

# Supporting Information for “Probabilistic point source inversion of strong-motion data in 3-D media using pattern recognition”

Paul Käuffl,<sup>1</sup> Andrew Valentine,<sup>1</sup> and Jeannot Trampert<sup>1</sup>

## Contents of this file

1. Section S1 Neural network training using a pre-computed database of Green’s functions
2. Section S2 Outline of the neural network training procedure using mini-batch training
3. Section S3. Description of the 3-D Green’s functions database for the Whittier-Chino fault system made available as a part of this publication.
4. Table S1. Layered crustal and upper mantle model used for calculating approximate P wave arrival times and 1-D Green’s functions.

---

Corresponding author: A. P. Valentine, Department of Earth Sciences, Utrecht University, Postbus 80.115, 3508TC Utrecht, The Netherlands. (a.p.valentine@uu.nl)

<sup>1</sup>Department of Earth Sciences, Utrecht University Utrecht, The Netherlands

5. Table S2. Layered crustal and upper mantle model used for calculating approximate P wave arrival times and 1-D Green's functions.

6. Figure S1. Improvement of the network performance by mini-batch training.

7. Figure S2. Topographic map of study area; counterpart to Figure 1 of main article.

8. Figure S3. Generation of pseudo-random source locations for the Whittier-Chino fault system.

9. Figure S4. Real and virtual receiver locations of the 3-D Green's functions database.

10. Figure S5. Wavefield snapshots for the synthetic example of Figure 3 of the main article.

11. Figure S6. Test set performance for networks using 6 s of waveform data.

12. Figure S7. Test set performance for networks using 15 s of waveform data.

13. Figure S8. Information gain distributions for networks trained using 6, 15 and 30 s of waveform data.

14. Figure S9. Increasing the robustness of results by increasing the number of ensemble members.

15. Figure S10. Invariance of the resolving power with respect to the onset time perturbation.

## **Introduction**

Sections S1 and S2 provide additional details regarding the generation of neural network training sets and the neural network training procedure. Section S3 is concerned with the 3-D Green's functions database for the Whittier-Chino fault system that is made available as a part of this publication. Table S1 provides an overview over the source

parameterisation and according prior distributions as described in the manuscript. Table S2 gives the 1-D layered Earth model used to calculate approximate first arrival times (see Section S1). Figure S1 provides results related to the neural network training procedure, Figures S3 and S4 provide additional visualizations regarding the dataset described in Section S3. Figure S5 shows a snapshot of a wavefield synthesized using the Green’s functions database and Figures S6 to S10 provide supplementary results for the Chino Hills source inversion scenario discussed in the manuscript.

### **Section S1: Neural network training using a pre-computed database of Green’s functions**

The neural network based approach requires us to obtain a training set of synthetic waveform data and corresponding source parameter vectors. However, solving the elastic wave equation in a realistic 3-D heterogeneous medium, as done in this article, is computationally intensive and limits the amount of simulations that can be performed with a given amount of CPU time. We therefore adopt a two-stage approach for generating neural network training sets in order to use available computational resources optimally.

Firstly, a set of  $J$  source locations  $\{\mathbf{x}_j = (\text{lat}, \text{lon}, \text{depth})_j\}$  is drawn at random from a suitable prior distribution, potentially acknowledging any prior knowledge we might have on known fault traces and past seismicity. Subsequently, for a receiver located at  $\boldsymbol{\xi}_k$  and for the  $j$ -th source location, unit Green’s functions  $G_{ijk}$  are calculated for the unit moment tensors  $\hat{\mathbf{M}}_i = (\delta_{1i}, \delta_{2i}, \delta_{3i}, \delta_{4i}, \delta_{5i}, \delta_{6i})$ , where  $i = 1 \dots 6$  and  $\delta_{hi}$  is the Kronecker delta. The  $l$ -th component of the synthetic seismogram for moment tensor  $\mathbf{M}$  corresponding to the  $j$ -th source location at the  $k$ -th receiver can then be computed according to equation (1)

of the main article. We thus have the ability to compute the seismogram corresponding to any moment tensor located at any one of the points  $\mathbf{x}_j$ .

In order to obtain a neural network training set, we pick  $N$  source locations from the set  $\{\mathbf{x}_j\}$  at random, where typically  $J < N$ , i.e. the same location may appear multiple times. Subsequently, we draw  $N$  random moment tensors  $\{\mathbf{M}_n\}$  and half-durations  $\{\tau_n\}$  according to a prior distribution and calculate synthetic seismograms using the set of pre-computed unit Green's functions according to equation (1) of the main article. Moreover, random noise is added to the synthetic waveforms in order to simulate observational noise and potential modelling errors. Data vectors  $\mathbf{d}$  for neural network training are then formed by selecting time-windows, concatenating all components at all receivers and applying a non-linear input transformation in the same way as in *Käufel et al.* [2015]. Source and corresponding data vectors are combined into the dataset  $\mathcal{D} = \{(\mathbf{m}, \mathbf{d})_n\}$ .

All time windows start at the onset time  $t_0$ , i.e. the time at which the first arrival of seismic energy is detected by the first receiver in the receiver network. Various approaches exist that allow real-time triggering, such as monitoring variations in short-term/long-term signal averages. For simplicity, when working with the synthetic dataset  $\mathcal{D}$ , we estimate  $t_0$  by calculating P wave arrival times  $T_{jk}$  for each source-receiver pair using the TauP toolkit [*Crotwell et al.*, 1999] in a simplified 1-D average Earth model (see Table S2). These are then distorted by random noise drawn from the uniform distribution  $\mathcal{U}(-\delta t_0, \delta t_0)$  in order to desensitize the neural network with respect to errors in the determination of  $t_0$  on the order of  $\delta t$ . Here we choose  $\delta t_0 = 3\text{s}$  to take into account that the 1-D P wave arrivals are only poor estimates of the true onset times in the 3-D model, and to account for potential

errors in the onset times determined for the real observations. This may potentially be done using an automatic picking algorithm, such as a STA/LTA trigger [e.g. *Trnkoczy, 2002*]. For the observed Chino Hills dataset used throughout the main article, we have picked first arrival times manually. Note that the amount of perturbation  $\delta t_0$  has only little influence on the posterior uncertainties, since the relative timing of the signal at different stations is not affected (see Figure S10).

We redraw a new dataset  $\mathcal{D}$ , as described above, several times during the training procedure, in order to increase the total amount of examples presented to the network and to obtain a more thorough sampling of the prior distribution of moment tensors and source-time functions. This measure has a positive influence on the network performance (see Figure S1). A similar effect could have been achieved by increasing the total number of examples  $N$ , but at the cost of significantly increased computational and memory requirements for the training procedure. The next section gives an overview over the network training procedure.

