

# NET-VISA: Network Processing Vertically Integrated Seismic Analysis

by Nimar S. Arora, Stuart Russell,\* and Erik Sudderth

**Abstract** The automated processing of multiple seismic signals to detect and localize seismic events is a central tool in both geophysics and nuclear treaty verification. This paper reports on a project begun in 2009 to reformulate this problem in a Bayesian framework. A Bayesian seismic monitoring system, NET-VISA, has been built comprising a spatial event prior and generative models of event transmission and detection, as well as an inference algorithm. The probabilistic model allows for seamless integration of various disparate sources of information. Applied in the context of the International Monitoring System (IMS), a global sensor network developed for the Comprehensive Nuclear Test Ban Treaty (CTBT), NET-VISA achieves a reduction of around 60% in the number of missed events compared with the currently deployed system. It also finds events that are missed by the human analysts who postprocess the IMS output.

## Introduction

The Comprehensive Nuclear Test Ban Treaty (CTBT), which bans all nuclear explosions on Earth, is gaining renewed attention in light of growing worldwide interest in mitigating the risks of nuclear weapons proliferation and testing. To monitor compliance with the treaty, the Preparatory Commission for the Comprehensive Nuclear Test Ban Treaty Organization (CTBTO) has installed a suite of sensors known as the International Monitoring System (IMS). The IMS includes waveform sensor stations (seismic, hydroacoustic, and infrasound) connected by a worldwide communications network to a centralized processing system in the International Data Center (IDC) in Vienna. The IDC operates continuously and in real time, performing station processing (analysis and reduction of raw seismic sensor data to detect and classify signal arrivals at each station) and network processing (association of signals from different stations that have presumably come from the same event). Perfect performance remains well beyond the reach of current technology: the IDC's automated system, a highly complex and well-tuned piece of software, misses nearly one-third of all seismic events in the magnitude range of interest, and about half of the reported events are spurious. A large team of expert analysts postprocesses the automatic bulletins to improve their accuracy to acceptable levels.

The IDC results indicate that the network processing problem is far from trivial. There are three primary sources of difficulty: (1) the travel time between any two points on Earth and the attenuation of various frequencies and wave types are not known accurately; (2) each detector is subject to local noise that may mask true signals and cause false

detections (as much as 90% of all detections are false); and (3) many thousands of detections are recorded per day, so the problem of proposing and comparing possible events (subsets of detections) is daunting. These considerations suggest that an approach based on probabilistic inference and combination of evidence might be effective, and this paper demonstrates that this is in fact the case. For example, such an approach automatically takes into account nondetections as negative evidence for a hypothesized event, something that classical methods typically do not do.

The existing network processing algorithm in use at the IDC, known as global association (GA) (Le Bras *et al.*, 1994), uses various heuristics to cluster arrivals, and then determines the location of the event in each cluster. The event location algorithm is based on the original iterative linear least squares method of Geiger (1910, 1912). Of course, the algorithm has been enhanced considerably over the years, for example, by the use of singular value decomposition to solve the resulting matrix equations (Menke, 1989; Lay and Wallace, 1995) and by the use of azimuth and slowness to constrain the solution (Magotra *et al.*, 1987; Roberts *et al.*, 1989). However, in the words of Myers *et al.* (2007) (p. 1049), “Seismic event location is—at its core—a minimization of the difference between observed and predicted arrival times.” Further, all of these classical methods rely only on the associated arrivals to locate an event. Even the multiple-event location algorithms, such as those due to Waldhauser and Ellsworth (2000) and Myers *et al.* (2007), ignore data from stations that fail to detect an event.

In this paper, we present a probabilistic model of seismic event occurrence,  $P_\theta$  (see Table 1 for a list of all mathematical notations), and another probabilistic model of seismic detections triggered by events (or noise),  $P_\phi$ . We also describe how the model parameters  $\theta$  and  $\phi$  are estimated from

\*Also at Bayesian Logic, Inc., 32432 Lighthouse Way, Union City, California 94587.

