Experiment Full-name: **Long Valley Caldera Full Waveform Tomography**

Experiment Nickname: **LVCFWI**

Reference:

Flinders, A. F., Shelly, D. R., Dawson, P. B., Hill, D. P., Tripoli, B. and Shen, Y. Seismic evidence for significant melt beneath the Long Valley Caldera, California, USA, *Geology*, (2018).

Ashton F. Flinders[1*], David R. Shelly[1], Philip B. Dawson[1], David P. Hill[1], Barbara Tripoli[2], Yang Shen[3]

[1] *U.S Geological Survey, California Volcano Observatory, Menlo Park, California, 94025*
[2] *Department of Earth and Planetary Science, University of California, Berkeley, California, 94720*
[3] *University of Rhode Island, Graduate School of Oceanography, Narragansett, Rhode Island, 02882*

DATA PROCESSING DESCRIPTION

Data described here was used in *Flinders et al. 2018* as source data for a full waveform ambient noise tomography project centered around the Long Valley Caldera, California. Source seismograms were Rayleigh-wave ambient noise cross-correlations, derived from all seismic stations within 150 km of the caldera over the past 26 years, supplemented with Rayleigh-wave records from eleven large regional earthquakes (Figure 1, Data Table 1/2). The compiled data set includes Empirical Green's Functions (EGF) from the ambient noise cross-correlations and earthquake seismograms. Earthquake records have had their instrument responses removed and bandpass filtered over the range 1-40s. Only waveforms used in at least one iteration for at least one tomographic frequency band are included; 2396 EGFs and 141 earthquake seismograms.

A full description of data methodology used to calculate the Empirical Green's Functions follows.

Seismic Data

All high-gain vertical component seismic stations within ~150 km of the Long Valley Caldera were investigated for their potential use, including broadband (BHZ), high-broadband (HHZ), short period (SHZ) and extremely-short period (EHZ) channels. Data requests were made for 161 stations, for the years 1990–2016, through the Incorporated Research Institution for Seismology (IRIS) Data Management Center and the Northern California Earthquake Data Center (NCEDC). Preliminary data discrimination was performed by calculating probabilistic power spectral densities[1,2] (PPSD) for all stations for all years. The general tomographic period band of interest (1–40 s) was divided into two subgroups; a broadband-period group representing the full tomographic bandwidth, and a short-period group of 1–14 s. Initially, each station/year PPSD was analyzed over the range of 1–40 s, to ensure that the dominant energy fell within the secondary microseism band of ~3–10 s. Stations that passed this criterion were considered members of both subgroups (96 stations). Stations that failed this discrimination due to significant low frequency noise (> 14 s), were members of only the short-period group (40 stations). Stations/years with significant noise between 1–14 s were not considered further. This preprocessing discrimination and period-band subdivision ensured that stations/instruments which can have significant low-frequency noise (HHZ/SHZ/EHZ), would be fully utilized over their valid frequency ranges. All subsequent processing was then performed on these two period-band subgroups independently.

Empirical Green's Functions

Source data for tomography were derived from the cross-correlation of ambient noise waveforms. Processing to generate station-station EGFs included removing the seismometer instrument response, cutting the records into daily segments (two-hour overlaps), normalizing their spectra, bandpass filtering from 1–40 s (broadband group) and 1–14 s (short-period group), and resampling all data to 10 Hz. Normalization was performed via the frequency-time normalization method using 6.25 mHz wide frequency bands for both period-band subgroups[3]. This method normalizes the record across all frequencies, attempting to compensate for frequency-dependent amplitude loss and energy partitioning—however, amplitudes were not used in the tomographic process beyond relative signal-to-noise discriminations. Segments of seismic records overlapping with local earthquakes–within 200 km of a station, depth < 15 km, and 3.5 < M < 5.0–were nulled and tapered. A similar process was performed for teleseismic earthquakes (M > 5.0). Earthquake locations and times were taken from the ANSS Comprehensive Earthquake Catalog[4] (ComCat), with arrival time-windows calculated using a minimum arrival-time surface-wave velocity of 4.5/8 kms^{-1} (local/global) and a maximum of 2 kms^{-1}. Daily records with > 30% earthquake related signal were not processed further. Daily records for unique station-station pair combinations were then cross-correlated, stacked into individual months, and EGFs calculated as their time derivative. Monthly EGFs were compared to a total-stacked EGF, and months with cross-correlation coefficients < .75 were removed. This discrimination removed months where instruments were behaving erratically or when there was localized coherent noise. The remaining months were stacked into new total EGFs, and signal-to-noise values (SNR) were calculated over a filtered band of 4–14 s. SNR was defined as the ratio of the variance of a surface-wave window to the variance of a noise window starting five minutes after and 4x longer than the surface-wave window. SNR was calculated for both the causal and acausal sides of the EGFs, and those with an average SNR < 80 were removed. Similarly, EGFs comprised of < 50 daily records or with interstation distances < 4 km were also removed. Station-station combinations used in the final EGFs were then compared between the two period-band subgroups and any combinations that occurred in both subgroups were removed from the short-period group.