## **Section S2: Outline of the neural network training procedure using mini-batch training**

During the neural network training procedure an iterative optimization algorithm is used to minimize the network prediction error

$$E[\mathcal{D}_{\text{tr}}] = - \sum_n \ln p[(m_k)_n | \mathbf{d}_n, \mathbf{w}], \quad (1)$$

where the sum runs over all examples in the training set  $\mathcal{D}_{\text{tr}}$  and  $p[(m_k)_n | \mathbf{d}_n, \mathbf{w}]$  is the likelihood of the the network parameter vector  $\mathbf{w}$  for the  $n$ -th training pattern  $(\mathbf{m}, \mathbf{d})_n$ .

See *Bishop* [1995] and *Käufel et al.* [2014] for a detailed motivation and description of the training procedure for feed-forward neural networks.

During training the error of a second independent validation dataset  $\mathcal{D}_{\text{val}}$  is monitored and after a pre-defined number of iterations has been reached the training procedure is stopped and the weight vector is chosen that maximizes the likelihood of the validation data set, i.e. that minimizes an equation similar to (1) for the validation set.

As mentioned in the previous section, we regenerate the training set  $\mathcal{D}_{\text{tr}}$  several times during the training procedure—a practice also known as *mini-batch* training. Here, the goal of mini-batch training is to limit the memory requirements and computational cost for each training iteration, which both scale linearly with the number of training set examples, while at the same time maximizing the amount of prior samples that can be presented during training. At iteration  $n_{\text{outer}}$ , we draw  $N$  source vectors from the prior distribution, as described in Section 1. Subsequently, we train the given network for  $n_{\text{inner}}$  iterations and repeat these two steps until a pre-defined maximum number of overall iterations is reached.

Figure S1 shows a comparison of 15 training runs that have been performed without training set re-generation and 15 runs that use mini-batch training as described above. As can be seen the minimum validation set errors are consistently lower for the networks trained using mini-batches for the same number of total iterations. In both cases the training sets consisted of 200000 examples. The training set has been regenerated 10 times in case of the mini-batch runs, which yields a total number of 2 Mio. prior samples, that are presented to the network, as opposed to 200000 samples in the case where a fixed

dataset is being used. Also the random-noise component that is added to each synthetic data vector, is redrawn every time the training set is regenerated.

### **Section S3, Description of the 3-D Green's functions database for the Whittier-Chino fault system made available as a part of this publication.**

We have generated a 3-D Green's functions database for the Whittier-Chino fault system for a set of 150 source locations located on and close to the fault. This database is made available to the public as a part of this paper and may aid future research on the regional seismicity in the Los Angeles basin. In the following, we provide details on the generation of the database.

The Whittier-Chino fault system consists of two northwest-trending sections with separate surface traces, which intersect at depth due to their opposing dip angles. See Figure S2 for the regional topographic context, and Figure S3 for a 3-D representation of the fault system taken from the Community Fault Model Version 4 (CFM) for Southern California [*Plesch et al.*, 2007]. Note that two alternative representations for the interaction between the two faults are present in the CFM, Figure S3 shows the option preferred by [*Plesch et al.*, 2007].

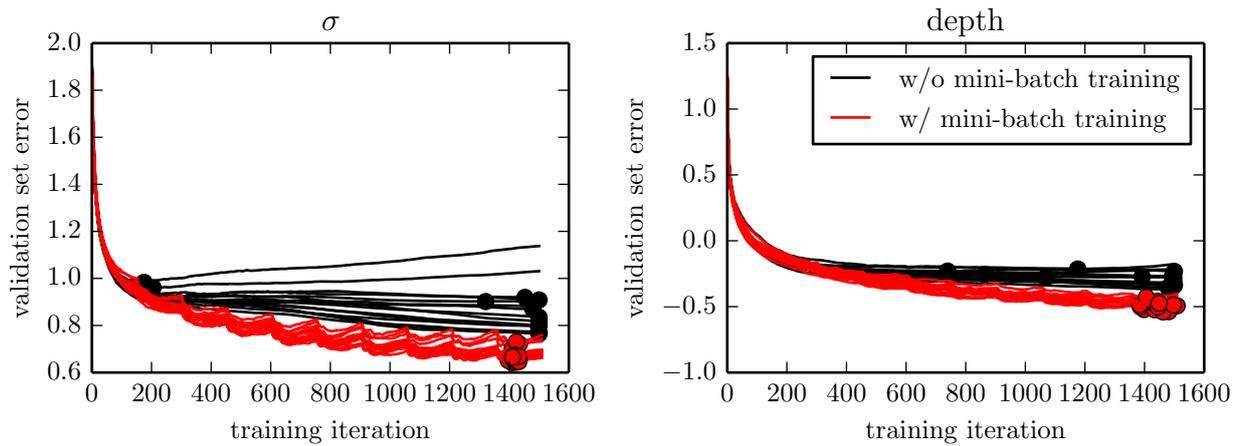
In order to obtain point source locations that follow the geometry of the fault sections, we draw locations uniformly on an auxiliary plane fitted to the larger of the two sections, the Whittier fault, as shown in Figure S3. The surface normal of the auxiliary plane is hereby determined by least-squares minimization, which is equivalent to finding the right-singular vector corresponding to the smallest singular value of the matrix containing all vertices of the fault mesh. In order to obtain a more even distribution with the

relatively small number of locations, we generate random points on the plane using a Sobol pseudo-random sequence. Subsequently we perturb each point in the direction of the surface normal projected onto the X-Y plane by a random distance drawn from a normal distribution with mean zero and whose standard deviation equals 5% of the total fault length. This is in order to account for potential uncertainties in the fault orientation, deviations of the actual fault geometry from the planar representation and the fact that earthquakes may occur on hitherto unmapped subsections of the fault. The locations of the resulting random source locations are visualized in Figure S3 as grey balls.

