x_1(t)$
 393 $=\cos (w(t - \frac{r_1}{c}) + \phi_1)$, and the recorded waveform from s_2 at station y is $y_1(t)=\cos (w(t -$
 394 $\frac{r_2}{c}) + \phi_2)$, where r_1 and r_2 are the distances from source to the two receivers and c is the
 395 velocity. The correlation between the two receivers is

$$396 \quad C_{xy} = \frac{1}{2T} \int_{-T}^T \cos(w(\tau - \frac{r_1}{c}) + \phi_1) \cos(w(t + \tau - \frac{r_2}{c}) + \phi_2) d\tau$$

$$397 \quad = \frac{1}{2} \cos(w(t - \frac{r_2-r_1}{c}) + \phi_2 - \phi_1) \quad (T \gg 1)$$

398 For ZZ correlation, $\phi_1 = \phi_2$; for ZR correlation, $\phi_1 = \phi_Z, \phi_2 = \phi_R$. $C_{ZR} = \frac{1}{2} \cos(w(t -$
 399 $\frac{r_2-r_1}{c}) + \phi_R - \phi_Z)$, $C_{ZZ} = \frac{1}{2} \cos(w(t - \frac{r_2-r_1}{c}))$. Therefore, we proved the phase difference
 400 between ZZ and ZR is $\phi_Z - \phi_R$, equal to the phase difference between the Z and R components
 401 of the source.

402

403 **Acknowledgments**

404 The authors thank Zhe Jia for help with the ambient noise correction code and helpful
 405 discussions. The authors thank all the volunteers who helped with deploying the dense nodal
 406 arrays. This research was supported by the National Science Foundation awards 2105358 and
 407 210532. The BASIN project was partly supported by U.S. Geological Survey awards
 408 GS17AP00002 and G19AP00015, and Southern California Earthquake Center awards 18029 and
 409 19033. Nodal instruments were provided by Incorporated Research Institutions for Seismology
 410 (IRIS), Portable Array Seismic Studies of the Continental Lithosphere (PASSCAL), University
 411 of Utah, Louisiana State University, and the University of Oklahoma.

412

413 **Data Availability**

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

418

419

420

421 **References**

- 422 Anderson, M., Matti, J., Jachens, R., 2004. Structural model of the San Bernardino basin,
423 California, from analysis of gravity, aeromagnetic, and seismicity data: STRUCTURE
424 OF THE SAN BERNARDINO BASIN. *J. Geophys. Res. Solid Earth* 109.
425 <https://doi.org/10.1029/2003JB002544>
- 426 Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro,
427 N.M., Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-
428 band surface wave dispersion measurements. *Geophys. J. Int.* 169, 1239–1260.
429 <https://doi.org/10.1111/j.1365-246X.2007.03374.x>
- 430 Brocher, T.M., Ruebel, A.L., Wright, T.L., Okaya, D.A., 1998. Compilation of 20 sonic and
431 density logs from 12 oil test wells along LARSE Lines 1 and 2, Los Angeles region,
432 California. US Geol Surv Open-File Rept 98 366, 53.
- 433 Buwalda, J.P., 1940. Geology of the Raymond Basin.
- 434 Castellanos, J.C., Clayton, R.W., 2021. The Fine-Scale Structure of Long Beach, California, and
435 Its Impact on Ground Motion Acceleration. *J. Geophys. Res. Solid Earth* 126.
436 <https://doi.org/10.1029/2021JB022462>
- 437 Clayton, R., Persaud, P., Denolle, M., Polet, J., 2019. Exposing Los Angeles's Shaky Geologic
438 Underbelly. *Eos* 100. <https://doi.org/10.1029/2019EO135099>
- 439 Davis, T., Namson, J., 2013. Southern California Cross Section Study Map Showing 2012
440 AAPG Annual Mtg Field Trip Stops.
- 441 Davis, T.L., Namson, J.S., 2017. Field excursion: Petroleum traps and structures along the San
442 Andreas convergent strike-slip plate boundary, California. *AAPG Bull.* 101, 607–615.
443 <https://doi.org/10.1306/011817DIG17040>
- 444 Denolle, M.A., Dunham, E.M., Prieto, G.A., Beroza, G.C., 2014. Strong Ground Motion
445 Prediction Using Virtual Earthquakes. *Science* 343, 399–403.
446 <https://doi.org/10.1126/science.1245678>
- 447 Ekström, G., 2006. Global detection and location of seismic sources by using surface waves.
448 *Bull. Seismol. Soc. Am.* 96, 1201–1212.
- 449 Graves, R., Jordan, T.H., Callaghan, S., Deelman, E., Field, E., Juve, G., Kesselman, C.,
450 Maechling, P., Mehta, G., Milner, K., Okaya, D., Small, P., Vahi, K., 2011. CyberShake:
451 A Physics-Based Seismic Hazard Model for Southern California. *Pure Appl. Geophys.*
452 168, 367–381. <https://doi.org/10.1007/s00024-010-0161-6>