Composite Stations

Clusters of stations where individual interstation distances were < 350 m (e.g. tomographic grid-spacing) were combined into composite stations and their EGFs stacked (labeled AF.M0X). This reduced the total number of stations from 135 to 126. Composite stations were made from the following;

AF.M01	=	8E.MB01, NC.MMS
AF.M02	=	8E.MB11, NC.MQ1P
AF.M03	=	IM.NV11, IM.NV31, NN.MNA, US.MNV
AF.M04	=	NC.MDP, NC.MDPB
AF.M05	=	NN.OMM, NN.OMMB
AF.M06	=	XE.SNP12, XJ.BRR
AF.M07	=	XE.SNP24, XJ.FLL

Sensitivity kernels from the first tomographic iteration showed bias caused by high-station density of 10 stations in the Nevada Seismic Array (NVAR), although with interstation distances greater > 350 m. Nine of these stations were removed from the inversion and all subsequent iterations (IM.NV06 was kept). This reduced the total number of stations from 126 to 117, with 3170 unique station-station combinations, 1381 in the broadband group and 1789 in the short-period group.

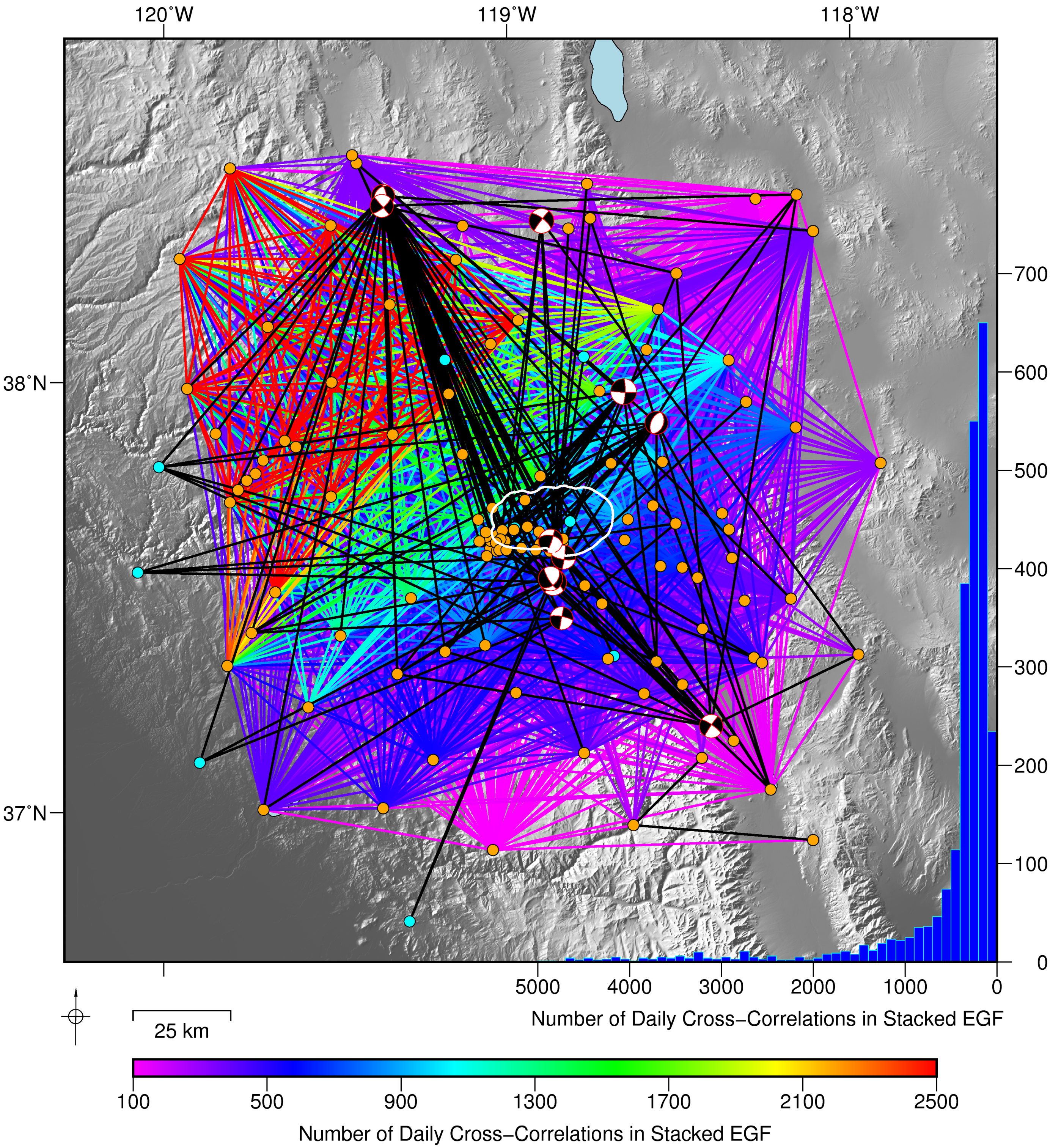

Figure 1. Station\event locations and distribution of Empirical Green's Functions. Cross-correlation (CC) of ambient noise waveforms from 117 seismic stations (orange circles) were used to reconstruct Empirical Green's Functions (EGFs) between seismic-station pairs. The number of daily CCs in the stacked EGFs, are shown by colored ray-paths and the inset histogram. Ray paths between eleven large regional earthquakes (Mw>4) and additional stations (cyan circles) shown in black. The Long Valley Caldera outlined in white.

