

# *Bulletin of the Seismological Society of America*

This copy is for distribution only by  
the authors of the article and their institutions  
in accordance with the Open Access Policy of the  
Seismological Society of America.

For more information see the publications section  
of the SSA website at [www.seismosoc.org](http://www.seismosoc.org)



THE SEISMOLOGICAL SOCIETY OF AMERICA  
400 Evelyn Ave., Suite 201  
Albany, CA 94706-1375  
(510) 525-5474; FAX (510) 525-7204  
[www.seismosoc.org](http://www.seismosoc.org)

# Robust and Fast Probabilistic Source Parameter Estimation from Near-Field Displacement Waveforms Using Pattern Recognition

by Paul Käüfl, Andrew Valentine, Ralph de Wit, and Jeannot Trampert

**Abstract** The robust and automated determination of earthquake source parameters on a global and regional scale is important for many applications in seismology. We present a novel probabilistic method to invert a wide variety of (waveform) data for point-source parameters in real time using pattern recognition. Inferences are made in the form of marginal probability density functions for point-source parameters and incorporate realistic posterior uncertainty estimates. The neural-network-based method is calibrated using samples from the prior distribution, which are synthetic data vectors, and corresponding sources located in a predefined monitoring volume. Once a set of trained neural networks is available, inversions are fast with very moderate demands on computational resources: an inversion takes less than a second on a standard desktop computer. Uncertainties in the layered Earth model are taken into account in the Bayesian framework and increase the robustness of the results with respect to neglected 3D heterogeneities. Moreover, we find that the method is very robust with respect to perturbations such as observational noise and missing data and therefore is potentially well suited for automated and real-time tasks, such as earthquake monitoring and early warning. We demonstrate the method by means of synthetic tests and by inverting an observed high-rate Global Positioning System displacement dataset for the 2010  $M_w$  7.2 El Mayor–Cucapah event. Our results are compatible with published point-source estimates for this event within the respective uncertainty bounds.

*Online Material:* Additional information on the neural network methodology and implementation details, tables on neural network parameters, crustal model and reference double-couple solution, and figures showing prediction error, crustal models, normalized displacements, and histograms of weighted parameters.

## Introduction

Earthquakes, and sometimes their associated tsunamis, are among the most serious natural disasters humanity has to face. Despite great advances in our understanding of the rupture process, earthquakes are unpredictable and will most likely remain so (Geller *et al.*, 1997). It is therefore important to provide information as quickly as possible after an event has occurred to coordinate responses appropriately and limit damage and casualties. A description of the seismic source process is the basis of any such effort, and a moment tensor point source (Madariaga, 2007) provides a useful approximation. Automatic moment tensor point-source solutions are inferred globally from long-period teleseismic surface and body waves (Ekström *et al.*, 2012) using the centroid moment tensor (CMT) method (Dziewonski *et al.*, 1981). Furthermore, the incorporation of the  $W$  phase (Kanamori, 1993), which arrives shortly before the  $S$  phase, allows

for more rapid global automatic CMT solutions. However, while being very robust and well understood, the global approach is limited because it can only be applied to relatively large ( $M_w \gtrsim 5.5$ ) events. Further, the need to obtain sufficient teleseismic records limits the speed with which solutions can be obtained. Therefore, the method has been adapted to the inversion of regional broadband data (e.g., Ritsema and Lay, 1993; Walter, 1993; Zhu and Helmberger, 1996; Bernardi *et al.*, 2004), allowing smaller events to be characterized and reducing the response times by using shorter epicentral distances.

More recently, the increasing distribution of Global Positioning System (GPS) receivers of geodetic quality operating at high sampling rates ( $\geq 1$  Hz) has opened up new possibilities to observe the displacement wavefield directly and without clipping, even for large events in the near field

(Larson *et al.*, 2003; Bock *et al.*, 2011). Thus, in combination with traditional seismic observations, continuous GPS receivers are a promising candidate to provide additional constraints for rapid source parameter estimation. In particular, they allow the observation of the wavefield at very low frequencies, which may help to overcome the well-known magnitude saturation problem caused by the fact that the corner frequency of the source spectrum shifts toward lower frequencies with increasing source dimensions (Aki and Richards, 1980). It has recently been demonstrated that high-rate GPS data can be used for rapid source inversions in seismically active regions of the world, such as southern California, Japan, and Indonesia (e.g., Blewitt *et al.*, 2009; Crowell *et al.*, 2012; Melgar *et al.*, 2012, 2013; Ohta *et al.*, 2012; O'Toole *et al.*, 2012; Wright *et al.*, 2012).

Ideally, we would be able to utilize all observations available up to a certain point in time to rapidly obtain a source estimate, regardless of whether the information comes from a strong-motion sensor, a broadband seismic station, or a GPS receiver. However, it is a challenging problem to combine heterogeneous data types in an optimal way for a real-time inversion. Bock *et al.* (2011) and Geng *et al.* (2013) used Kalman filters to couple the GPS displacement time series to strong-motion data in order to obtain a broadband displacement time series with a high signal-to-noise ratio (SNR) in real time. A drawback of this approach is that it requires the strong-motion instruments to be collocated with GPS receivers, which is currently not the case for most operational high-rate GPS stations, at least in the western United States (Geng *et al.*, 2013). Making full use of these heterogeneous data types therefore requires new inversion algorithms that are able to jointly invert data from multiple, potentially spatially separated, sources. Furthermore, automated systems are required to deal with observational noise and missing data, for example, due to clipping or physical damage.

Realistic uncertainty estimates are required for any interpretation of the results, notably if they are to be used as an input for subsequent simulations such as finite-fault modeling or ground-motion prediction. The expected uncertainties strongly vary from event-to-event, depending on factors such as station coverage, the knowledge of Earth structure, and the event magnitude. This requires us to quantify uncertainties automatically in each case and indicates a probabilistic approach is needed. A consistent use of probabilistic methods would enable us to optimally combine the output of different independent estimations (Cua and Heaton, 2007) and propagate observational and modeling uncertainties through the entire processing chain using the laws of probability, a process that has been described as the formalization of logical reasoning (Jaynes, 2005).

Recently, we introduced a rapid, neural-network-based, probabilistic point-source inversion algorithm and demonstrated how it can be used to invert static displacement observations (Käüfl *et al.*, 2014). Here, we present an extension to our approach, enabling us to directly invert waveform data

as provided by accelerometers, broadband seismometers, or continuously operating GPS receivers. It can operate automatically, in (near) real time, as the data arrives. It is based on a nonlinear pattern recognition algorithm and has several advantages over (often linearized) optimization schemes, such as least-squares inversions or Markov chain Monte Carlo methods. First, our method is very flexible with respect to the type of observations. Any combination of observables can in principle be used as an input pattern for a neural network, as long as a sufficient quantity of examples of input and output patterns can be provided. Second, inversions are fast and very robust with respect to outliers and observational noise. Similar probabilistic neural network algorithms have recently been applied to other geophysical inverse problems (e.g., Meier *et al.*, 2007; Shahraneen *et al.*, 2012; de Wit *et al.*, 2013, 2014; Walker and Curtis, 2014).

This article is organized as follows. First, we give a brief introduction to the neural-network-based method and describe the source and Earth model parametrization. Second, we apply the method to an observed GPS displacement waveform dataset for the 2010 El Mayor–Cucapah event. Third, we introduce a framework to analyze the resolving power of the data and assess the robustness of inversions by performing a number of synthetic experiments, followed by a discussion on more general consequences and limitations of the approach.

## Methodology

The goal of an inversion in general is to determine a set of model parameters  $\mathbf{m}$ , given a vector of observations  $\mathbf{d}$ . Conventionally, this problem is approached by fitting a model to an observed datum  $\mathbf{d}$  by repeatedly evaluating a forward problem  $\mathbf{d} = \mathbf{g}(\mathbf{m})$ , relating model parameters to observations, for example, by running a wave propagation code until a candidate model (or a set of candidate models) is found that matches the observations sufficiently. Examples are the CMT algorithm for moment tensor inversion (Dziwonski *et al.*, 1981) or the adjoint method often adopted for tomographic imaging (e.g., Fichtner *et al.*, 2006). This approach is typically prohibitive for real-time applications, because it involves repeated evaluation of the often computationally expensive forward problem  $\mathbf{g}(\mathbf{m})$ . Furthermore, if similar inversions must be performed repeatedly, for example in monitoring tasks such as earthquake early warning (EEW), repeated evaluation of the forward problem is required for each independent observation.

A second family of algorithms is based on a direct approximation of the inverse mapping  $\mathbf{g}^{-1}(\mathbf{d})$ . Typically, this is achieved by means of an empirical regression model, sometimes dubbed an empirical scaling law, which has to be calibrated beforehand, often using a database of past observations. These models are computationally cheap at evaluation time and this route is thus often taken for rapid source parameter estimation in the context of EEW. An example is the rapid estimation of the event magnitude using proxy

variables such as peak ground movement or the predominant period (e.g., Nakamura, 1988; Allen and Kanamori, 2003; Kanamori, 2005).

Our approach belongs to the second group of methods. It is based on a neural network forming a regression model, which relates observable waveform data with source parameters. However, it is calibrated using a synthetic dataset accurately capturing the physics of wave propagation and giving us full control over source distributions and the noise model. Moreover, it provides a fully nonlinear and probabilistic description of the possibly nonunique inverse mapping, thereby relaxing many of the restrictions associated with conventional regression-based approaches, which often require measurements to be defined in such a way that the relation of observables to parameters takes a simple form. For instance, when rapidly estimating the moment magnitude from the predominant period of a short initial waveform window (e.g., Wu and Kanamori, 2005), a log-linear regression relation is fitted to a database of historic observations and catalog events.