- 453 Herrmann, R.B., 2013. Computer programs in seismology: An evolving tool for instruction and
 454 research. *Seismol. Res. Lett.* 84, 1081–1088.
- 455 Jia, Z., Clayton, R.W., 2021. Determination of Near Surface Shear-Wave Velocities in the
 456 Central Los Angeles Basin With Dense Arrays. *J. Geophys. Res. Solid Earth* 126,
 457 e2020JB021369.
- 458 Jones, L.M., Bernknopf, R., Cox, D., Goltz, J., Hudnut, K., Mileti, D., Perry, S., Ponti, D.,
 459 Porter, K., Reichle, M., Seligson, H., Shoaf, K., Treiman, J., Wein, A., 2008. The
 460 ShakeOut Scenario (Open-File Report No. 2008–1150), USGS Numbered Series. U.S.
 461 Geological Survey, Reston, VA.
- 462 Kohler, M.D., Magistrale, H., Clayton, R.W., 2003. Mantle heterogeneities and the SCEC
 463 reference three-dimensional seismic velocity model version 3. *Bull. Seismol. Soc. Am.*
 464 93, 757–774.
- 465 Lee, E.-J., Chen, P., Jordan, T.H., Maechling, P.B., Denolle, M.A., Beroza, G.C., 2014. Full-3-D
 466 tomography for crustal structure in southern California based on the scattering-integral
 467 and the adjoint-wavefield methods. *J. Geophys. Res. Solid Earth* 119, 6421–6451.
- 468 Lin, F.-C., Li, D., Clayton, R.W., Hollis, D., 2013. High-resolution 3D shallow crustal structure
 469 in Long Beach, California: Application of ambient noise tomography on a dense seismic
 470 array. *GEOPHYSICS* 78, Q45–Q56. <https://doi.org/10.1190/geo2012-0453.1>
- 471 Lin, F.-C., Moschetti, M.P., Ritzwoller, M.H., 2008. Surface wave tomography of the western
 472 United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps.
 473 *Geophys. J. Int.* 173, 281–298. <https://doi.org/10.1111/j.1365-246X.2008.03720.x>
- 474 Linda R. Woolfenden, D.K., 1997. Geohydrology and water chemistry in the Rialto-Colton
 475 Basin, San Bernardino County, California (USGS Numbered Series), Water-Resources
 476 Investigations Report. <https://doi.org/10.3133/wri974012>
- 477 Liu, G., Persaud, P., Clayton, R.W., 2018. Structure of the Northern Los Angeles Basins
 478 Revealed in Teleseismic Receiver Functions from Short-Term Nodal Seismic Arrays.
 479 *Seismol. Res. Lett.* 89, 1680–1689. <https://doi.org/10.1785/0220180071>
- 480 Lu, Z., Danskin, W.R., 2001. InSAR analysis of natural recharge to define structure of a ground-
 481 water basin, San Bernardino, California. *Geophys. Res. Lett.* 28, 2661–2664.
 482 <https://doi.org/10.1029/2000GL012753>
- 483 Ma, Y., Clayton, R.W., 2016. Structure of the Los Angeles Basin from ambient noise and
 484 receiver functions. *Geophys. J. Int.* 206, 1645–1651. <https://doi.org/10.1093/gji/ggw236>
- 485 Ma, Y., Clayton, R.W., Li, D., 2016. Higher-mode ambient-noise Rayleigh waves in sedimentary
 486 basins. *Geophys. J. Int.* 206, 1634–1644.
- 487 Magistrale, H., 2000. The SCEC Southern California Reference Three-Dimensional Seismic
 488 Velocity Model Version 2. *Bull. Seismol. Soc. Am.* 90, S65–S76.
 489 <https://doi.org/10.1785/0120000510>
- 490 Magistrale, H., McLaughlin, K., Day, S., 1996. A geology-based 3D velocity model of the Los
 491 Angeles basin sediments. *Bull. Seismol. Soc. Am.* 86, 1161–1166.
- 492 Mao, S., Mordret, A., Campillo, M., Fang, H., van der Hilst, R.D., 2020. On the measurement of
 493 seismic traveltimes changes in the time–frequency domain with wavelet cross-spectrum
 494 analysis. *Geophys. J. Int.* 221, 550–568.
- 495 Olsen, K.B., Day, S.M., Minster, J.B., Cui, Y., Chourasia, A., Faerman, M., Moore, R.,
 496 Maechling, P., Jordan, T., 2006. Strong shaking in Los Angeles expected from southern
 497 San Andreas earthquake. *Geophys. Res. Lett.* 33, L07305.
 498 <https://doi.org/10.1029/2005GL025472>

499 Paulinski, S., 2012. Structural, hydrogeologic framework, and textural model of the Rialto-
 500 Colton Basin and the Chino and North Riverside area. Dr. Diss. Calif. State Univ.
 501 Sacram.