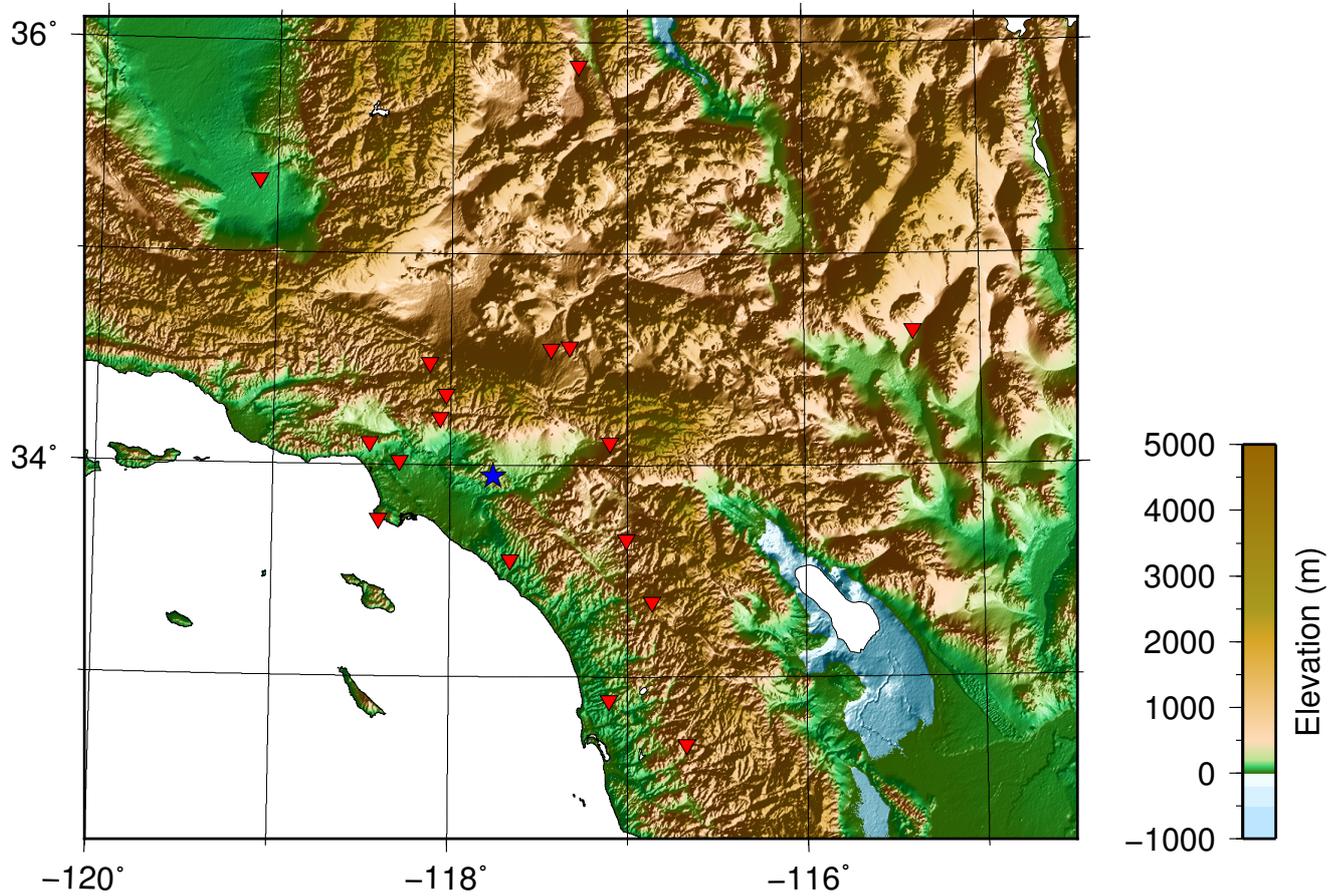
Subsequently, for each source location, we generate the 6 unit Green's functions  $G_{ijkl}$  (see eq. 1 in the main text) for the 150 source locations and a set of 1866 existing and virtual receivers in Southern California. Please find the full list of receiver names and locations in the provided STATIONS file. The receiver distribution is visualized in Figure S4. The synthetic displacement, velocity and acceleration waveforms of length 200 s are calculated using SPECFEM3D in the tomographic California model CVMH 11.9 [Tape *et al.*, 2009], in the domain  $32.2^\circ \leq \text{lat} \leq 36.8^\circ$ ,  $-121.6^\circ \leq \text{lon} \leq -114.4^\circ$ ,  $0 \leq \text{depth} \leq 400\text{km}$ . A time-step of 0.009 s is chosen and oceans, attenuation and gravity are not taken into account for the simulations. A heaviside source-time function is used and the resulting synthetic recordings are low-pass filtered using a cosine low-pass filter with corner frequencies 0.4 and 0.5 Hz and down-sampled to 0.999 s sampling period using integer decimation.

The data archive can be downloaded from [http://www.geo.uu.nl/~jeannot/My\\_web\\_pages/Downloads.html](http://www.geo.uu.nl/~jeannot/My_web_pages/Downloads.html) and is structured as follows: The main archive contains 150 sub-

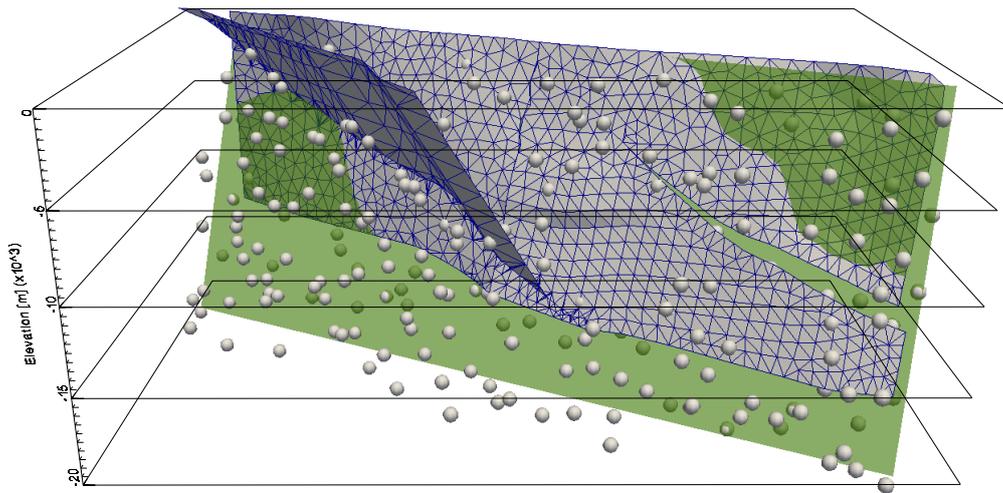
archives corresponding to the 150 source locations. Each sub-archive contains 6 folders corresponding to the 6 unit moment-tensors, which in turn contain a DATA directory holding SPECFEM3D input files, and an OUTPUT\_FILES directory containing the individual waveform files in SAC format. Note that the network code XX is used for the stations of the virtual receiver grid.