We follow a Bayesian approach (e.g., Tarantola, 2005), that is we describe our state of knowledge on a set of earthquake source parameters as the probability density function (PDF)  $p(\mathbf{m}^f | \mathbf{d})$ , in which  $\mathbf{m}^f$  is a vector of source parameters and  $\mathbf{d}$  is the totality of (noisy) observations available for inversion at a certain point in time. An inference about a specific parameter  $m_k$  can be made by means of the marginal posterior PDF,

$$p(m_k^f | \mathbf{d}) = \int \prod_{j \neq k} dm_j p(\mathbf{m}^f | \mathbf{d}). \quad (1)$$

The marginal PDF tells us what we can learn about parameter  $m_k^f$  while taking into account the possible variations in all other parameters. As in Käufel *et al.* (2014), we directly model marginal posterior PDFs (equation 1) as mixtures of 1D Gaussian kernels,

$$p(m_k^f | \mathbf{d}) = \sum_{i=1}^M \frac{\alpha_i(\mathbf{d})}{\sqrt{2\pi}\sigma_i(\mathbf{d})} \exp\left\{-\frac{[m_k^f - \mu_i(\mathbf{d})]^2}{2\sigma_i(\mathbf{d})^2}\right\}, \quad (2)$$

in which  $\alpha_i(\mathbf{d})$ ,  $\mu_i(\mathbf{d})$ , and  $\sigma_i(\mathbf{d})$  are functions of the data  $\mathbf{d}$  that are approximated by a two-layer feedforward neural network. Any PDF can be represented to arbitrary accuracy by a superposition of a fixed number of Gaussian kernels if enough kernels are being used (McLachlan and Basford, 1988). Our methodology is based on the mixture density network (MDN) introduced by Bishop (1995). An MDN can learn a parametric approximation to any smooth, potentially multimodal, conditional probability density from a set of samples. We repeat the training procedure several times for each parameter and combine the independently obtained network models into ensembles, which increases the accuracy and robustness of the predictions. Naturally the number of Gaussian kernels  $M$  in equation (2) has an influence on the accuracy of the approximation. However, note that the effective

number of kernels is much larger than  $M$ , due to the ensemble averaging of several independently obtained neural networks, which possess  $M$  Gaussian kernels each. We thus conclude that in practice, the limit imposed by using a finite number of Gaussian kernels is negligible compared to other limiting factors, such as the finite number of samples in the training set (see the *Inference with Prior Samples* section).

The methodology is described in full detail in Käufel *et al.* (2014). The main development in this article is the use of waveform data; the underlying method is unchanged. <sup>Ⓔ</sup> Details on the neural network architecture used throughout this work are provided in the Neural Network Architecture and Implementation Details section and in Table S1 in the electronic supplement to this article.

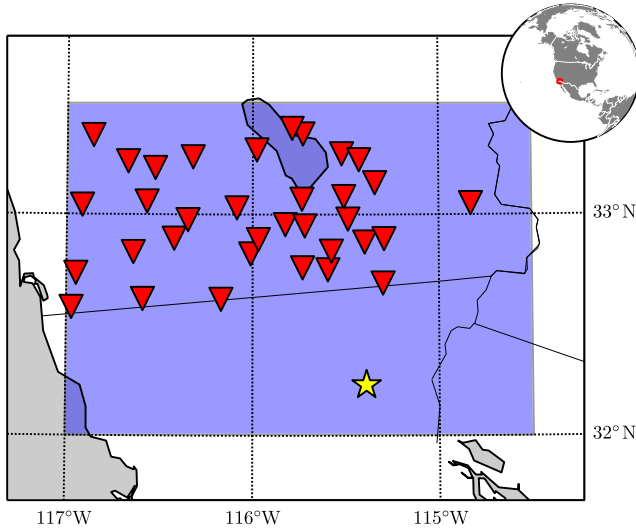
We represent seismic sources as moment tensor point sources (Madariaga, 2007). As in Käufel *et al.* (2014), we parametrize the space of moment tensors using a geometric parametrization suggested by Chapman and Leaney (2012) and Tape and Tape (2012) and additionally restrict ourselves to double-couple sources. The latter choice is mainly motivated by the consideration that a double couple is potentially easier to interpret than a full moment tensor source in the context of an early warning application. Moreover, this assumption simplifies the inverse problem by reducing the number of free model parameters. As a consequence, fewer synthetic examples are required for neural network training, which reduces the computational cost of the training procedure. The framework can easily be extended to include non-double-couple (as in Käufel *et al.*, 2014), as well as nondeviatoric sources; and, because of our parametrization, such constraints are straightforward to implement.

The probabilistic framework requires us to define suitable prior distributions for each parameter, representing the knowledge we have about a parameter before the data have arrived. This involves the assumption that the event has occurred in a particular monitoring area, which must be determined in advance. We work with uniform distributions on orientational parameters, source location, depth, and logarithmic magnitude. The temporal evolution of the source is parametrized by a triangular source time function with varying half duration. Considering that earthquakes appear to some extent to be self-similar (e.g., Aki and Richards, 1980), it is reasonable to assume that the length of the source wavelet should scale with the event size. For simplicity, here we correlate the half duration to the event size using the relation

$$\tau = a \times 10^{-8} M_0^{1/3} \quad (3)$$

(Ekström *et al.*, 2012) and note that any other relation can easily be replaced for equation (3) if we have reason to make other assumptions on the rupture process. The proportionality factor  $a$  is drawn from a normal distribution with mean 1.05 and standard deviation 0.1. This is intended to account for uncertainties in the empirical determination of equation (3).

In summary, we parametrize moment tensor point sources using the eight independent model parameters given



**Figure 1.** Station configuration and source region according to Table 1 (shaded area). The triangles show the position of 34 real-time Global Positioning System (GPS) receivers, and the star denotes the epicenter of the 2010 El Mayor–Cucapah event. The color version of this figure is available only in the electronic edition.

in Table 1. The corresponding prior distributions cover all possible double-couple point sources in a predefined monitoring volume. For the latter, we chose a rectangular region in southern California, that spans the Salton trough and the Mexicali Valley, thereby covering several major fault zones, such as the southern tip of the San Andreas fault, the Brawley seismic zone, the Imperial fault, the Laguna Salada, and the Cerro Prieto (Lovely *et al.*, 2006). In depth, the volume extends from 1.5 to 20 km. Figure 1 shows an overview of the source region and receiver configuration.

### Observational Noise and Modeling Errors

According to equation (2), the result of any inference about a source parameter is made in the form of a PDF. The posterior uncertainty estimates incorporate several sources of error, such as the observational noise propagated into the model space, trade-offs between parameters, limitations in sensitivity of the data to the model parameter in question, and also potential limitations in the neural network model used to approximate the inverse mapping (as will be discussed in detail in the [Inference with Prior Samples](#) section).

We assume that observational noise is additive and independent of  $\mathbf{m}$  and write

$$\mathbf{d} = \mathbf{g}(\mathbf{m}) + \boldsymbol{\epsilon}, \quad (4)$$

in which  $\mathbf{g}(\mathbf{m})$  is an approximation to the forward problem, that relates physical model parameters  $\mathbf{m}$  to observable data vectors  $\mathbf{d}$  and  $\mathbf{m} = (\mathbf{m}^f, \mathbf{m}^s)$ , with the joint vector of source parameters  $\mathbf{m}^f$  and structural parameters  $\mathbf{m}^s$ , respectively, representing a 1D layered crustal model over a half-space. The vector  $\boldsymbol{\epsilon}$  accounts for any part of the observed signal that is not accurately represented by  $\mathbf{g}(\mathbf{m})$ . It includes both errors

Table 1

Point-Source Parametrization and Prior Distributions

Parameter	Prior Distribution*	Description
$\kappa$	$\mathcal{U}(0, 2\pi)$	Strike
$\sigma$	$\mathcal{U}(-\pi/2, \pi/2)$	Rake
$h$	$\mathcal{U}(0, 1)$	$\cos(\text{dip})^\dagger$
$M_w = (\log M_0/1.5) - 10.7$	$\mathcal{U}(5.0, 8.0)$	Moment magnitude
Latitude	$\mathcal{U}(32.0, 33.5)$	Centroid latitude ( $^\circ$ )
Longitude	$\mathcal{U}(-117.0, -114.5)$	Centroid longitude ( $^\circ$ )
Depth	$\mathcal{U}(1.5, 20)$	Centroid depth (km)
$\tau = a \times 10^{-8} M_0^{1/3}$	$a \sim \mathcal{N}(1.05, 0.1)$	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  $\sigma^2$ .

$^\dagger \cos(\text{dip})$  refers to cosine of dip.

due to neglected (source) physics, as well as observational errors such as ambient noise. The noise vector  $\boldsymbol{\epsilon}$  is drawn from a normal distribution with covariance  $\mathbf{C}_d$ .