502 Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: A
 503 derivation based on stationary phase. *Phys. Rev. E* 69, 046610.
 504 <https://doi.org/10.1103/PhysRevE.69.046610>

505 Stephenson, W.J., Odum, J.K., Williams, R.A., Anderson, M.L., 2002. Delineation of faulting
 506 and basin geometry along a seismic reflection transect in urbanized San Bernardino
 507 Valley, California. *Bull. Seismol. Soc. Am.* 92, 2504–2520.

508 Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. *Bull. Am. Meteorol.*
 509 *Soc.* 79, 61–78.

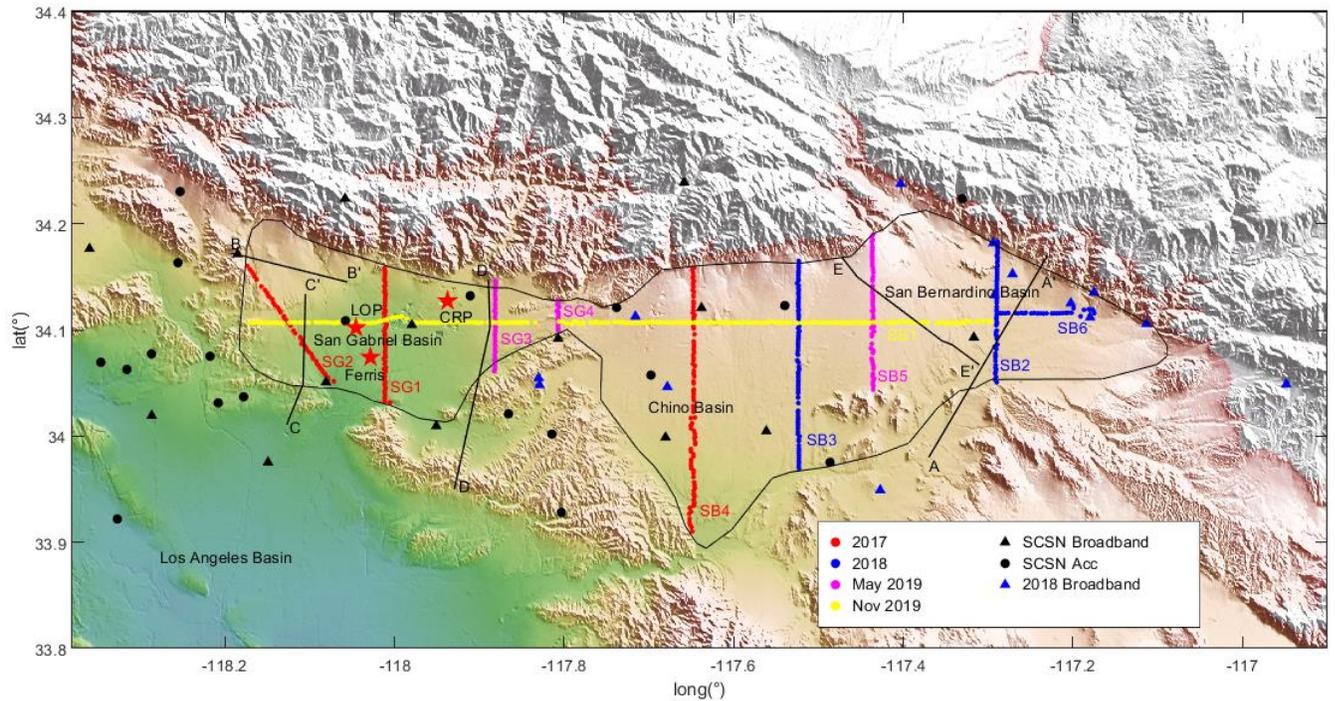
510 Wang, X., Zhan, Z., Zhong, M., Persaud, P., Clayton, R.W., 2021. Urban basin structure imaging
 511 based on dense arrays and bayesian array-based coherent receiver functions. *J. Geophys.*
 512 *Res. Solid Earth* 126, e2021JB022279.

513 West, J.C., Redin, T.W., Manings, G.C., Bartling, W.A., Green, K.E., 1988. Correlation section
 514 across Los Angeles basin from Palos Verdes Hills to San Gabriel Mountains. *Correl.*
 515 *Sect. CS* 3R.