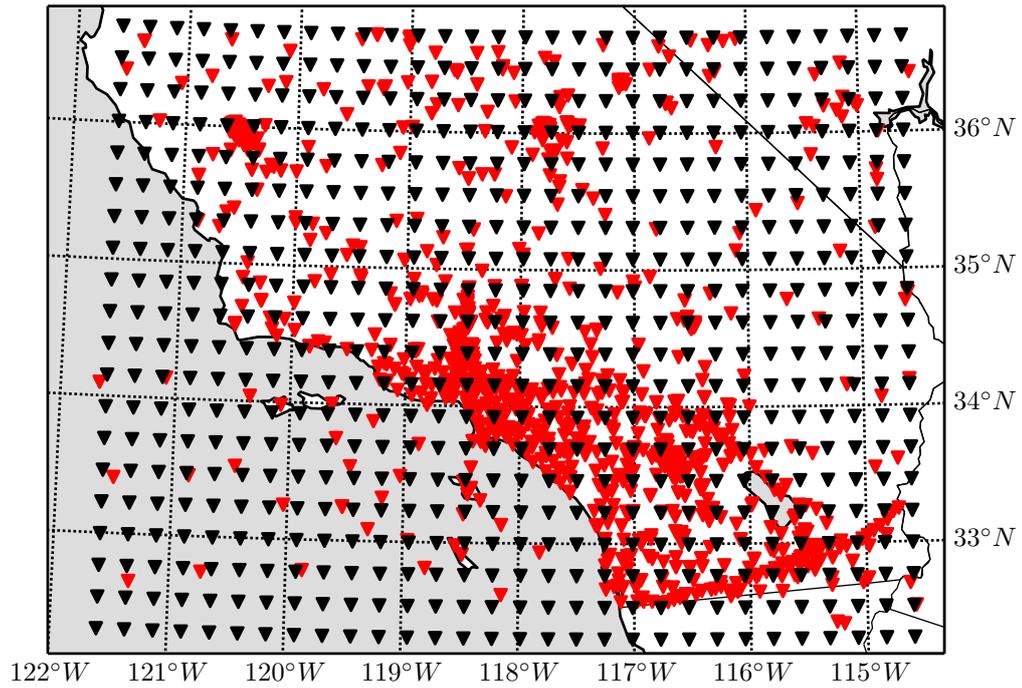
**Figure S1.** Improvement of the network performance by mini-batch training for parameters  $\sigma$  (left) and *depth* (right). Networks for other parameters show a similar behaviour. Black curves correspond to 15 neural networks trained w/o re-drawing the training set, red curves correspond to 15 runs, where the training set has been regenerated from the prior every 150 iterations. Filled circles indicate the validation error minima for each trace.



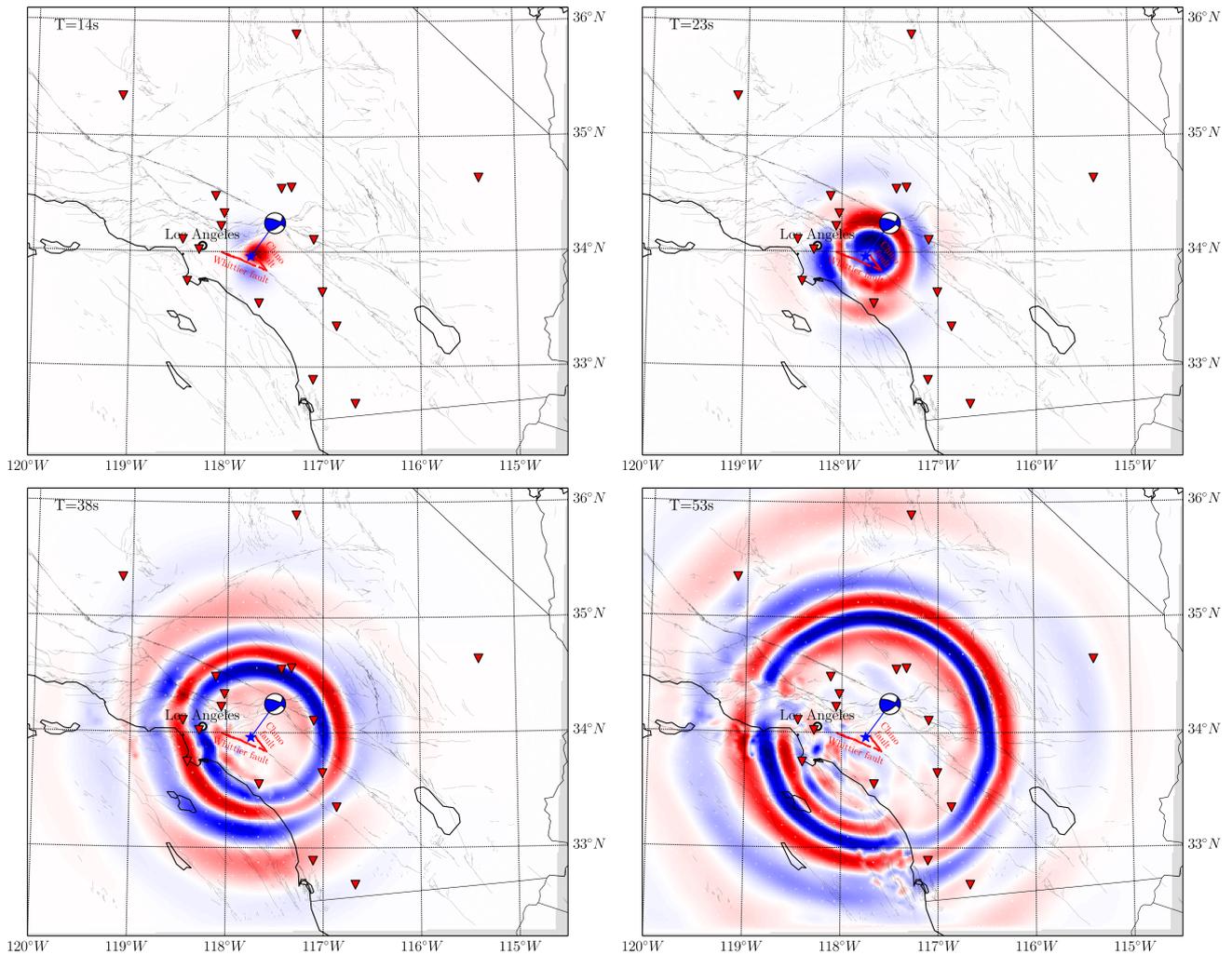
**Figure S2.** Map of study region showing topography [SRTM, *Farr et al.*, 2007]; see also Figure 1 of main article. Red triangles denote seismic stations used; blue star denotes the Chino Hills earthquake.