The fact that we only simulate wave propagation in a 1D-layered medium requires us to find a 1D representation of the highly heterogeneous Earth structure in the study region. Because there are multiple ways to define a 1D average of a 3D structure, and it is not clear that we should give preference to any particular choice, we marginalize over a range of 1D models, rather than conditioning our posterior estimates on one specific average structure. Although the main purpose of this measure is not to accurately represent existing 3D heterogeneities in the study region, it helps to desensitize the neural networks with respect to effects caused by local structural variations. We obtain distributions of 1D crustal structure from the 3D crustal model CRUST 1.0 (Laske *et al.*, 2013). CRUST 1.0 is defined on a  $1^\circ \times 1^\circ$  grid featuring an ocean, two sedimentary and three crystalline rock, and an upper-mantle layer (s) at each grid point. We ignore the ocean layer and topography and average the two sedimentary layers to obtain a single sedimentary layer. Thus, we have a five-layer model in every grid point in the model region ranging from  $31.5^\circ$  to  $34.5^\circ$  in latitude and from  $-118.5^\circ$  to  $-113.5^\circ$  in longitude. For each layer, the minimum and maximum values of each parameter (layer thickness,  $V_P$ ,  $V_S$ , density) taken over all grid points provide the limits of a uniform distribution from which we independently draw 1D layered crustal models  $m_k^g \sim \mathcal{U}(m_{k,\min}^g, m_{k,\max}^g)$ . The ranges are summarized in Table 2 and  $\textcircled{E}$  a set of test models is visualized in Figure S2.

We do not attempt to estimate other modeling errors, such as neglected 3D heterogeneities or attenuation. However, as we will show, our results are in general very robust with respect to unmodeled noise and other perturbations (see the [Discussion](#) section).

### Formation of Input Vectors and Preprocessing

$\textcircled{E}$  To use the observed and synthetic data for neural network training and evaluation, several preprocessing steps are required, which are described in detail in the text of

Table 2  
Prior 1D Earth Model Ranges

Layer	Thickness (km)	$V_p$ (km/s)	$V_s$ (km/s)	$\rho$ (kg/cm <sup>3</sup> )
Sediments	[0.11,2.94]	[1.75,3.47]	[0.34,1.76]	[1.82,2.28]
Upper crust	[1.61,11.8]	[5.00,6.30]	[2.707,3.63]	[2.55,2.79]
Middle crust	[3.54,10.7]	[6.30,6.60]	[3.65,3.80]	[2.78,2.86]
Lower crust	[5.90,14.2]	[6.60,7.10]	[3.60,4.05]	[2.86,3.05]
Upper mantle (half-space)	[10000,10000]	[7.68,8.09]	[4.28,4.49]	[3.17,3.33]

the electronic supplement. In essence, three-component displacement waveforms at all receivers are low-pass filtered, downsampled to the Nyquist sampling rate, windowed, and concatenated to form input data vectors  $\mathbf{d}$ . We include all data that arrive at the receiver network from the moment  $t_0$  of the first signal detected, typically at the station closest to the epicenter, until  $t_0 + \Delta T$ . That is, receivers that have not yet triggered or are nonoperational are nevertheless included in the inversion. This is to simulate an automated monitoring situation, where no manual data quality check can be performed. Clearly, the fact that certain receivers did not record any signal yet at time  $t_0 + \Delta T$  represents information on the hypocenter location (Satriano *et al.*, 2008).

#### Application to Displacement Waveform Data

The increasing availability of high-rate GPS stations provides us with the potential to directly observe the dynamic displacement field caused by major earthquakes in real time. We use displacement waveforms observed by the California Real Time Network (CRTN), a network of high-rate GPS receivers in southern California, during the 2010 El Mayor earthquake (Hauksson *et al.*, 2010) to determine point-source parameters. This serves as an example and enables us to assess the potential of GPS waveform data to constrain point-source parameters. However, other time-series data, such as accelerograms or velocity seismograms, or a combination thereof, could likewise have been used.

Previously, a good agreement has been reported between integrated data of collocated accelerometer and broadband instruments with the GPS displacement time series for this event (Allen and Ziv, 2011; Bock *et al.*, 2011). Point-source inversions using either a similar GPS displacement waveform dataset or using static offsets extracted from the displacement time series have recently been conducted by Melgar *et al.* (2012), Zheng *et al.* (2012), O’Toole *et al.* (2013), and Käufel *et al.* (2014) and have shown a good agreement to catalog solutions, although with large uncertainties attached to some of the source parameters.

We draw a total of 100,000 source and crustal models from the prior distributions (Table 1) and solve the forward problem  $\mathbf{g}(\mathbf{m})$  using a method developed by O’Toole and Woodhouse (2011). The synthetic waveforms contain body- and surface-wave phases and include the remaining static offset. We low-pass filter the observed and synthetic traces using a cosine low-pass filter with corner frequencies 0.07

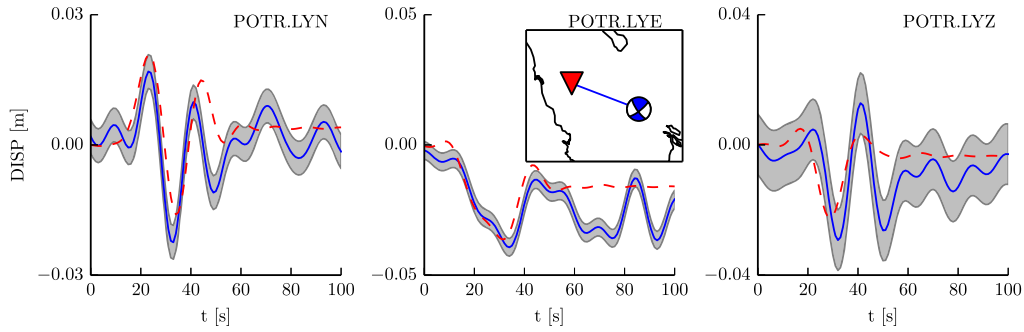
and 0.08 Hz and therefore work with a sampling interval of 6 s. This choice is motivated by the fact that we cannot expect to explain any of the higher frequency signal with a 1D average model in the highly heterogeneous southern California crust (Tape, 2009). Moreover, a higher sampling rate would significantly increase the computational effort required for training set generation and network training, because a higher sampling rate would in turn lead to a larger input dimensionality and therefore bigger neural networks (see also the Limitations, Computational Cost and Training Set Size section in the electronic supplement).

In Figure 2, examples of filtered, three-component observed, and corresponding synthetic displacement waveforms recorded at station POTR are given. We set  $\Delta T = 60$  s, that is we collect data for one minute after the earthquake has been detected at the first receiver. We assume that the noise is temporally and spatially uncorrelated, that is  $\mathbf{C}_d$  is diagonal, and we assign a standard deviation of 0.004 m to the horizontal and 0.01 m to the vertical component, estimates obtained from 600 s of pre-event data at the 34 receivers, filtered in the same way as the synthetic data. The realistic estimation of the observational noise covariance is in general difficult and outside the scope of this article. A discussion in the context of high-rate GPS time series is given by Genrich and Bock (2006). In general, we found that reported noise levels for this type of data vary between a fraction of millimeters and centimeters for the horizontal components and millimeters to several centimeters in the vertical components, depending on which preprocessing steps are applied to the raw GPS time series.

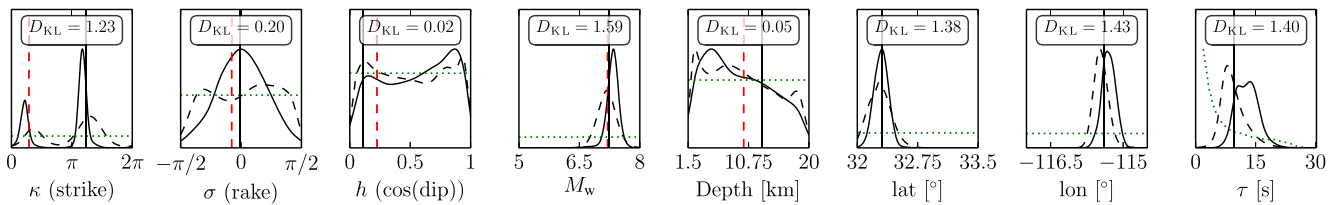
Subsequently, we train ensembles of MDNs using 95% of the synthetic examples during training and keep the remaining 5% as an independent test set. This is used to assess the prediction accuracy and resolving power of the data as discussed below. The training procedure is described in Käufel *et al.* (2014).

#### Inversion of the El Mayor–Cucapah GPS Dataset

Having obtained a set of trained MDN ensembles, we can perform inversions by presenting the preprocessed observed input vectors to the neural networks. We thus invert one minute of waveform data recorded after the 2010  $M_w$  7.2 El Mayor–Cucapah earthquake and obtain the posterior PDFs shown in Figure 3. Examples of observed displacement waveforms are depicted in Figure S3. We use all 34 CRTN receivers located within the modeling domain and



**Figure 2.** The synthetic (dashed line) and observed (solid line) GPS waveforms at station POTR, at an epicentral distance of 121 km of the  $M_w$  7.2 El Mayor–Cucapah event. The remaining static offset is included in the synthetic waveforms. The shaded area corresponds to one standard deviation of the observational noise, estimated from 600 s of pre-event data. The synthetics have been calculated for the 1D crustal model and source given in Zheng *et al.* (2012). Inset shows the source–receiver geometry. The color version of this figure is available only in the electronic edition.