516 Yao, H., van Der Hilst, R.D., De Hoop, M.V., 2006. Surface-wave array tomography in SE Tibet
 517 from ambient seismic noise and two-station analysis—I. Phase velocity maps. *Geophys.*
 518 *J. Int.* 166, 732–744.

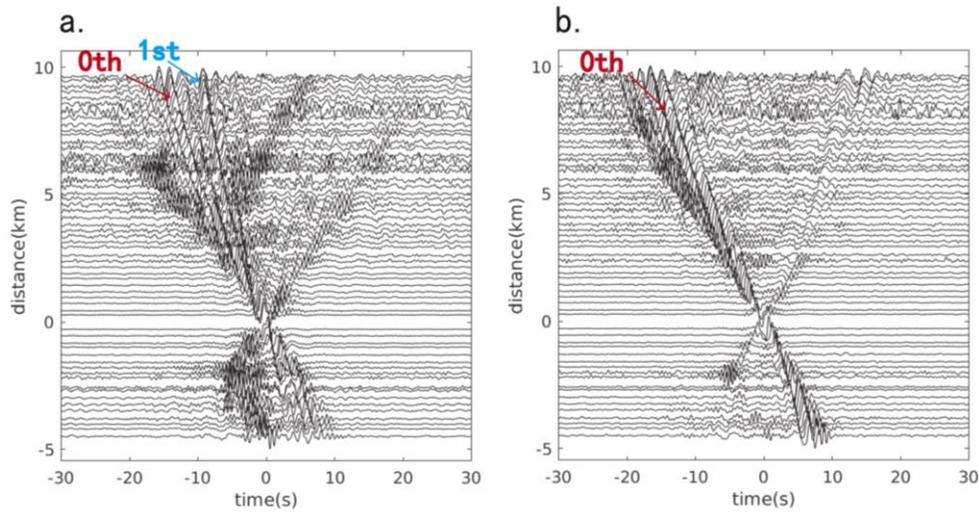
519 Yeats, R.S., 2004. Tectonics of the San Gabriel Basin and surroundings, southern California.
 520 *Geol. Soc. Am. Bull.* 116, 1158. <https://doi.org/10.1130/B25346.1>