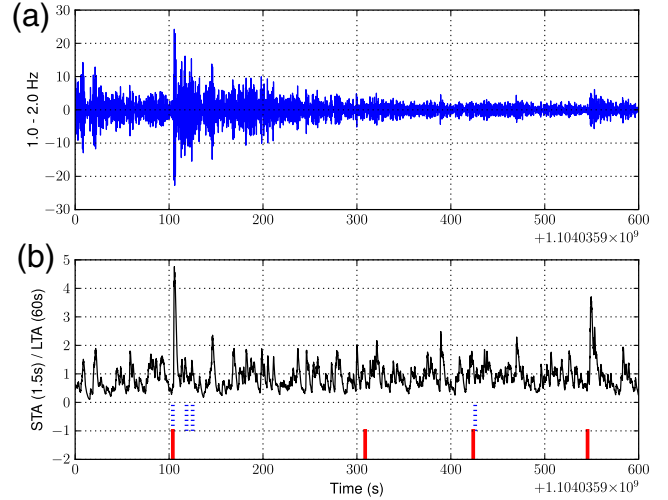
Table 1  
Mathematical Notations

Symbol	Description
$P_\theta$	Prior probability distribution of events
$P_{\theta,n}$	Prior probability of the number of events
$P_{\theta,t}, P_{\theta,d}, P_{\theta,l}, P_{\theta,m}$	Prior probability of event time, depth, location, and magnitude
$e$	Set of events; $e^i$ is the $i$ th event
$e^i, e^i_d, e^i_l, e^i_m$	Event time, depth, location, and magnitude
$P_\phi$	Probability of arrivals given an event
$P_{\phi,d}$	Detection probability of an event
$P_{\phi,z}, P_{\phi,s}, P_{\phi,a}, P_{\phi,h}$	Probability of arrival azimuth, slowness, amplitude, and phase given event
$\Lambda$	Set of arrivals associated to events
$\Lambda^{ijk}$	The arrival of event $i$ 's $j$ th phase at station $k$
$\Lambda_t^{ijk}, \Lambda_z^{ijk}, \Lambda_s^{ijk}, \Lambda_a^{ijk}, \Lambda_h^{ijk}$	Arrival time, azimuth, slowness, amplitude, and phase
$\zeta$	A special symbol that signifies the lack of an arrival
$P_\omega$	Probability of false arrivals
$P_{\omega,n}$	Probability of number of false arrivals
$P_{\omega,t}, P_{\omega,z}, P_{\omega,s}, P_{\omega,a}, P_{\omega,h}$	Probability of false arrival time, azimuth, slowness, amplitude, and phase
$\xi$	All false arrivals. $\xi^k$ are the false arrivals at station $k$ , and $\xi^{kl}$ is one of these false arrivals
$\xi_t^{kl}, \xi_z^{kl}, \xi_s^{kl}, \xi_a^{kl}, \xi_h^{kl}$	False arrival time, azimuth, slowness, amplitude, and phase
$P_\gamma$	Probability distribution of coda arrivals
$P_{\gamma,d}$	Probability that a coda arrival is detected
$P_{\gamma,z}, P_{\gamma,s}, P_{\gamma,a}$	Probability of coda azimuth, slowness, and amplitude given previous arrival
$P_{\gamma,h}$	Probability of coda phase
$\eta^a$	Coda arrival generated by arrival $a$ or $\zeta$ if no coda
$\eta_t^a, \eta_z^a, \eta_s^a, \eta_a^a$	Coda arrival's time, azimuth, slowness, and amplitude
$\eta$	Set of all coda arrivals
$A$	Set of all observed arrivals
$\Delta_{ik}$	Great-circle distance between event $i$ and station $k$

historical data and present an algorithm that infers the set of seismic events given the observed detections. In simple terms, let  $X$  be a random variable ranging over all possible collections of seismic events, with each event defined by time, location, depth, and magnitude. Let  $Y$  range over all possible signal observations at all seismic stations. Then  $P_\theta(X)$  describes a parameterized generative prior over events, and  $P_\phi(Y|X)$  describes how the signal is propagated and measured (including travel time, selective absorption and scattering, noise, artifacts, sensor bias, sensor failures, etc.). Given observations  $Y = y$ , we are interested in the posterior distribution over events,  $P(X|Y = y)$ , which is given by Bayes rule:

$$P(X|Y = y) \propto P_\theta(X)P_\phi(Y = y|X).$$

For the CTBT monitoring application, we are interested in the value  $x^*$  that maximizes this expression, that is, the most likely explanation for all the sensor readings:



**Figure 1.** Example of seismic waveform (station ASAR, channel SE), STA/LTA, and arrivals. At the bottom of the lower panel the dotted lines are the automatically detected arrivals, while the solid lines are the analyst marked arrivals. The color version of this figure is available only in the electronic edition.

$$x^* = \arg \max_x P_\theta(X = x)P_\phi(Y = y|X = x).$$

Assuming that an algorithm can be devised to solve this optimization problem, which turns out to be nontrivial, the key determinant of the success of the Bayesian approach is the accuracy of the probability models  $P_\theta$  and  $P_\phi$ . These models embody, in explicit form, elementary knowledge of seismology and seismometry, as well as the residual uncertainty inherent in that knowledge. Adding to and refining the knowledge embodied in the models reduces residual uncertainty and improves system performance.