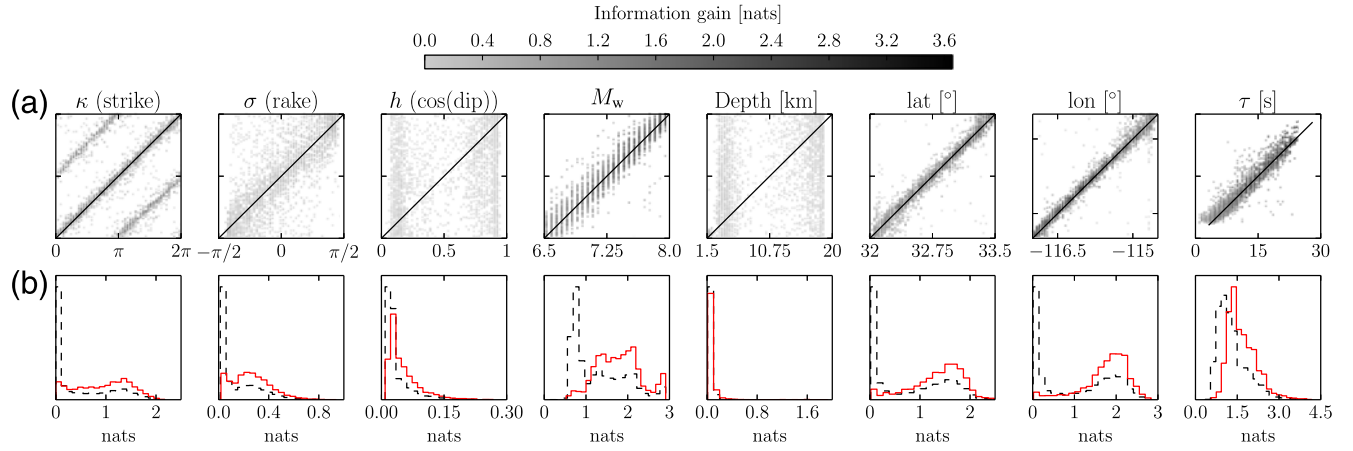
**Figure 3.** The posterior probability density functions (PDFs) for the El Mayor–Cucapah observed GPS waveform dataset using 60 s (solid curve) and 12 s of data (dashed curve). The vertical lines denote the position of the centroid moment tensor catalog solution (solid line) and the solution by Zheng *et al.* (2012) (dashed line) for this event. Zheng *et al.* (2012) did not invert for epicentral location and half duration. The dotted curve shows the prior distribution for each parameter. The information gain  $D_{KL}$  (see the Discussion section) represents a measure of how much the posterior distribution differs from the prior distribution. The color version of this figure is available only in the electronic edition.

deliberately do not remove any stations with poor data quality or missing data. This is to simulate a (near-) real-time setup, in which we have to deal with missing data and independent quality criteria might be unavailable. In fact stations P496, P492, P503, and P481 did not record any data at all during the event and contain zeros, but they are presented to the networks nonetheless. Although the posterior modes agree well with the CMT catalog solution (solid vertical lines in Fig. 3) within the uncertainty bounds, our result suggests a slightly increased half duration compared to the value obtained using relation (3) with  $a = 1.05$ . This is in agreement with finite-fault inversion results, obtained using Interferometric Synthetic Aperture Radar, geodetic and seismic data, that have revealed a relatively long rupture over a complex system of subfaults with total length of  $\sim 120$  km and a duration of more than 40 s (Hauksson *et al.*, 2010; Wei *et al.*, 2011).

The ambiguity in strike has been observed before for this event by Zheng *et al.* (2012) and Käüfl *et al.* (2014). For a perfectly vertical fault plane ( $h = 0$ ) and the slip being in strike direction ( $\sigma = 0$ ), the radiation pattern becomes invariant under a rotation of strike by  $180^\circ$ . Given that the fault orientation is close to such a strike-slip mechanism for this event, we cannot discriminate between the two orientations given the uncertainties in  $\sigma$  and  $h$ .

Furthermore, we compare our results to those obtained recently by Zheng *et al.* (2012), who inverted a 5 Hz GPS displacement waveform dataset for focal mechanism, magnitude, and depth using the cut-and-paste method (Zhu and Helmberger, 1996) for the same earthquake. Their source parameter estimates are indicated by vertical dashed lines in Figure 3. Our results agree well with theirs, and the differences between their solution and the CMT catalog solution are entirely explained by our uncertainty bounds.

We conclude that it is possible to infer a meaningful point-source description from near-field displacement data in near-real time, even if, mainly due to neglected 3D structure and source finiteness, the synthetics are far from being able to explain all the complexity of the observed waveforms (an example is given in Fig. 2). However, we find that not all parameters of a double-couple centroid moment tensor can be constrained well under the given assumptions on the observational noise and structural uncertainties using a 12 s period range and above. In particular, we cannot constrain source depth and dip. The depth resolution is limited, mainly due to the relatively low-frequency data corresponding to average wavelengths of around 60 km. Moreover, because centroid depth is known to trade off with 1D structure (Fan and Wallace, 1991), we expect that the marginalization over 1D crustal structure limits the depth resolution even further.



**Figure 4.** Resolution analysis using 5000 synthetic examples drawn from the prior distribution. (a) Position of the mode of the posterior distribution versus the known target value for the portion of the test set with  $M_w \geq 6.5$ . The gray scale shows the information gain  $D_{\text{KL}}$  (equation 5). (b) Distribution of the information gain for each parameter for all test set examples (dashed curve) and only the portion with  $M_w \geq 6.5$  (solid curve). The parameters of larger magnitude events are typically predicted with greater precision due to their increased signal-to-noise ratio. The color version of this figure is available only in the electronic edition.

## Discussion

### Resolution Analysis

To obtain a more complete picture of the achievable resolution under the given assumptions and to evaluate the performance of the method, we perform a large number of synthetic inversions using the 5000 test examples that were not used during network training. As in Käufel *et al.* (2014), we use the Kullback–Leibler divergence  $D_{\text{KL}}$  to quantify the difference between the posterior and prior marginal PDF for each parameter and thus measure how much we have learned from seeing a particular observation:

$$D_{\text{KL}}(m_k) = \int \ln \left( \frac{p(m_k|\mathbf{d})}{p(m_k)} \right) p(m_k|\mathbf{d}) dm_k, \quad (5)$$

in which  $D_{\text{KL}} \geq 0$  with equality only if  $p(m_k|\mathbf{d}) = p(m_k)$ , that is, the posterior equals the prior.

We present each test set example to the MDN ensemble for each parameter and obtain a prediction in the form of a set of posterior PDFs. Figure 4a shows the mode of the predicted PDF plotted against the true target value for each example. If a parameter were to be perfectly resolved, we would expect to see a linear correlation indicating that predictions are precise. However, a larger scatter indicates that the mode of the distribution deviates from the target value, typically accompanied by a broadened distribution. To quantify the precision of the prediction for each example, we calculate the information gain  $D_{\text{KL}}$  between prior and posterior distribution. The information gain associated with each example corresponds to the gray scale in Figure 4a to enable a visual quality assessment. Examples with a lower prediction accuracy, for which the posterior mode differs more from the target value, consistently have a lower information gain, indicating a broader PDF as expected. Histograms of the information gain

over the entire test set can give an impression of how well a parameter can be resolved (Fig. 4b). It is also insightful to consider the average conditional information gain  $D_{\text{KL}}(m_i|\mathbf{d}, m_j \in [a, b])$ , that is, the information gain for a parameter  $m_i$ , given that parameter  $m_j$  falls into the bin  $[a, b]$ . This enables us to analyze to what extent parameter  $m_i$  is constrained, depending on the value of parameter  $m_j$ . Results for a number of parameter combinations are depicted in Figure 5. See Figure S4 for a complete overview.

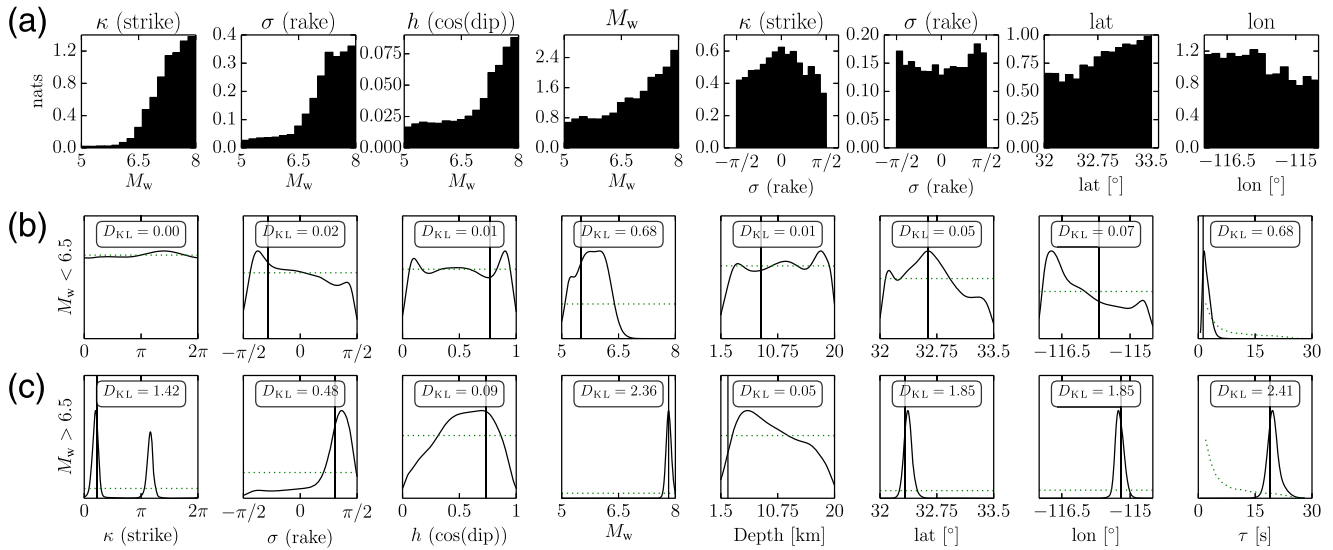
This analysis reveals that we cannot resolve dip and depth at all, because  $D_{\text{KL}} \lesssim 0.2$  nats for almost all examples. All other parameters can only be resolved well for events larger than  $M_w 6.5$  (Figs. 4b and 5), which is due to the poor SNR for the smaller events. Furthermore, the source orientation and the source location have an influence on the resolution of some parameters (Fig. 5a, panels 5–8), which is a consequence of the uneven station distribution (compare to Fig. 1). Note that  $D_{\text{KL}}(M_w|\mathbf{d}) > 0$ , even for events smaller than  $M_w 6.5$  (Fig. 5c, panel 4), indicating that we are able to discriminate smaller from larger events even if the signal is below noise level for most or all of the stations. Figure 5b and 5c shows posterior PDFs for small and large examples, respectively. If the signal is below noise level at (almost) all of the stations, all we can say is that an event is certainly smaller than  $M_w 6.5$ , as is the case for the event shown in Figure 5b.