522

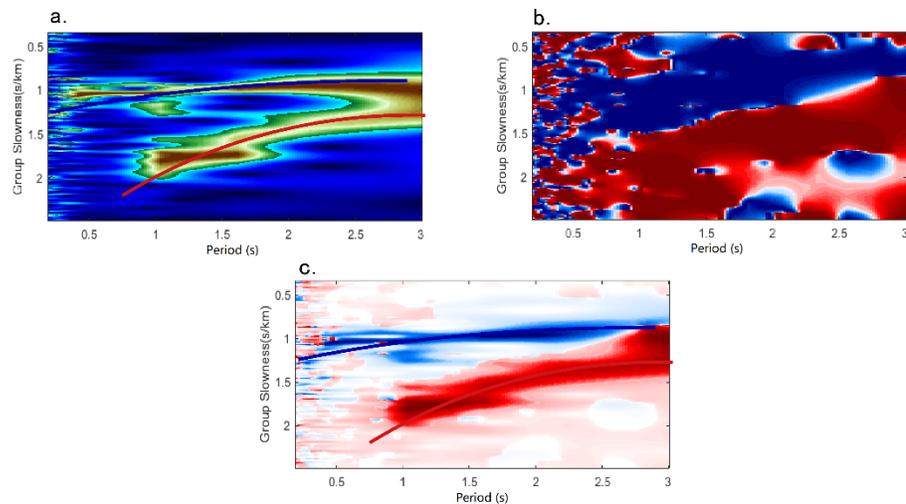


523

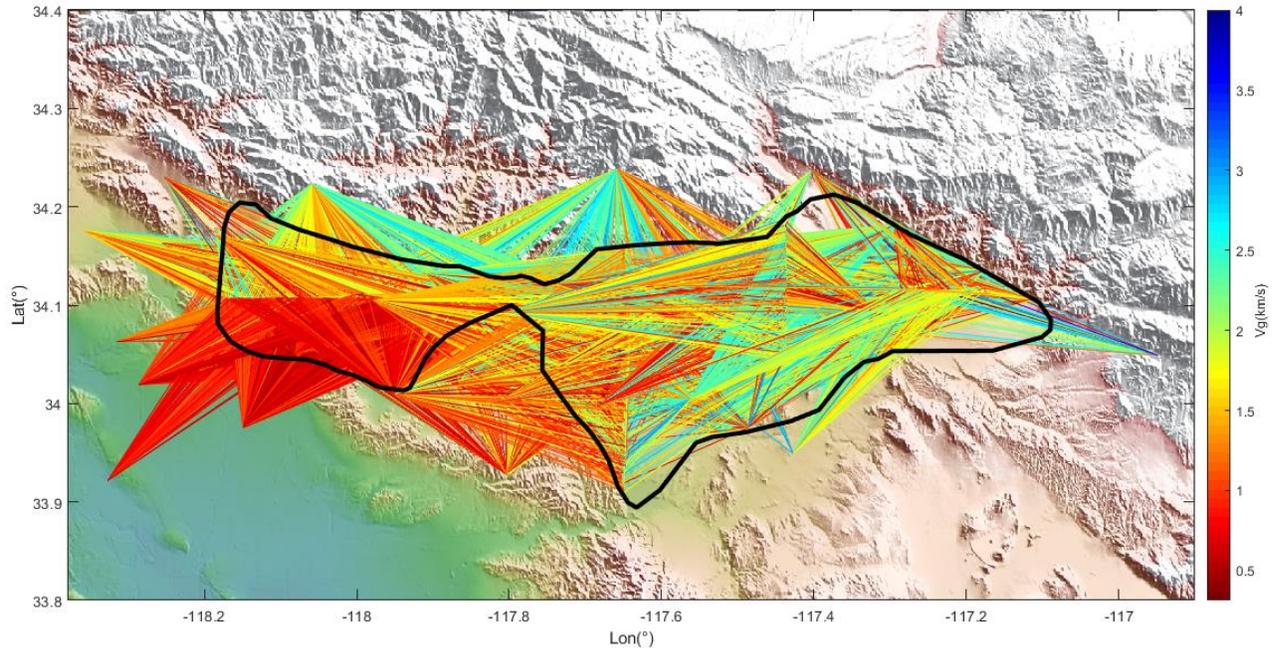
524 **Figure 1.** Distribution of BASIN nodal arrays (colored dots), broadband stations (black and blue
 525 triangles), and SCSN accelerometers (black dots). Color represents the deployment time for the
 526 temporary node stations. Black triangles are the permanent Southern California Seismic Network
 527 (SCSN) stations. Black lines (AA' to EE') are geological cross-sections, and red stars in the San
 528 Gabriel basin are borehole well logs.
 529



530
 531 **Figure 2.** Intra-array correlation function from the SG1 dense linear array. (a) The ZZ
 532 component depicts Rayleigh waves. (b) The TT component with virtual source SG120.
 533 Correlation functions are filtered between 0.2 and 2 Hz.

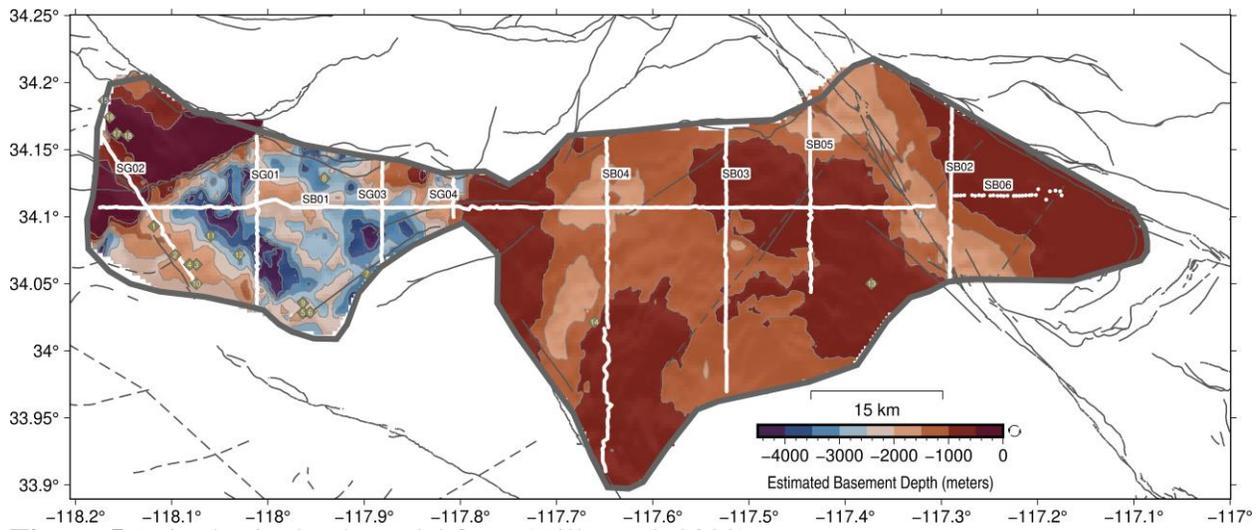


534
 535 **Figure 3.** An example of Rayleigh wave group velocity dispersion analysis in the frequency-time
 536 domain. a) Hilbert transform of the ZZ correlation function. b) Phase difference δ between ZR
 537 and ZZ from the Wavelet transform. Red for δ between $[0, \pi]$, retrograde particle motion. Blue
 538 for δ between $[-\pi, 0]$, prograde particle motion. c) Combination of a) and b). Red and blue lines
 539 are inferred retrograde fundamental mode and prograde first higher mode dispersion curves. The
 540 correlation is from station pair SG102-SG160.
 541