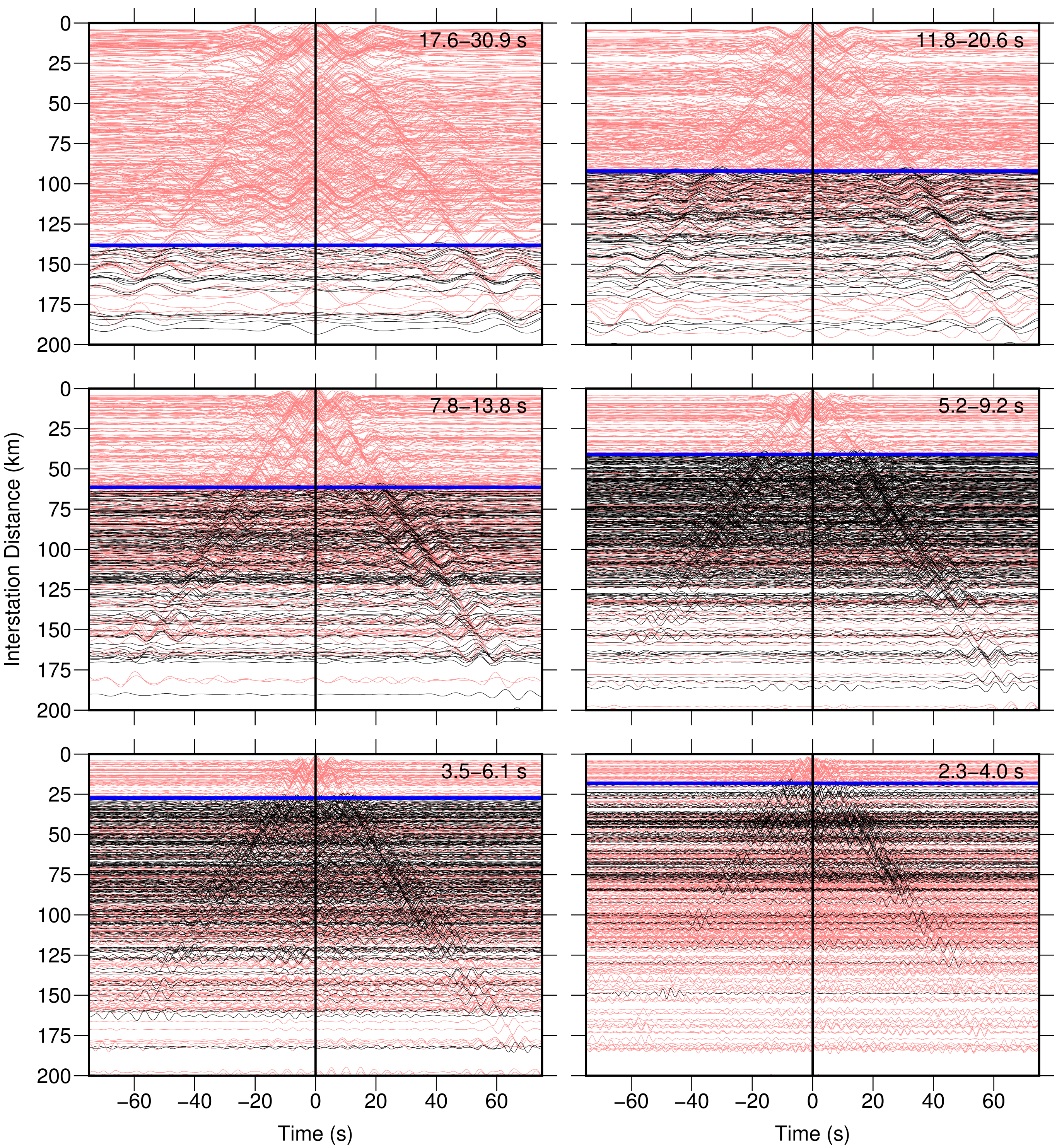

Figure 2. Empirical Green's Functions versus distance and data discrimination.
Ambient noise-derived Empirical Green's Functions for the six different tomographic frequency bands used, showing Rayleigh wave move out. Waveforms in red failed minimum requirements for (i) signal to noise, (ii) cycle skipping, (iii) cross-correlation coefficient or (iii) interstation distance, and were not used in the tomography and are not included in the compiled dataset. Waveforms in black were used in at least one iteration and are included in this dataset. Horizontal blue lines mark the frequency dependent minimum interstation distance cutoffs, equal to 1.75 times the mean frequency wavelength. For clarity only every tenth unused and every fifth used waveform is shown in each band.

DATA FORMAT DESCRIPTION – EMPIRICAL GREENS FUNCTIONS

Empirical Green's Functions are provided in ***SAC*** format with the following naming convention, for station name "*StnA*" from network "*Ntwk1*" and station name "*StnB*" from network "*NtwkB*";

*Ntwk1.StnA.Ntwk2.StnB.*SAC

SAC records are 3600.1 seconds in duration, beginning at *t = -1800 s* and ending at *t = 1800 s*, with 0.1 s sample spacing. The negative times of the record (*t* < 0) correspond to the acausal portion of the Empirical Greens Function, and the positive times to the causal portion. SAC header dates (KZDATE, KZTIME) for all files are set to the same arbitrary time (January 1, 2000, 12:00:00) to allow for easy plotting. Important SAC header values to note, following the same *Ntwk1.StnA.Ntwk2.StnB.*SAC, convention described previously:

KNETWK = *Network 1 name*
KSTNM = *Station A name*
STLA = *Station A latitude*
STLO = *Station A longitude*
STEL = *Station A elevation (meters)*
STDP = NOT USED