### Length of the Data Window

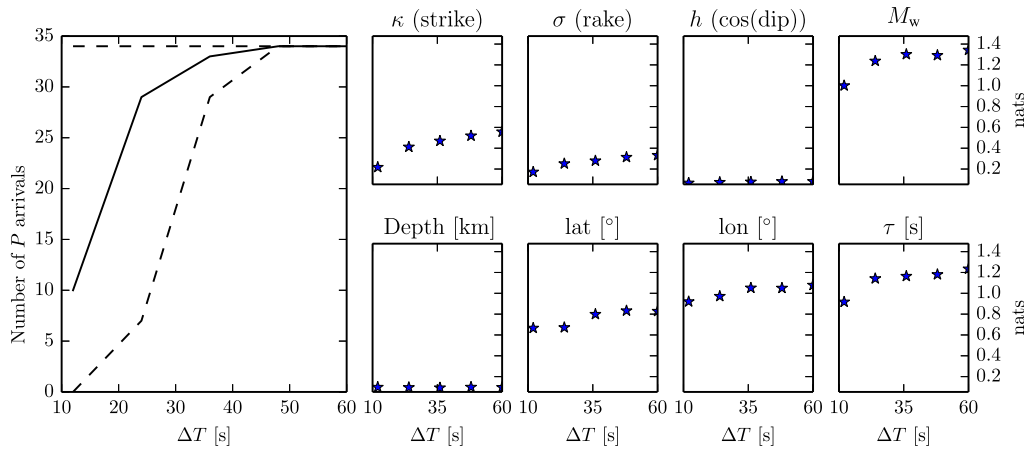
For EEW, it is crucial to obtain a solution as rapidly as possible. We investigate the effect of the window length on the resolution by training four additional sets of networks for  $\Delta T \in \{12, 24, 36, 48 \text{ s}\}$  and evaluating the average information gain of the test set

$$\overline{D_{\text{KL}}} = \sum_i D_{\text{KL}}(m_k|\mathbf{d}_i)/N, \quad (6)$$





**Figure 5.** (a) The information gain for the parameter in the axis title averaged over all test examples falling into the bins for the parameter plotted on the horizontal axis (see the [Resolution Analysis](#) section). (b) Posterior PDFs for a relatively small example and (c) a relatively large example. The vertical line indicates the true (target) value; the dotted curve indicates the prior distribution. As expected, we are able to discern more information about the large event than about the small event. The color version of this figure is available only in the electronic edition.

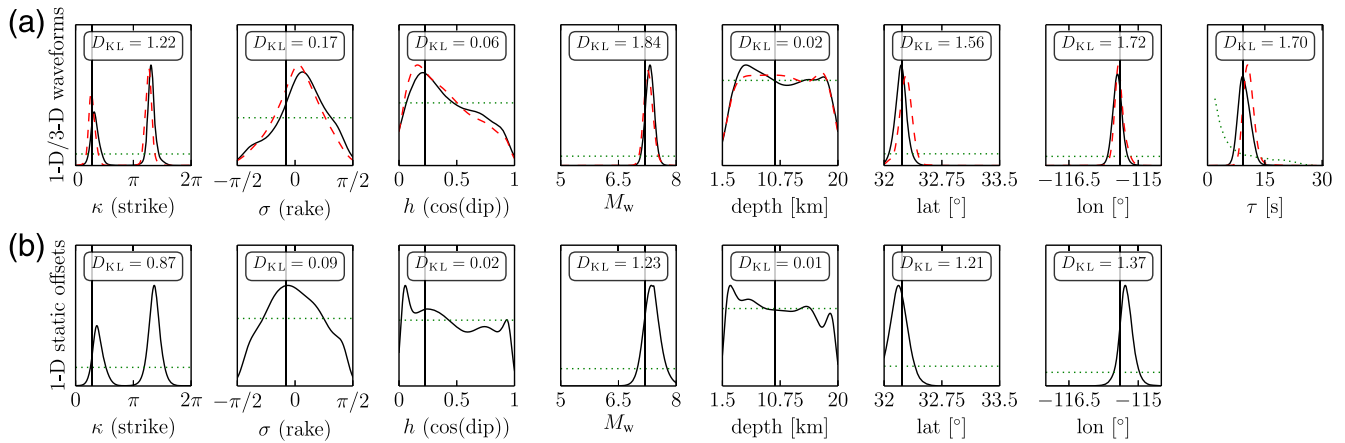


**Figure 6.** (left) Average (solid line), minimum and maximum (dashed line) number of  $P$ -wave arrivals that occur within the data window of length  $\Delta T$ . (right) Average information gain over the test set in nats plotted versus  $\Delta T$  for each parameter. The color version of this figure is available only in the electronic edition.

in which  $N$  is the number of test set examples. It serves as a proxy for the resolving power of the data. An increase in  $\overline{D}_{\text{KL}}$  indicates that the uncertainties on the test set predictions have decreased on average. The results are shown in Figure 6 and indicate that, for all resolved parameters, a significant increase in average information gain with the length of the data window can be observed. This is expected, because as time progresses, an increasing number of stations provide data (left-most panel in Fig. 6). Furthermore, stations closer to the epicenter will gradually include later arriving phases and reach the final static offset. This analysis indicates that we can potentially reduce the window length to  $\Delta T \approx 40$  s without losing much of the resolving power.

#### Information Provided by the Waveform Data

Recently, Käüfl *et al.*, (2014) inverted static offsets using the same methodology for the El Mayor–Cucapah event. It is interesting to compare the results obtained by inverting full waveforms with those obtained by inverting the coseismic static offsets only. Because we are using slightly different prior ranges, a different noise estimate (owing to the real-time nature of the data in this work) and a different station configuration from that adopted in Käüfl *et al.* (2014), these two sets of results are not directly comparable. Instead, we perform a synthetic experiment, in which we simulate noisy static displacements, invert these as in Käüfl *et al.* (2014) and compare the result to those obtained for the respective



**Figure 7.** Inversion of (a) synthetic 1D waveforms (solid line) and 3D waveforms (dashed line) and (b) static displacements only for the El Mayor event. (a) The differences between 1D and 3D waveform inversions are negligible. The  $D_{KL}$  values correspond to the 1D dataset. (b) Results of inverting static offsets only, as per Käufel *et al.* (2014). A comparison of the information gain with the waveform inversion results reveals, that although the  $D_{KL}$  values are slightly higher in the latter cases, we cannot learn significantly more from the waveform data in this case. The color version of this figure is available only in the electronic edition.

synthetic waveform datasets. Posterior PDFs are shown in Figure 7b. We assume that the uncertainty of the static offset measurements is similar to the noise level of the dynamic movement and use a noise level of 4 mm for the horizontal and 10 mm for the vertical component. As can be seen, using waveform data in general increases the information gain, but dip and source depth remain unresolved.

However, even if the additional information gained using the full waveforms over static offsets may be limited in this case, it remains advantageous to use the waveform data directly. To invert static offsets, we must first wait until the dynamic ground movement has ceased at all stations. It is possible to obtain a first estimate of the final static offset roughly 10 s after the detection of the first dynamic movement (Allen and Ziv, 2011), which is typically at the time of  $S$ -wave arrival. Therefore, we could obtain early estimates of the static offset at all of the 34 stations in the case of the El Mayor event at around 50–60 s after the  $P$  wave has been detected at the first station. This is similar to the length of the waveform time window used here. However, the noise level of these early estimates of the static offset is likely to be significantly higher than in our synthetic test, with consequent deterioration of the achievable resolution. Moreover, we have seen that shorter-waveform time windows could be used without much loss of precision (Fig. 6).