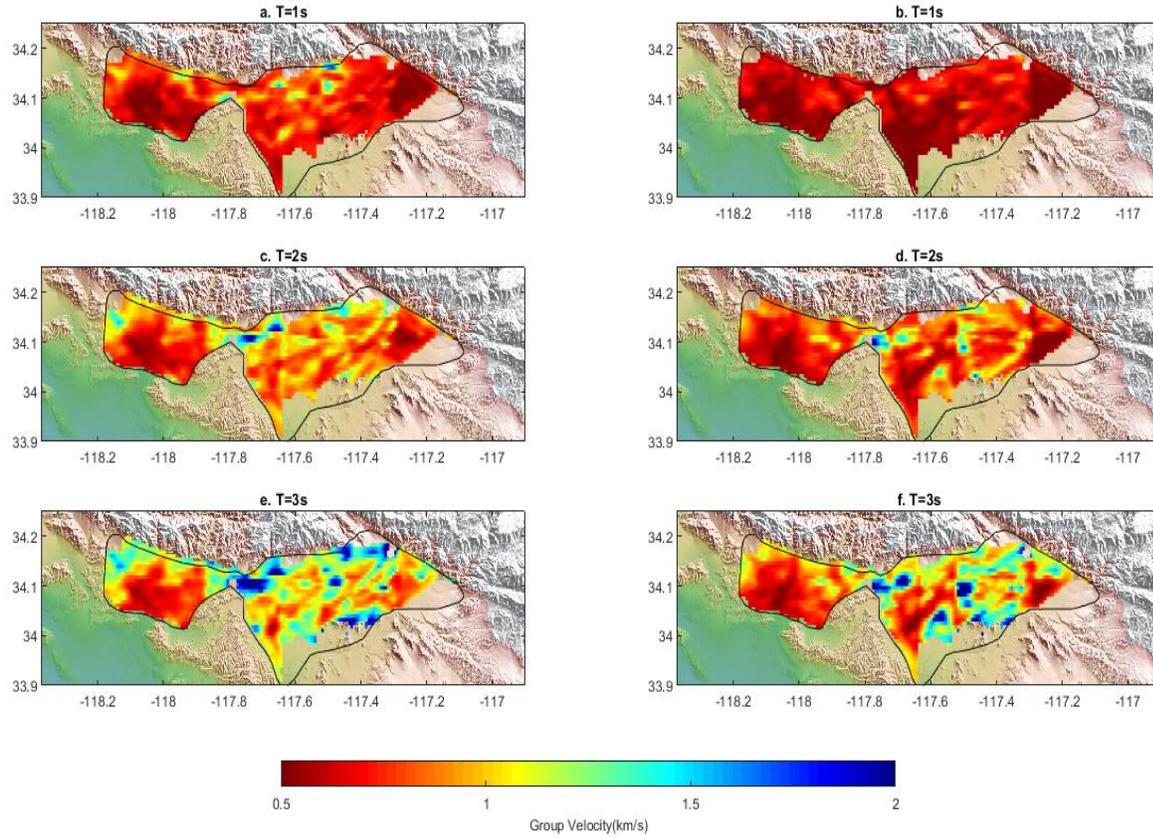


542
543 **Figure 4.** Ray coverage of Rayleigh wave fundamental mode group velocity at T=1s.
544