KEVNM = *Ntwk2.StnB*
EVLA = *Station B latitude*
EVLO = *Station B longitude*
EVDP = *Station B elevation (meters; NOT DEPTH!)*

DIST = *Interstation distance from Station A to Station B (kilometers)*
AZ = *Azimuth from Station B to Station A*

USER1 = *The number of days in the stacked cross-correlation*
USER2 = *Broadband (1) or short-period group (2) code number*
USER3 = *Station A channel code (1 = BHZ, 2 = HHZ, 3 = SHZ, 4 = EHZ)*
USER3 = *Station B channel code (1 = BHZ, 2 = HHZ, 3 = SHZ, 4 = EHZ)*

KCMPNM = ?HZ *(indicates high-gain vertical stations)*

A full Ambient Noise EGF SAC header example is provided below;

```
FILE: NN.BHP.TA.R08A.SAC
------------------------
    NPTS = 36001
    B = -1.800000e+03
    E = 1.800000e+03
    IFTYPE = TIME SERIES FILE
    LEVEN = TRUE
    DELTA = 1.000000e-01
    IDEP = DISPLACEMENT (NM)
    DEPMIN = -9.894372e-01
    DEPMAX = 1.000000e+00
    DEPMEN = -1.012630e-09
    KZDATE = JAN 01 (001), 2000
    KZTIME = 12:00:00.000
    KSTNM = BHP
    CMPAZ = 0.000000e+00
    CMPINC = 0.000000e+00
    STLA = 3.729950e+01
    STLO = -1.184873e+02
    STEL = 2.171000e+03
    STDP = 0.000000e+00
    KEVNM = TA.R08A
    EVLA = 3.834890e+01
    EVLO = -1.181064e+02
    EVDP = 1.419800e+03
    KHOLE =
    DIST = 1.212087e+02
    AZ = 1.961799e+02
    BAZ = 1.594728e+01
    GCARC = 1.090194e+00
    LOVROK = TRUE
    USER1 = 1.490000e+02
    USER2 = 2.000000e+00
    USER3 = 3.000000e+00
    USER4 = 1.000000e+00
    NVHDR = 6
    NORID = 0
    NEVID = 0
    LPSPOL = FALSE
    LCALDA = TRUE
    KCMPNM = ?HZ
    KNETWK = NN
```

DATA FORMAT DESCRIPTION – EARTHQUAKE WAVEFORM

Earthquake waveforms are provided in ***SAC*** format with the following naming convention, for event name "*XXXXXXXX*" from the NCSN event catalog "*nc*" and station name "*Stn*" from network "*Ntwk*";

*nc. XXXXXXXX.Ntwk.Stn.*SAC

SAC records are 600 seconds in duration, beginning at ~*t = 0 s* and ending at *t = 600 s*, with 0.1 s sample spacing. Important SAC header values to note, following the same *nc.XXXXXXXX.Ntwk.Stn.*SAC, convention described previously:

KNETWK = *Network 1 name*
KSTNM = *Station A name*
STLA = *Station A latitude*
STLO = *Station A longitude*
STEL = *Station A elevation (meters)*
STDP = NOT USED

KEVNM = *Event/catalog ID*
EVLA = *Event latitude*
EVLO = *Event longitude*
EVDP = *Event depth (below sea level; kilometers)*

DIST = *Interstation distance from event to Station A (kilometers)*
AZ = *Azimuth from event to Station A*