Many other researchers have previously applied Bayesian techniques to geophysical problems (see [Duijndam, 1988a,b](#) for examples). Most of these applications involve inference over a fixed set of continuous-valued parameters; the event localization problem in particular is addressed by [Myers et al. \(2007\)](#). The full monitoring problem involves inference over a combinatorial discrete space (the number of events and the association between events and observations), as well as a continuous parameter space for each event. In this respect, it resembles the data association problems arising in multitarget tracking ([Bar-Shalom and Fortmann, 1988](#)).

Our overall project, VISA (vertically integrated seismic analysis), is divided into two stages. The first stage, NET-VISA, is the subject of the current paper. As the name suggests, NET-VISA deals only with network processing and relies upon the IDC's pre-existing signal detection algorithm, which converts the raw waveforms into a sequence of arrivals. Thus, the observations  $Y$  correspond to sets of arrivals with measured attributes rather than raw waveform signals. An example of the signal processing is shown in [Figure 1](#). [Figure 1a](#) displays a filtered seismic waveform, and [Figure 1b](#) shows the typical short-term average to long-term average

ratio (STA/LTA) used to detect the arrivals. The dotted lines in Figure 1b show the automatically detected arrivals, while the solid lines show the analyst marked arrivals. The second stage, SIG-VISA, will incorporate a signal waveform model and thereby subsume the detection function.

As noted previously, NET-VISA computes a single most likely explanation: a set of hypothesized events with their associated arrivals, marking all other arrivals as noise. This input–output specification, while not fully Bayesian in spirit, enables direct comparison with the current automated system bulletin, Standard Event List 3 (SEL3). Using the final expert-generated bulletin, LEB, as ground truth, we compared the two systems on seven days of held-out data. NET-VISA reduces the number of missed events by 60% without introducing more false events. Furthermore, taking data from the more comprehensive NEIC (National Earthquake Information Center) bulletin as ground truth for the continental United States, we find that NET-VISA is able to detect events in the IMS data that are not in the LEB report produced by IDC’s expert analysts. We find similar results by corroborating with other regional networks as well. Thus, NET-VISA’s true performance may be higher than the LEB-based calculation would suggest. Independent IDC evaluation confirms the gist of our findings, as well as demonstrates that NET-VISA can handle large aftershock sequences such as the one following the March 2011 Tohoku earthquake in near real time.

In the rest of this paper we describe the problem and our approach in detail including some elementary seismology (for readers who are not seismologists), the probability model, the inference algorithm, the results, and finally, future plans and conclusions.

### The Seismic Association and Localization Problem

Seismic events are disturbances in the earth’s crust. Our work is concerned primarily with earthquakes and explosions (nuclear and conventional), but other types of events, such as waves breaking, trees falling, or ice falling, may generate seismic waves, too. All such waves occur in a variety of types: body waves that travel through the earth’s interior and surface waves that travel on the surface. There are two types of body waves: compression or  $P$  waves and shear or  $S$  waves. There are also two types of surface waves: Love and Rayleigh. Further, body waves may be reflected off different layers of the earth’s interior; these are labeled distinctly by seismologists. Each particular wave type generated by a given event is called a phase. More than 100 distinct phases are identified in standard tables (Storchak *et al.*, 2003); in our work, we consider 14 phases making up the vast majority of IMS seismic detections:  $P$ ,  $Pn$ ,  $PKP$ ,  $S$ ,  $PKPbc$ ,  $PcP$ ,  $pP$ ,  $Lg$ ,  $PKPab$ ,  $PKKPbc$ ,  $Pg$ ,  $Rg$ ,  $Sn$ , and  $ScP$ .

After traveling through the earth, these waves are picked up in seismic stations as ground vibrations. Typically, seismic stations have either a single three-axis detector or an array of vertical-axis detectors spread over a scale of many

kilometers. Most detectors are sensitive to nanometer-scale displacements and are quite susceptible to noise.