545

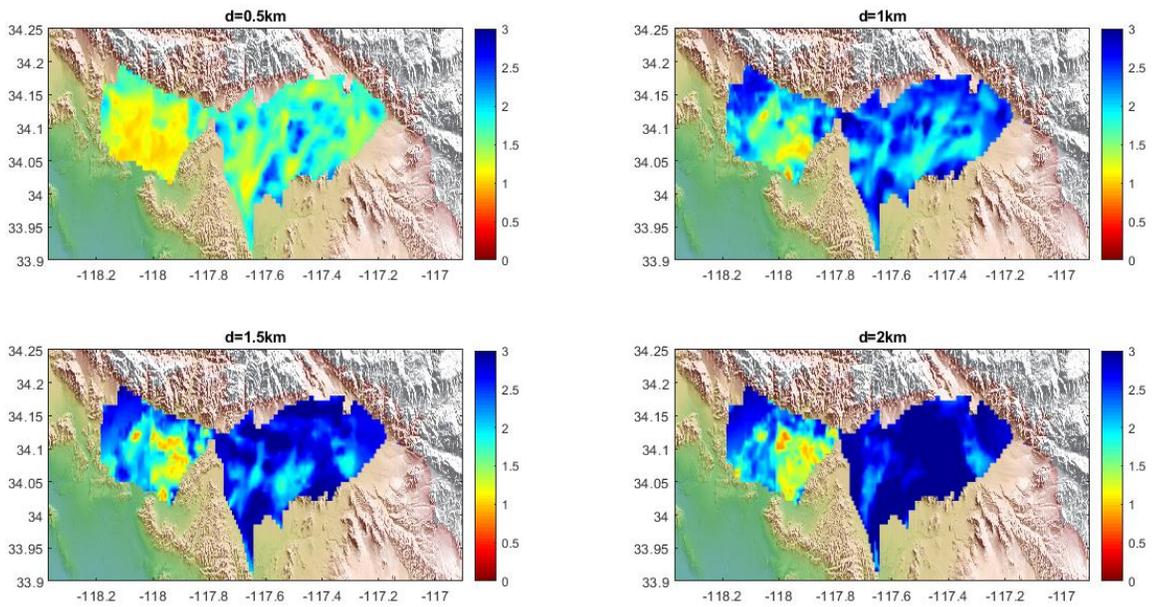


546
547 **Figure 5.** Prior basin depth model from (Villa et al, 2022)
548



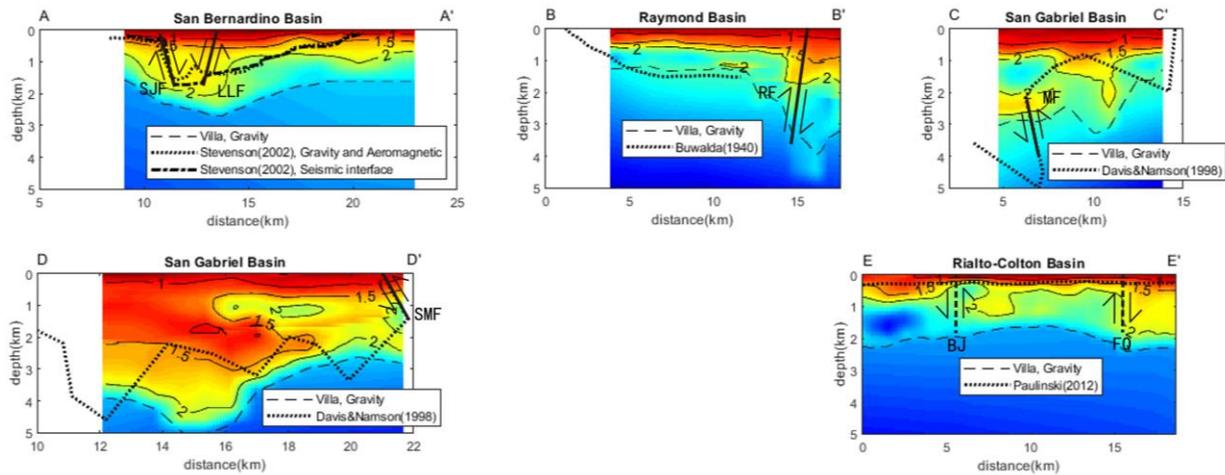
549
550
551
552
553
554
555

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.