USER2 = *Broadband (1) or short-period group (2) code number*

A full Earthquake waveform SAC header example is provided below;

```
FILE: nc.71625245.NC.MBE.SAC
----------------------------
   NPTS = 6000
   B = 9.000000e-04
   E = 5.999009e+02
   IFTYPE = TIME SERIES FILE
   LEVEN = TRUE
   DELTA = 1.000000e-01
   DEPMIN = -1.001520e-05
   DEPMAX = 1.064946e-05
   DEPMEN = 1.027970e-09
   KZDATE = AUG 24 (236), 2011
   KZTIME = 11:59:51.399
   IZTYPE = BEGIN TIME
   KSTNM = MBE
   STLA = 3.674538e+01
   STLO = -1.192829e+02
   STEL = 9.900000e-01
   KEVNM = 71625245
   EVLA = 3.754849e+01
   EVLO = -1.188712e+02
   EVDP = -1.029700e+01
   DIST = 9.634160e+01
   AZ = 2.024356e+02
   BAZ = 2.218804e+01
   LOVROK = TRUE
   USER2 = 2.000000e+00
   NVHDR = 6
   SCALE = 1.000000e+00
   LPSPOL = FALSE
   LCALDA = TRUE
   KCMPNM = EHZ
   KNETWK = NC
```

Data Table 1. Earthquake source parameters (used in iterations 7–13)

Event ID	**Latitude**[b]	**Longitude**[b]	**Depth**[c]	**M_0** (10^{15})	**Mpp**[d]	**Mtt**[d]	**Mrr**[d]	**Mtp**[d]	**Mrp**[d]	**Mrt**[d]
NCSN[a]			(*km*)	(*N m*)	Mxx	Myy	Mzz	-Myx	Mzx	-Myz
30069317	37.601	-118.832	-11.4	2.25	0.9138	-0.4181	-0.4957	-2.0673	-0.5374	-0.1303
54958200	37.632	-118.866	-6.5	2.29	0.9235	-1.0952	0.1717	-1.4004	1.5255	0.0295
21405798	37.981	-118.658	-7.3	7.58	-0.0713	1.0502	-0.9792	7.4005	0.7965	1.4864
21445149	37.452	-118.841	-7.7	1.32	0.5649	-0.3709	-0.1940	-1.1999	0.3034	0.0058
51179422	38.430	-119.359	-8.7	13.3	1.0688	-0.1355	-0.9333	0.2668	-0.7934	0.2366
51182810	37.538	-118.866	-9.8	9.51	11.0080	-6.9079	-4.1001	-1.0017	3.9786	0.0017
71625245	37.548	-118.871	-10.3	2.71	2.5548	-1.5430	-1.0119	0.2113	1.4400	-0.7980
71664296	37.905	-118.570	-10.0	1.38	1.2229	0.0051	-1.2280	0.6017	0.2557	0.1263
72229261	38.409	-119.361	-4.4	1.34	1.2574	-1.2825	0.0251	0.2050	0.2745	-0.2315
72592670	37.206	-118.406	-15.7	17.6	1.6415	-1.5633	-0.0783	-0.6322	0.2015	0.2708
72744480	38.372	-118.899	-10.0	364	3.6251	-3.1146	-0.5105	-0.8722	0.7601	0.6727

Data Table 2. Supplementary earthquake information.

Event ID	**Date**	**Magnitude**[b]
NCSN[a]		(Mw)
30069317	3/5/1995	4.2
54958200	1/5/1998	4.2
21405798	10/9/2004	4.52
21445149	3/13/2005	4
51179422	3/9/2007	4.7
51182810	6/12/2007	4.6
71625245	8/24/2011	4.23
71664296	10/15/2011	4.03
72229261	5/30/2014	4.02
72592670	2/16/2016	4.77
72744480	12/28/2016	5.64

[a] Northern California Seismic Network (NCSN)
[b] Locations from NCEDC *Double Difference Catalog*
[c] Depth (relative to sea level) provided by Felix Waldhauser at Lamont-Doherty Earth Observatory
[d] Moment Tensors from Northern California Earthquake Data Center (NCEDC) *Mechanism Catalog*

References

1. McNamara, D. E. Ambient Noise Levels in the Continental United States. *Bull. Seismol. Soc. Am.* **94,** 1517–1527 (2004).

2. Beyreuther, M. *et al.* ObsPy: A Python Toolbox for Seismology. *Seismol. Res. Lett.* **81,** 530–533 (2010).

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

4. Young, J. B., Presgrave, B. W., Aichele, H., Wiens, D. A. & Flinn, E. A. The Flinn-Engdahl Regionalisation Scheme: The 1995 revision. *Phys. Earth Planet. Inter.* **96,** 223–297 (1996).