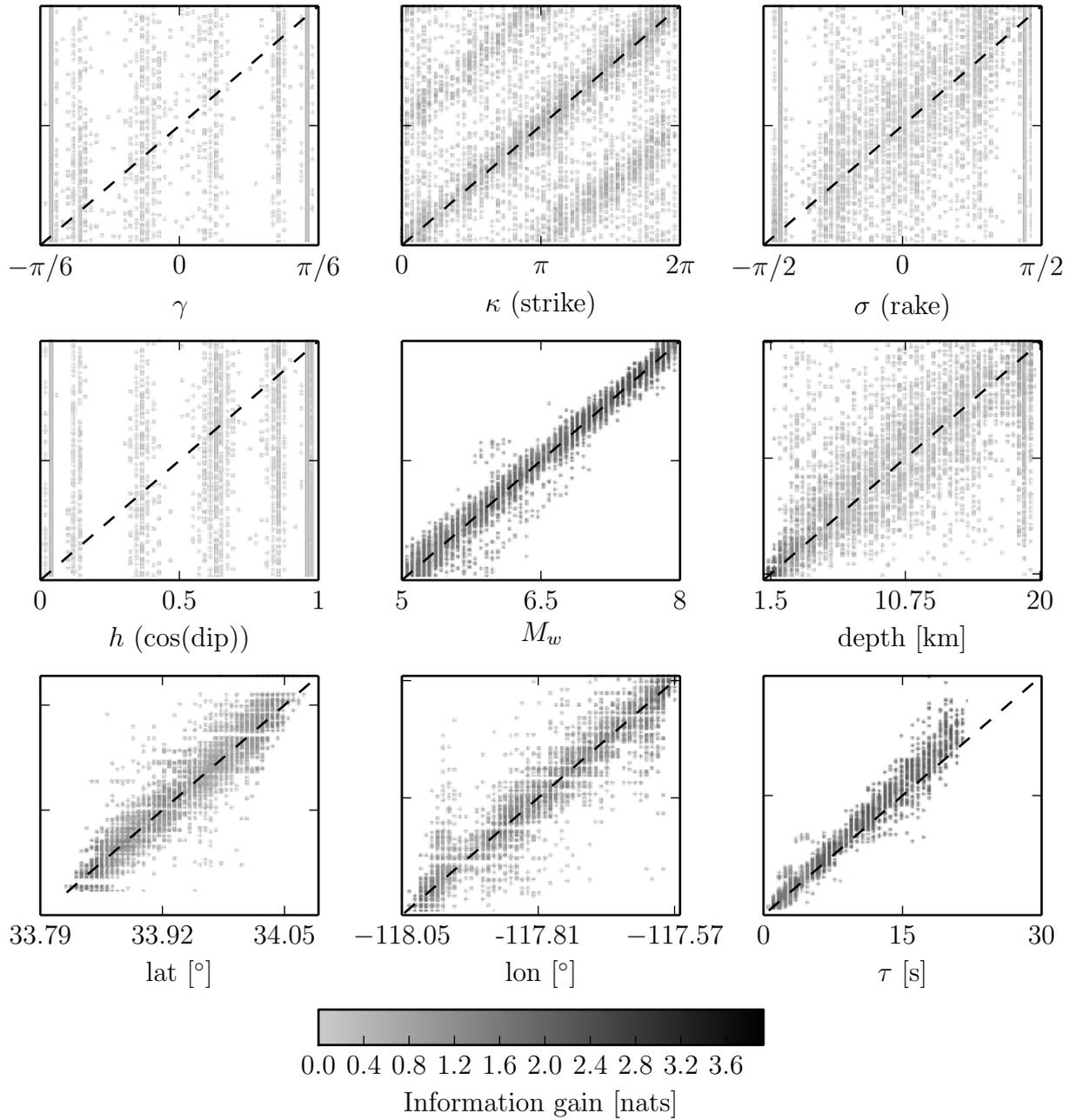
**Figure S3.** Whittier and Chino fault geometry taken from the Community Fault Model (CFM) for Southern California [*Plesch et al.*, 2007] (blue and grey mesh), a planar representation of the Whittier fault (green) and the locations of 150 pseudo-randomly distributed source locations (grey balls).



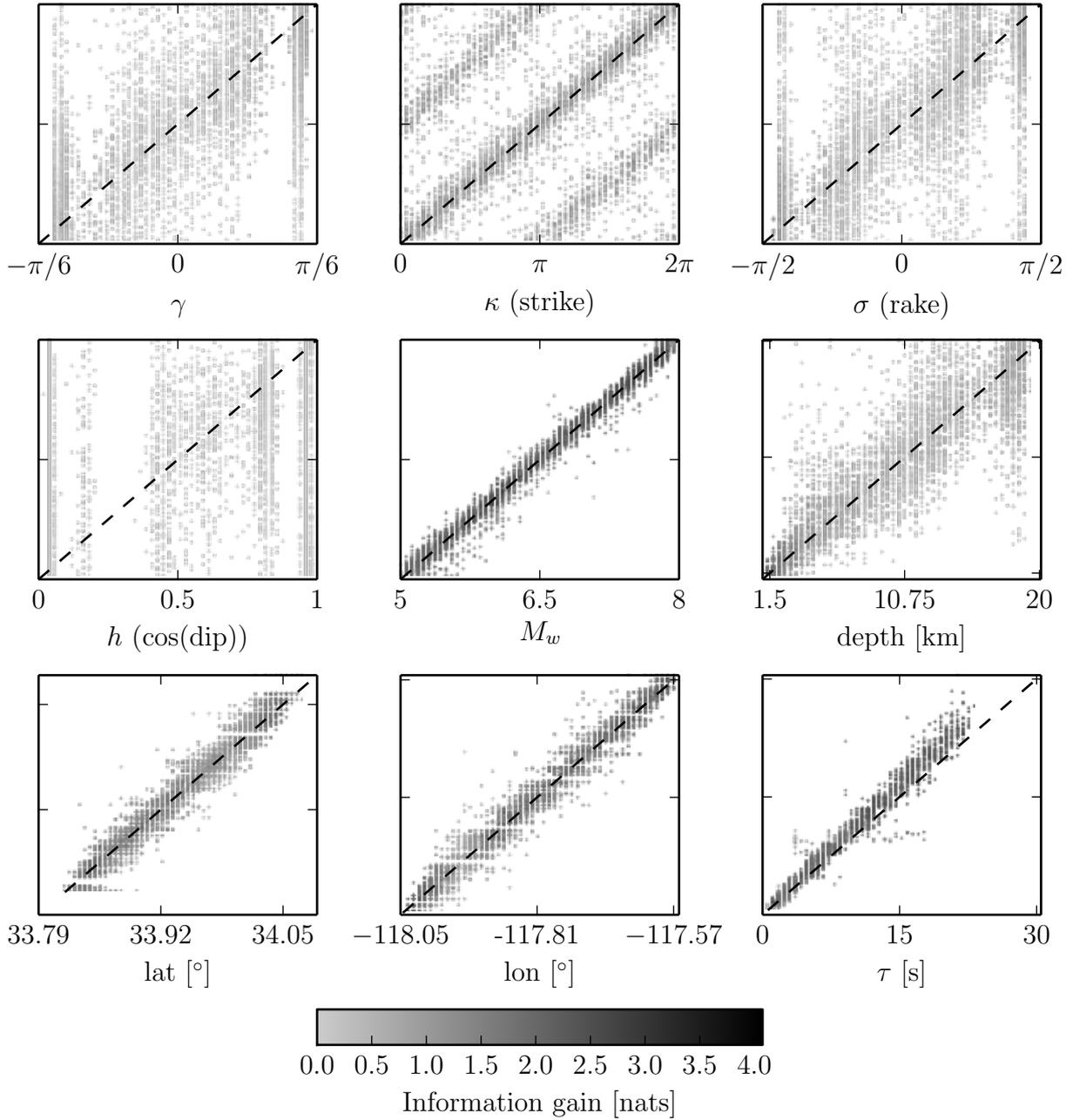
**Figure S4.** Real (red) and virtual (black) receiver locations at which synthetic displacement, velocity and acceleration waveforms are stored.