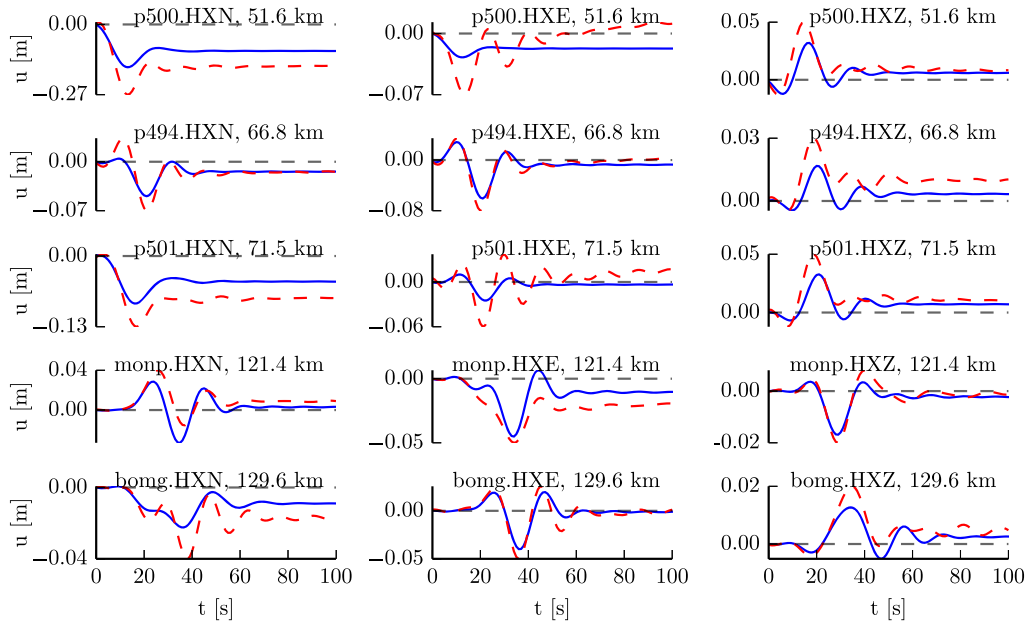
#### Robustness with Respect to Noise and Unmodeled Physical Effects

**Neglected 3D Crustal and Upper-Mantle Structure.** Because of the high computational demands of accurate 3D simulations in the highly heterogeneous southern California crust, we restricted ourselves to 1D layered Earth models. However, incorporating 3D synthetics would not alter the time required for evaluation of the trained neural networks and would be straightforward, given enough computational

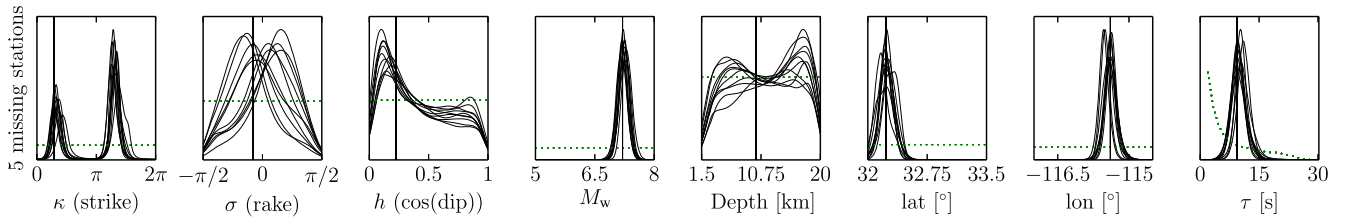
resources are assigned for the generation of the training set, which has to be done only once for a particular region.

To investigate how the neglect of 3D structure affects the neural network output, we first generate synthetics in a 1D layered Earth model (Table S2) using a double-couple moment tensor obtained by Zheng *et al.* (2012) (Table S3). Second, we use the 3D southern California crustal and upper-mantle model CVMH 11.9 (Tape, 2009; C. Tape, personal comm., 2014) and generate accurate 3D synthetic waveforms using SPEC-FEM3D (Tromp *et al.*, 2008). Some examples for 1D and 3D synthetic waveforms are depicted in Figure 8 and indicate a substantial difference between 1D and 3D synthetics. Both, the dynamic and the static wavefield are affected by the 3D structure. The linear trend, which can be observed at some of the 3D traces (e.g., p500.HXE) is an artifact most likely caused by the finiteness of the modeling volume, which affects the static offsets much more than the dynamic wavefield (Tape *et al.*, 2011). This artifact is unlikely to bias our results, because we are using relatively short initial time windows, in which the linear build up is not significant. Moreover, while the static part of the 3D synthetics might not represent a physical signal, this inversion merely serves as a test to assess the robustness of the trained networks, and we expect that the general characteristics of the 3D synthetics are representative of the effects of a heterogeneous mantle and crust.

Both datasets are presented to the neural networks trained using only 1D data and yield the posterior PDFs shown in Figure 7a. The differences between the resulting PDFs are minor, and we conclude that our inversion is robust with respect to neglected 3D crustal and upper-mantle structure for this example. However, this invariance does not indicate that a training set based on 3D synthetics would not be beneficial. Such a training set could significantly improve the resolution, in particular for the centroid location and depth,



**Figure 8.** Examples of synthetic displacement waveforms, low-pass filtered at 12 s, calculated in a 1D layered model (solid line) using a wave propagation code by O’Toole and Woodhouse (2011) and a 3D Earth model using SPECFEM3D (dashed line). The color version of this figure is available only in the electronic edition.



**Figure 9.** The signal at five stations selected at random has been replaced by random noise in 10 cases. The well-resolved parameters prove robust with respect to this type of distortion. The color version of this figure is available only in the electronic edition.

because it would potentially enable us to use higher-frequency data and forego the marginalization over 1D crustal structure.

*Missing Data.* Although errors caused by using a 1D, rather than a 3D Earth model, have partly been accounted for during training by marginalizing over 1D average Earth models, other sources of error have not been taken into account. For example, the observed displacement waveform dataset used throughout this article includes several stations that do not have any clear arrivals and show spurious signal or no data at all. To test to what extent such perturbations affect the output of the neural networks, we replaced the signal at five stations selected at random by random noise to simulate missing data. Although depending on the actual implementation, missing data might in fact rather be represented by zeros or other fixed values, we could easily modify any implementation to replace those values by random noise. The resulting posterior PDFs in 10 cases are shown in Figure 9. The well-resolved parameters prove very robust

with respect to this type of perturbation, in the sense that the posterior PDF assigns a high probability in the vicinity of the target value. Most unstable are the results for  $\sigma$ ,  $h$ , and depth, which are the least well-constrained parameters. The event latitude seems to be more sensitive to the perturbation than the longitude, which is a consequence of the one-sided station distribution. We expect that the robustness could be further improved upon by explicitly including examples with missing or distorted data in the training set, which has not been done here.

*Finiteness of the Earthquake Source.* Near-field observations for midsize to large events typically show sensitivity to the finite extent of the rupture area. This effect is not represented in the neural network training set, which purely consists of point-source examples. In the context of EEW, naturally we are interested in utilizing the first observations that are available for an ongoing earthquake. In particular in well-instrumented areas such as southern California, this involves the evaluation of signal from stations close to the

earthquake rupture front. Clearly, we cannot expect to be able to explain the complex near-field radiation pattern of a realistic finite-source distribution with a point source, but note that finding a best-fitting point source is still a theoretically well-defined concept. However, traditional approaches typically require strong regularization if near-field data are to be used, because the associated optimization problem becomes unstable and highly nonunique. In particular, in the case of the El Mayor–Cucapah event, other researchers have found a need to apply strong damping to suppress spurious non-double-couple components and to stabilize the iterative optimization process when inverting a similar static displacement dataset as the one used here for point-source parameters (O’Toole *et al.*, 2013). Here and in Käufel *et al.* (2014), there appears to be no need for additional measures to account for finite-source effects in the case of this event. This can potentially be understood by noting that our approach is based on samples that have been obtained prior to the inversion (see also the [Inference with Prior Samples](#) section). As a consequence, the relative weight given to the signal at different stations is determined during the training stage and independent of any particular observation that may be contaminated by an imprint of the spatial extent of the source. The inversion is therefore less likely to be dominated by the subset of stations that are strongly affected by finite-source effects than it would be in a traditional inversion framework, which involves the minimization of a misfit functional given a particular observation. Of course the problem is expected to become more severe if the rupture length becomes comparable to the extent of the station network and the majority of the receivers record a radiation pattern much different from that of a point source.

Undoubtedly, however, for certain applications a point-source description will not be sufficient. In particular, if the source model is to be used for ground-motion prediction in the close vicinity of the rupture area, the point-source radiation is unlikely to yield realistic results. Therefore, we suggest that future research should focus on the extension of the presented methodology to account for finite-fault effects. This appears to be a realistic goal, if a relatively low-dimensional parametrization is chosen, for example, by using an approach as in Vallée and Bouchon (2004) or Twardzik *et al.* (2014), where the rupture process is described by a finite number of elliptical subfault patches of varying size and orientation.

### Inference with Prior Samples

The observed robustness with respect to perturbations of the input is mainly a consequence of our approach relying on prior rather than posterior sampling. The inversion process does not involve the minimization of a misfit functional, which is prone to overfit to unmodeled signal. Rather the neural networks encapsulate a compact representation of a database of synthetic waveforms, to which each observed wavefield pattern is compared implicitly. A consequence of this approach is that we have no opportunity to refine

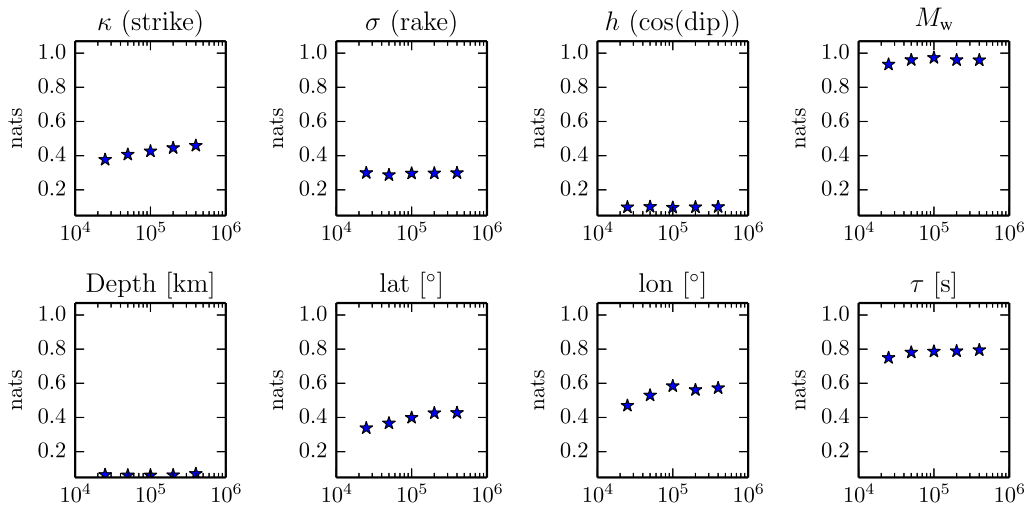
our estimate in high-likelihood regions, which would require generating samples from the actual posterior distribution. However, it has the advantage that we separate the time-consuming sampling stage from the inversion. Furthermore, we are less likely to be affected by outliers and noise. Most misfit measures can easily be dominated by outliers caused by missing stations or unmodeled physical effects (as discussed above). In our approach, the contribution of each individual trace to the network output is determined *a priori* during the training stage and thus is largely independent of the actual observation. The use of prior rather than posterior samples has further implications, which we discuss next.