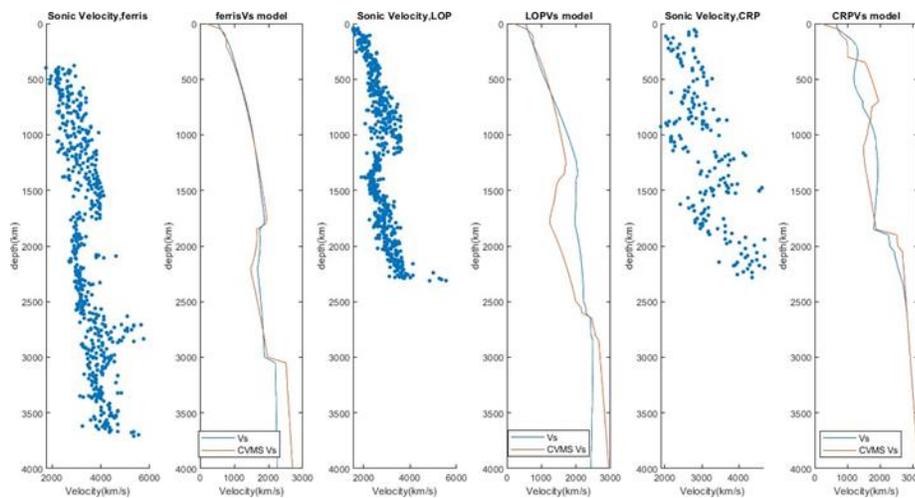


556

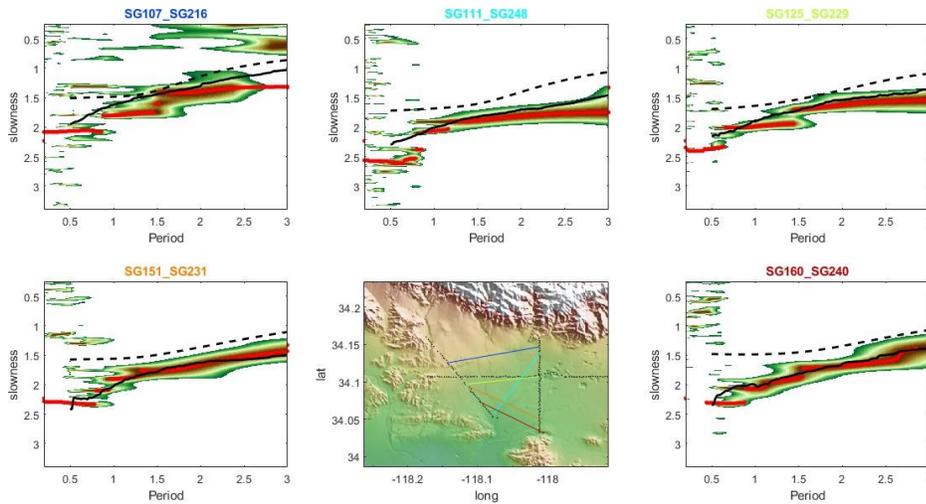
557 **Figure 7.** Vs model at the depths of 0.5, 1, 1.5, and 2 km.
558



559 **Figure 8.** Cross-sections of Vs model compared against prior basin model (dashed line) and
560 basin model constraint from other references (dotted line). Locations of the cross sections are
561 shown in Figure 1 with black lines. Abbreviations for faults: SJF-San Jacinto fault; LLF-Loma
562 Linda fault; RF-Raymond fault; SMF-Sierra Madre fault; BJ-Barrier J; FQ- fault Q.
563
564
565
566
567
568

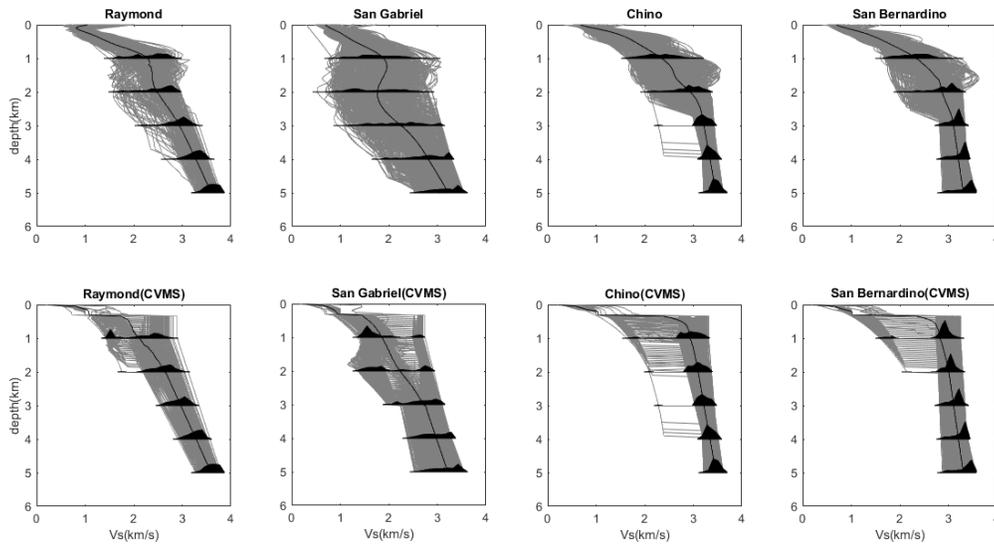


569 **Figure 9.** Sonic well logs from the Ferris, LOP, and CRP boreholes compared with our Vs
570 model (blue) and CVMS 4.26 (orange). Locations of the boreholes are shown with red stars in
571 Figure 1.
572
573
574



575
576
577
578
579
580
581
582
583
584

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.



585
586
587
588
589
590
591
592
593

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.