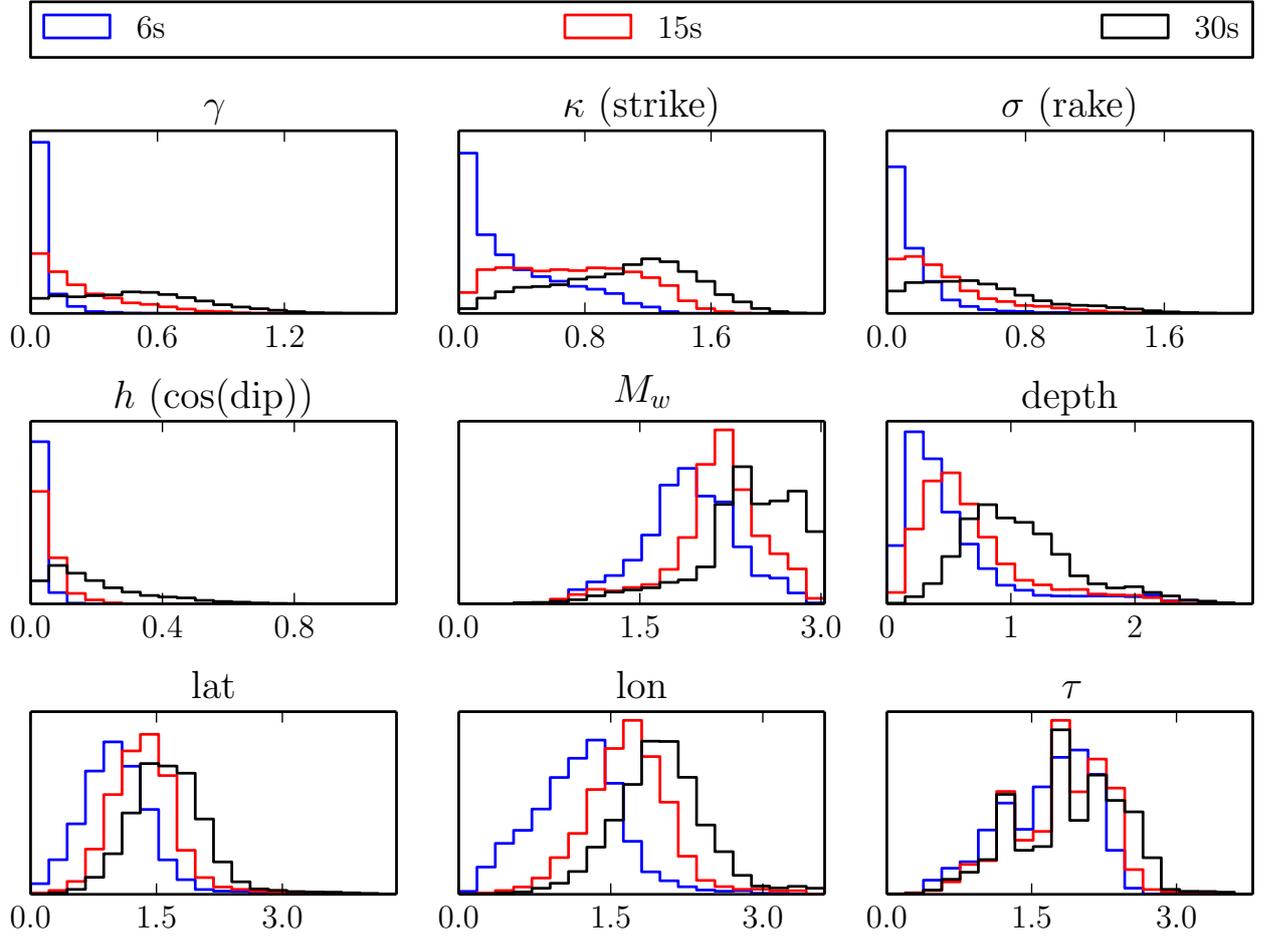
**Figure S5.** Snapshots of the vertical component surface velocity field for the synthetic example of Figure 3 of the main article with a source half-duration of 6 s. Points in time are with respect to origin time and correspond to the endpoints of the data windows of length 6 s, 15 s, 30 s and 45 s, respectively. Note that first arrivals are not clearly visible due to their small amplitudes. The first P wave arrival calculated using a 1-D average model is at 7.6 s after origin time at the closest receiver.