We find that not all parameters can be resolved equally well. Factors that may influence the resolution include the observational uncertainties propagated into the model space; the nonlinearity of the inverse problem, which may lead to an intrinsic nonuniqueness; the differing sensitivity of the data with respect to a specific parameter; and the limited spatial data coverage and the frequency content of the dataset. However, in a framework that relies solely on prior samples, the density of prior samples in the joint data model space also imposes limitations on the resolving power of the model.

First, consider the extreme case of a complete lack of training samples. This would result in the MDN output being independent of the input and mimicking the prior distribution (due to the initialization, see Bishop, 1995; Käufel *et al.*, 2014). As more samples are presented during training, the output distribution will progressively be shaped to mimic the distribution of the training set, until the quality of predictions for an independent validation set (which is constantly evaluated during the training process) deteriorates, at which point we halt the training procedure. This prevents the model from overfitting to the training set, a technique known as early stopping. Because the training procedure involves a nonconvex optimization problem (Bishop, 1995), it is prone to local minimums. Therefore, we train sets of networks with different numbers of hidden units and different random initializations of the network weights and form ensembles, which average out the differences of the individual networks (Käufel *et al.*, 2014).

Our experiments suggest that these measures lead to a situation in which an undersampling of the model space will not result in biased, but rather less precise predictions (that is, broader distributions). In that sense, our results can be considered to be conservative.

*Required Number of Prior Samples.* We have seen that the number of prior samples is a limiting factor on the posterior uncertainty. The prior sampling density necessary to make meaningful inferences is related to the expected precision of the observations via the inverse mapping. That is, differences between two arbitrary data vectors only have to be explained up to the order of the noise, which imposes a lower limit on the required sampling density in the model space. Unfortunately, there is no straightforward way to estimate the required sampling density without explicitly knowing the in-



**Figure 10.** Information gain as a function of the number of prior samples in the 25,000–400,000 range. The color version of this figure is available only in the electronic edition.

verse mapping. A pragmatic approach is therefore to train networks with an increasing quantity of prior samples, until the posterior uncertainties of the test set examples no longer decrease. Figure 10 shows the average information gain (equation 6) as a function of the number of prior samples. These prior samples are replicated, so that the total number of training examples is comparable in each case (see also the Limitations, Computational Cost and Training Set Size section in the electronic supplement). These experiments were performed using a reduced set of 15 receivers to limit the computational cost. Therefore, all information gain values are lower than those shown in Figure 4, for which 34 stations were used.

For all parameters, the average information gain increases with the number of prior samples, particularly for the better-resolved parameters. This indicates that a coarser prior sampling results in a lower resolution. For the other parameters, the influence of the prior sampling density is insignificant in the examined range. This can indicate that with the given noise model a denser sampling would not improve the resolving power, but we cannot exclude the possibility that regions with high-likelihood models are missed out completely. From this experiment, we conclude that a number of  $\sim 100,000$  samples seems to suffice. In general, an improvement in precision will always require a larger investment in computational resources during training. For further discussion, see also the Limitations, Computational Cost and Training Set Size section in the electronic supplement.

### Conclusions

We have introduced a nonlinear method for the rapid characterization of earthquakes based on noisy near-field and regional observations. The method is fully probabilistic and posterior probabilities are modeled as mixtures of Gaussian distributions. It can cope well with the strongly

event-dependent SNR of the waveform data. Uncertainties in the crustal model are taken into account, as is observational noise. Synthetic tests suggest that the results are very robust with respect to unmodeled effects, such as those caused by 3D structure and missing input data.

We have demonstrated that our approach is flexible with respect to the type of input data. It can potentially be used in a straightforward way to jointly invert different data types, such as strong-motion and GPS waveform data, even if stations are not collocated. Finally, the method in principle allows for the incorporation of highly accurate and large synthetic waveform databases (e.g., obtained by running computationally expensive wave-propagation simulations—without significantly increasing the computational cost for the inversions itself) because the requirements of the trained neural networks are modest in terms of memory and computation time, even for relatively large networks. If implemented efficiently, a single inversion can be done within a fraction of a second on a current standard desktop computer.

### Data and Resources

The network-adjusted GPS displacement waveform dataset for the El Mayor–Cucapah event can be downloaded from <http://scedc.caltech.edu/research-tools/MayorCucapah20100404/> (last accessed June 2015). The 3D crustal model CRUST 1.0 can be obtained from <http://igppweb.ucsd.edu/gabi/crust1.html> (last accessed September 2014). The following software packages and libraries have been used for this work: the 1D wave propagation code by O’Toole and Woodhouse (2011), SPEC-FEM3D (Tromp *et al.*, 2008), the TauP package (Crotwell *et al.*, 1999), ObsPy ([www.obspy.org](http://www.obspy.org), last accessed June 2015; Beyreuther *et al.*, 2010), ALGLIB ([www.alglib.net](http://www.alglib.net); last accessed June 2015), and PyBrain ([www.pybrain.org](http://www.pybrain.org), last accessed June 2015; Schaul *et al.*, 2010).

## Acknowledgments

We thank Carl Tape for his support with setting up the 3D southern California spectral element simulations, Tom O'Toole for providing his 1D numerical wave propagation code, and Andreas Fichtner and an anonymous reviewer for their useful comments that helped to improve the article. Paul Käüfl, Andrew Valentine, and Ralph De Wit are funded by the Netherlands Organisation for Scientific Research (NWO) under Grant ALW Top-subsidy 854.10.002.