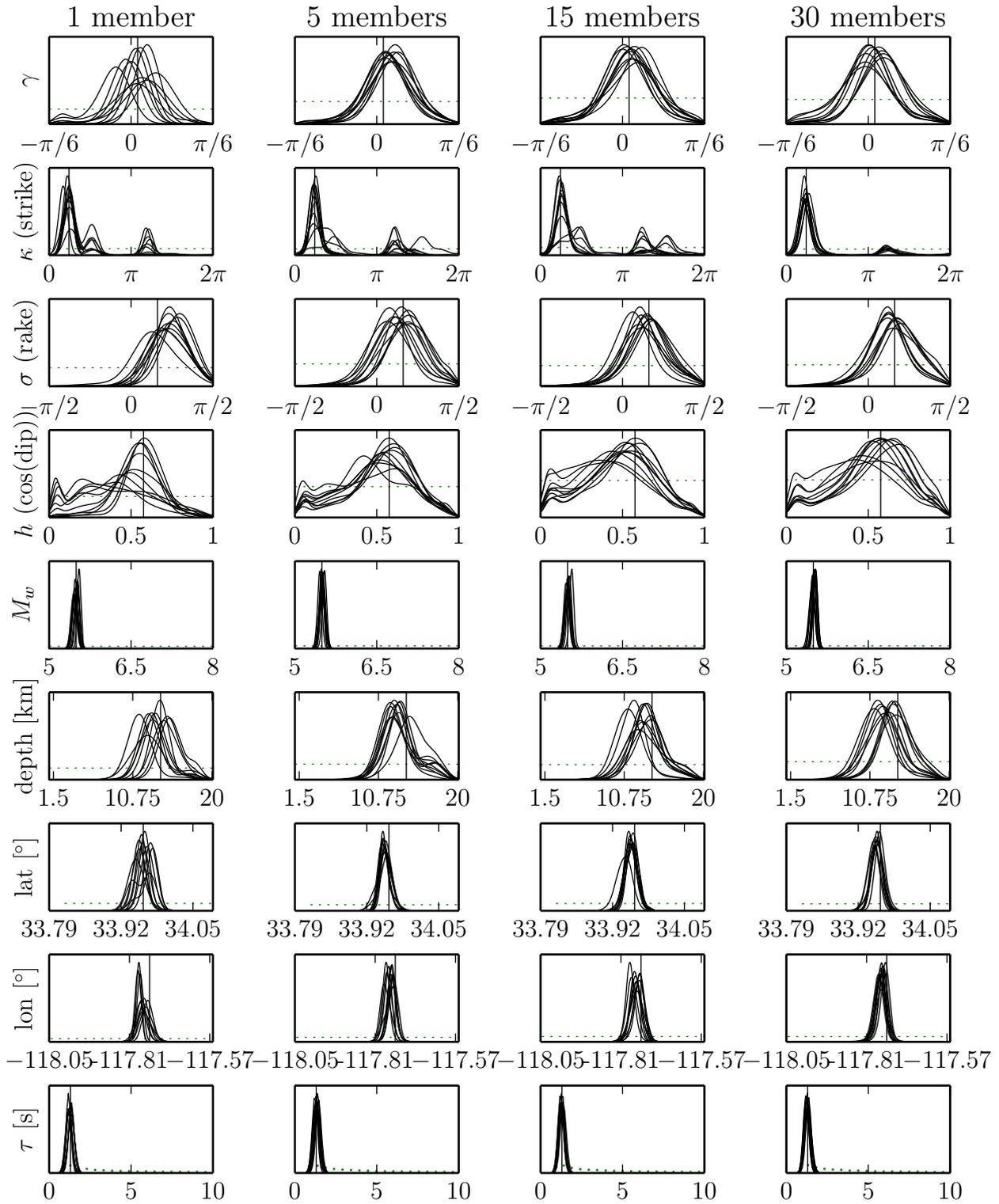
**Figure S6.** Test set performance for networks using 6 s of waveform data. Axes and gray-scale is the same as for Figure 2 in the main article.



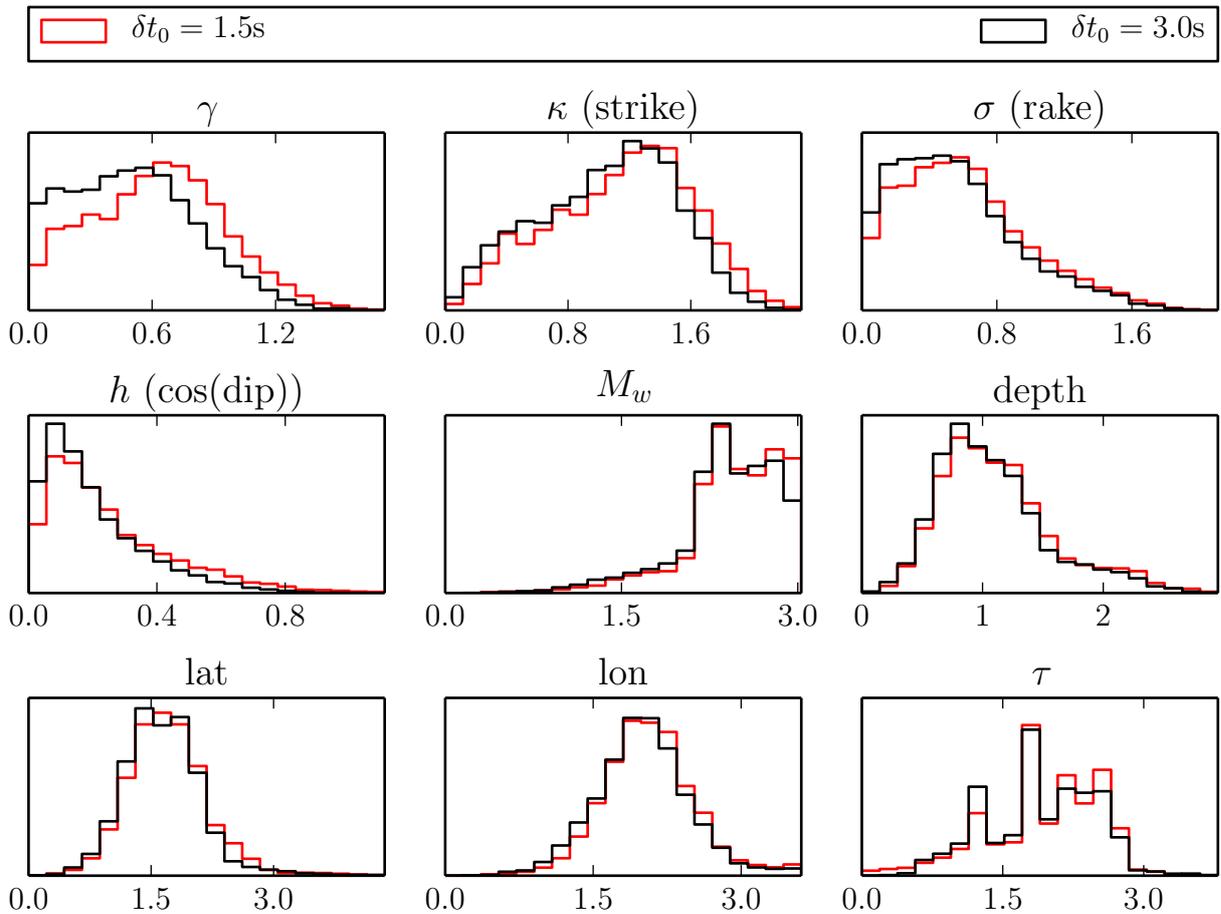
**Figure S7.** Test set performance for networks using 15 s of waveform data. Axes and gray-scale is the same as for Figure 2 in the main article.



**Figure S8.** Information gain distribution of 10000 synthetic test set examples for 6s (blue), 15s (red) and 30s (black). The horizontal axis shows the information gain with respect to the prior distribution  $D_{KL}$  in nats. Higher values correspond to a more certain answer.



**Figure S9.** Each panel shows marginal pdfs for the same synthetic example and 30s of data, but for 10 different random noise realizations (additive observational noise and a random perturbation of the onset time  $t_0$ ). Columns correspond to a different number of neural network ensemble members, rows to the different source parameters. In particular in the case of the lesser constrained parameters, the ensemble averaging helps to stabilize the results. Based on this observation, we chose to work with 30 members.



**Figure S10.** Information gain distribution for the synthetic test set under two different choices of the random onset time perturbation  $\delta t_0$ . Ideally, the neural networks would be invariant with respect to small perturbations of the onset time  $t_0$  and the prediction accuracy should not depend on the level of perturbation. This is the case except for parameter  $\gamma$ , where a smaller perturbation seems to improve the posterior certainty on average. However, for this parameter the bulk of the test set shows a very small information gain ( $< 1$  nat) in both cases and we conclude that the deterioration in prediction performance is acceptable.

**Table S1.** Point source parameters and prior ranges.

Parameter	Prior distribution*	Description
$\gamma$	$\mathcal{U}(-\pi/6, \pi/6)$	Non-DC component
$\kappa$	$\mathcal{U}(0, 2\pi)$	Strike
$\sigma$	$\mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$	Rake
$h$	$\mathcal{U}(0, 1)$	$\cos(\text{dip})$
$M_w = \frac{\log M_0}{1.5} - 10.7$	$\mathcal{U}(5.0, 8.0)$	Moment magnitude,
lat	see main text	Centroid latitude [°]
lon	see main text	Centroid longitude [°]
depth	see main text	Centroid depth [km]
$\tau = a \cdot 10^{-8} M_0^{1/3}$	$a \sim \mathcal{N}(1.05, 0.1)$	Half-duration

\*  $\mathcal{U}(a, b)$  denotes a uniform distribution on the interval  $[a, b]$ ;  $\mathcal{N}(\mu, \sigma)$  is a normal distribution with mean  $\mu$  and variance  $\sigma^2$ .

**Table S2.** The 1-D layered crustal and upper mantle model used for the calculation of 1-D Green’s functions and first arrival times. For obtaining approximate P wave arrival times and 1-D synthetic Green’s functions we have used a model, obtained by layer-wise averaging the global 3-D crustal model CRUST1.0 [Laske *et al.*, 2013] in the domain  $32.2 \leq \text{lat} \leq 36.8$ ,  $-121.6 \leq \text{lon} \leq -114.4$ .

Layer	Thickness [km]	$v_p$ [km/s]	$v_s$ [km/s]	$\rho$ [kg/cm <sup>3</sup> ]
1	1.56	2.38	0.91	2.04
2	7.89	5.97	3.44	2.72
3	8.33	6.46	3.72	2.82
4	8.81	6.89	3.84	2.95
5	10000 (half-space)	7.94	4.42	3.28

## References

- Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, vol. 92, Oxford University Press, Oxford, UK.
- Crotwell, H. P., T. J. Owens, and J. Ritsema (1999), The TauP Toolkit: Flexible seismic travel-time and ray-path utilities, *Seismological Research Letters*, *70*(2), 154–160.
- Farr, T., P. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank and D. Alsdorf (2007), The Shuttle Radar Topography Mission, *Reviews of Geophysics*, *45*, RG2004, doi:10.1029/2005RG000183.
- Käuffl, P., A. P. Valentine, T. B. O’Toole, and J. Trampert (2014), A framework for fast probabilistic centroid-moment-tensor determination–inversion of regional static displacement measurements, *Geophysical Journal International*, *196*(3), 1676–1693.
- Käuffl, P., A. Valentine, R. de Wit, and J. Trampert (2015), Robust and fast probabilistic source parameter estimation from near-field displacement waveforms using pattern recognition, *Bulletin of the Seismological Society of America*, *105*(4).
- Laske, G., G. Masters, Z. Ma, and M. Pasyanos (2013), Update on CRUST1.0 - A 1-degree Global Model of Earth’s Crust, *Geophysical Research Abstracts*, *15*, EGU2013–2658.
- Plesch, A., J. H. Shaw, C. Benson, W. a. Bryant, S. Carena, M. Cooke, J. Dolan, G. Fuis, E. Gath, L. Grant, E. Hauksson, T. Jordan, M. Kamerling, M. Legg, S. Lindvall, H. Magistrale, C. Nicholson, N. Niemi, M. Oskin, S. Perry, G. Planansky, T. Rockwell, P. Shearer, C. Sorlien, M. P. Suss, J. Suppe, J. Treiman, and R. Yeats (2007), Community Fault Model (CFM) for Southern California, *Bulletin of the Seismological Society of America*, *97*(6), 1793–1802,

doi:10.1785/0120050211.

Tape, C., Q. Liu, A. Maggi, and J. Tromp (2009), Adjoint Tomography of the Southern California Crust, *Science*, *325*(August), 988–992.

Tape, W., and C. Tape (2012), A geometric setting for moment tensors, *Geophysical Journal International*, *190*(1), 476–498, doi:10.1111/j.1365-246X.2012.05491.x.

Trnkoczy, A. (2002), Understanding and parameter setting of sta/lta trigger algorithm, *IASPEI New Manual of Seismological Observatory Practice*, *2*, 1–19.