## References

- Aki, K., and P. G. Richards (1980). *Quantitative Seismology*, Vol. II, University Science Books, Sausalito, California.
- Allen, R. M., and H. Kanamori (2003). The potential for earthquake early warning in southern California, *Science* **300**, no. 5620, 786–789.
- Allen, R. M., and A. Ziv (2011). Application of real-time GPS to earthquake early warning, *Geophys. Res. Lett.* **38**, no. 16, 1–7.
- Bernardi, F., J. Braunmiller, U. Kradolfer, and D. Giardini (2004). Automatic regional moment tensor inversion in the European-Mediterranean region, *Geophys. J. Int.* **157**, no. 2, 703–716.
- Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.* **81**, no. 3, 530–533.
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, Vol. 92, Oxford University Press, Oxford, United Kingdom.
- Blewitt, G., W. C. Hammond, C. Kreemer, H.-P. Plag, S. Stein, and E. Okal (2009). GPS for real-time earthquake source determination and tsunami warning systems, *J. Geodes.* **83**, nos. 3/4, 335–343.
- Bock, Y., D. Melgar, and B. W. Crowell (2011). Real-time strong-motion broadband displacements from collocated GPS and accelerometers, *Bull. Seismol. Soc. Am.* **101**, no. 6, 2904–2925.
- Chapman, C. H., and W. S. Leaney (2012). A new moment-tensor decomposition for seismic events in anisotropic media, *Geophys. J. Int.* **188**, no. 1, 343–370.
- Crotwell, H. P., T. J. Owens, and J. Ritsema (1999). The TauP toolkit: Flexible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.* **70**, no. 2, 154–160.
- Crowell, B. W., Y. Bock, and D. Melgar (2012). Real-time inversion of GPS data for finite fault modeling and rapid hazard assessment, *Geophys. Res. Lett.* **39**, no. 9, L09305, doi: [10.1029/2012GL051318](https://doi.org/10.1029/2012GL051318).
- Cua, G., and T. H. Heaton (2007). The virtual seismologist (VS) method: A Bayesian approach to earthquake early warning, in *Earthquake Early Warning Systems*, P. Gasparini, G. Manfredi, and J. Zschau (Editors), Springer, Berlin, Germany, 97–132.
- de Wit, R. W., P. Käüfl, A. P. Valentine, and J. Trampert (2014). Bayesian inversion of free oscillations for Earth's radial (an)elastic structure, *Phys. Earth Planet. In.* **237**, 1–17.
- de Wit, R. W., A. P. Valentine, and J. Trampert (2013). Bayesian inference of Earth's radial seismic structure from body-wave travel times using neural networks, *Geophys. J. Int.* **195**, no. 1, 408–422.
- Dziewonski, A. M., T.-A. Chou, and J. H. Woodhouse (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. Geophys. Res.* **86**, no. B4, 2825–2852.
- Ekström, G., M. Nettles, and A. M. Dziewonski (2012). The Global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Phys. Earth Planet. In.* **200/201**, 1–9.
- Fan, G., and T. C. Wallace (1991). The determination of source parameters for small earthquakes from a single, very broadband seismic station, *Geophys. Res. Lett.* **18**, no. 8, 1385–1388.
- Fichtner, A., H.-P. Bunge, and H. Igel (2006). The adjoint method in seismology, *Phys. Earth Planet. In.* **157**, nos. 1/2, 86–104.
- Geller, R., D. D. Jackson, Y. Y. Kagan, and F. Mulargia (1997). Earthquakes cannot be predicted, *Science* **275**, 1616–1618.
- Geng, J., Y. Bock, D. Melgar, B. W. Crowell, J. S. Haase, H. Cecil, and M. Ida (2013). A new seismogeodetic approach applied to GPS and accelerometer observations of the 2012 Brawley seismic swarm: Implications for earthquake early warning, *Geochem. Geophys. Geosyst.* **14**, 2124–2142.
- Genrich, J. F., and Y. Bock (2006). Instantaneous geodetic positioning with 10–50 Hz GPS measurements: Noise characteristics and implications for monitoring networks, *J. Geophys. Res.* **111**, no. B03403, doi: [10.1029/2005JB003617](https://doi.org/10.1029/2005JB003617).
- Hauksson, E., J. Stock, K. Hutton, W. Yang, J. A. Vidal-Villegas, and H. Kanamori (2010). The 2010  $M_w$  7.2 El Mayor–Cucapah earthquake sequence, Baja California, Mexico and southernmost California, USA: Active seismotectonics along the Mexican Pacific margin, *Pure Appl. Geophys.* **168**, nos. 8/9, 1255–1277.
- Jaynes, E. T. (2005). Probability theory: The logic of science, *Math. Intel.* **27**, no. 2, 83–83.
- Kanamori, H. (1993). Locating earthquakes with amplitude: Application to real-time seismology, *Bull. Seismol. Soc. Am.* **83**, no. 1, 264–268.
- Kanamori, H. (2005). Real-time seismology and earthquake damage mitigation, *Ann. Rev. Earth Planet. Sci.* **33**, no. 1, 195–214.
- Käüfl, 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, *Geophys. J. Int.* **196**, no. 3, 1676–1693.
- Larson, K. M., P. Bodin, and J. Gomberg (2003). Using 1-Hz GPS data to measure deformations caused by the Denali fault earthquake, *Science* **300**, no. 5624, 1421–1424.
- Laske, G., G. Masters, Z. Ma, and M. Pasyanos (2013). Update on CRUST1.0—A 1-degree global model of Earth's crust, *Geophys. Res. Abstr.* **15**, abstract EGU2013-2658.
- Lovely, P., J. H. Shaw, Q. Liu, and J. Tromp (2006). A structural  $V_p$  model of the Salton Trough, California, and its implications for seismic hazard, *Bull. Seismol. Soc. Am.* **96**, no. 5, 1882–1896.
- Madariaga, R. (2007). Seismic source theory, *Treatise Geophys.* **4**, 59–82.
- McLachlan, G. J., and K. E. Basford (1988). *Mixture Models. Inference and Applications to Clustering*, Statistics: Textbooks and Monographs, Dekker, New York.
- Meier, U., A. Curtis, and J. Trampert (2007). Global crustal thickness from neural network inversion of surface wave data, *Geophys. J. Int.* **169**, no. 2, 706–722.
- Melgar, D., Y. Bock, and B. W. Crowell (2012). Real-time centroid moment tensor determination for large earthquakes from local and regional displacement records, *Geophys. J. Int.* **188**, no. 2, 703–718.
- Melgar, D., B. W. Crowell, Y. Bock, and J. S. Haase (2013). Rapid modeling of the 2011  $M_w$  9.0 Tohoku-Oki earthquake with seismogeodesy, *Geophys. Res. Lett.* **40**, no. 12, 2963–2968.
- Nakamura, Y. (1988). On the urgent earthquake detection and alarm system (UrEDAS), *Proc. of the 9th World Conference on Earthquake Engineering*, Tokyo-Kyoto, 2–9 August 1988, Vol. 7.
- Ohta, Y., T. Kobayashi, H. Tsushima, S. Miura, R. Hino, T. Takasu, H. Fujimoto, T. Iinuma, K. Tachibana, T. Demachi, et al. (2012). Quasi real-time fault model estimation for near-field tsunami forecasting based on RTK-GPS analysis: Application to the 2011 Tohoku-Oki earthquake ( $M_w$  9.0), *J. Geophys. Res.* **117**, no. B02311, doi: [10.1029/2011JB008750](https://doi.org/10.1029/2011JB008750).
- O'Toole, T. B., and J. H. Woodhouse (2011). Numerically stable computation of complete synthetic seismograms including the static displacement in plane layered media, *Geophys. J. Int.* **187**, no. 3, 1516–1536.
- O'Toole, T. B., A. P. Valentine, and J. H. Woodhouse (2012). Centroid-moment tensor inversions using high-rate GPS waveforms, *Geophys. J. Int.* **191**, no. 1, 257–270.
- O'Toole, T. B., A. P. Valentine, and J. Woodhouse (2013). Earthquake source parameters from GPS-measured static displacements with potential for real-time application, *Geophys. Res. Lett.* **40**, 60–65.
- Ritsema, J., and T. Lay (1993). Rapid source mechanism determination of large ( $M_w \geq 5$ ) earthquakes in the western United States, *Geophys. Res. Lett.* **20**, no. 15, 1611–1614.
- Satriano, C., A. Lomax, and A. Zollo (2008). Real-time evolutionary earthquake location for seismic early warning, *Bull. Seismol. Soc. Am.* **98**, no. 3, 1482–1494.

- Schaul, T., J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Sehnke, T. Rückstieß, and J. Schmidhuber (2010). PyBrain, *J. Machine Learn. Res.* **11**, 743–746.
- Shahraeeni, M. S., A. Curtis, and G. Chao (2012). Fast probabilistic petrophysical mapping of reservoirs from 3D seismic data, *Geophysics* **77**, no. 3, O1–O19.
- Tape, C. (2009). Adjoint tomography of the southern California crust, *Science* **325**, 988–992.
- Tape, C., J. Loveless, and B. Meade (2011). Seismogram-based assessment of the southern California seismic velocity model CVM-H 11.9 with 234 reference earthquakes, in *Southern California Earthquake Center Annual Meeting*, Proceedings and Abstracts, Palm Springs, California, 11–14 September 2011, 235.
- Tape, W., and C. Tape (2012). A geometric setting for moment tensors, *Geophys. J. Int.* **190**, no. 1, 476–498.
- Tarantola, A. (2005). *Inverse Problem Theory*, Vol. 4, SIAM, Philadelphia, Pennsylvania.
- Tromp, J., D. Komattisch, and Q. Liu (2008). Spectral-element and adjoint methods in seismology, *Comm. Comput. Phys.* **3**, no. 1, 1–32.
- Twardzik, C., S. Das, and R. Madariaga (2014). Inversion for the physical parameters that control the source dynamics of the 2004 Parkfield earthquake, *J. Geophys. Res.* **119**, no. 9, 7010–7027.
- Vallée, M., and M. Bouchon (2004). Imaging coseismic rupture in far field by slip patches, *Geophys. J. Int.* **156**, no. 3, 615–630.
- Walker, M., and A. Curtis (2014). Varying prior information in Bayesian inversion, *Inverse Probl.* **30**, no. 6, doi: [10.1088/0266-5611/30/6/065002](https://doi.org/10.1088/0266-5611/30/6/065002).
- Walter, W. (1993). Source parameters of the June 29, 1992 Little Skull Mountain earthquake from complete regional waveforms at a single station, *Geophys. Res. Lett.* **20**, no. 5, 403–406.
- Wei, S., E. Fielding, S. Leprince, A. Sladen, J.-P. Avouac, D. V. Helmberger, E. Hauksson, R. Chu, M. Simons, K. Hudnut, *et al.* (2011). Superficial simplicity of the 2010 El Mayor–Cucapah earthquake of Baja California in Mexico, *Nat. Geosci.* **4**, no. 9, 615–618.
- Wright, T. J., N. Houlié, M. Hildyard, and T. Iwabuchi (2012). Real-time, reliable magnitudes for large earthquakes from 1 Hz GPS precise point positioning: The 2011 Tohoku-Oki (Japan) earthquake, *Geophys. Res. Lett.* **39**, no. 12, doi: [10.1029/2012GL051894](https://doi.org/10.1029/2012GL051894).
- Wu, Y.-M., and H. Kanamori (2005). Experiment on an onsite early warning method for the Taiwan early warning system, *Bull. Seismol. Soc. Am.* **95**, no. 1, 347–353.
- Zheng, Y., J. Li, Z. Xie, and M. H. Ritzwoller (2012). 5 Hz GPS seismology of the El Mayor–Cucapah earthquake: Estimating the earthquake focal mechanism, *Geophys. J. Int.* **190**, no. 3, 1723–1732.
- Zhu, L., and D. V. Helmberger (1996). Advancement in source estimation techniques using broadband regional seismograms, *Bull. Seismol. Soc. Am.* **86**, no. 5, 1634–1641.

Department of Earth Sciences  
 Universiteit Utrecht  
 Budapestlaan 4  
 3584 CD Utrecht, The Netherlands

Manuscript received 11 January 2015;  
 Published Online 30 June 2015