A typical monitoring system runs station-processing software that preprocesses the raw seismometer measurements, filtering out information that is not of interest and computing short-term and long-term averages of the signal amplitude. When the ratio of these averages exceeds a fixed threshold, an arrival (also called a trigger or onset) is announced. (Please refer to Allen, 1978 for an early description of this approach.) Various parameters of the arrival are measured: onset time, azimuth (direction from the station to the source of the wave), slowness (related to the angle of declination of the signal path), amplitude, signal-to-noise ratio (SNR), etc. Based on these parameters, a phase label may be assigned to the arrival based on the standard International Association of Seismology and Physics of the Earth’s Interior (IASPEI) phase catalog (Storchak *et al.*, 2003). It is important to note that none of these measurements are perfect and have uncertainty associated with them.

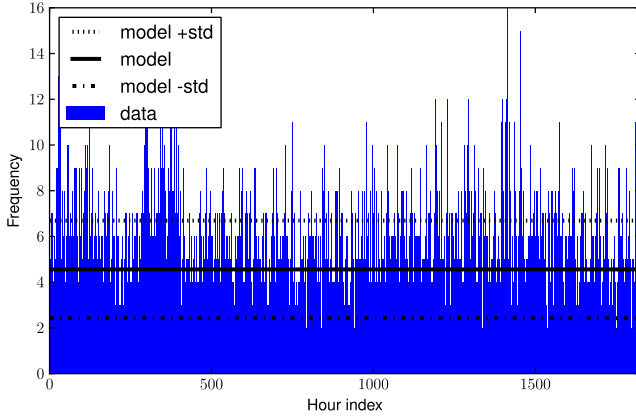
The problem that we attempt to solve in this paper is to take a continuous stream of arrivals (with onset time, azimuth, slowness, amplitude, SNR, and phase label) from the roughly 120 IMS seismic stations as input and produce a bulletin, a continuous stream of events and associations between events and arrivals. The parameters of an event are its longitude, latitude, depth, time, and magnitude ( $m_b$  or body-wave magnitude). Precise details of the metrics for evaluating a bulletin are given in the [Experimental Results](#) section.

### Generative Probabilistic Model

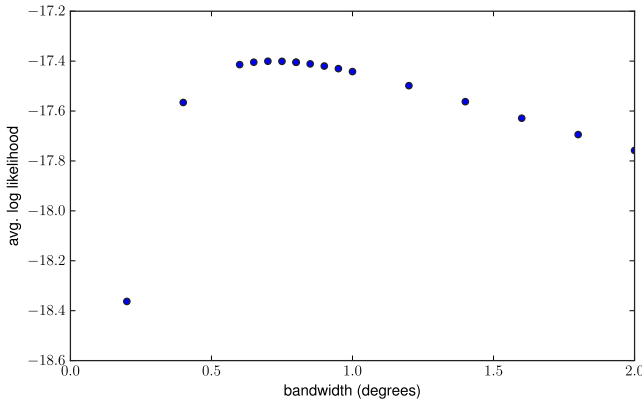
A generative model is one that describes the relationships among all variables of interest (here, events and observations) in such a way that complete samples can be generated from the joint distribution. The model is built from several components, each of which is a conditional probability distribution describing how some aspect of the underlying physical process operates. For example, the conditional distribution that predicts the onset time of an arrival, given the time of the event that caused it, embodies what seismologists call the travel-time model as well as its associated residual uncertainty.

The major components of our model are as follows:

- Events: This component describes the rate of occurrence, geographical distribution, depth distribution, and magnitude distribution of seismic events (natural and man-made).
- True arrivals: Each event is assumed to produce the 14 phases noted earlier; the detection probability model specifies the chance that a given phase from an event will be detected by a given station; the various component models for arrival attributes (onset time, azimuth, slowness, amplitude, and phase label) describe the predicted measurements resulting from station processing for detected phases. Note



**Figure 2.** Estimating event rate. The color version of this figure is available only in the electronic edition.



**Figure 3.** Event location average leave-one-out log likelihood versus bandwidth. The color version of this figure is available only in the electronic edition.

that measurements may include errors; for example, the phase label may not match the actual phase.

- False arrivals: The model describes the rate of false (noise-generated) arrivals at each station and their attribute distributions.
- Coda arrivals: Any arrival may trigger further spurious arrivals detected by the station-processing algorithms in the coda or tail of the waveform, due to additional peaks in the arriving amplitude. Attributes of coda arrivals are usually correlated with those of the triggering arrival.

The following sections describe these components in detail, giving their mathematical forms and methods of parameter estimation.

## Events

In the following, we only consider events with body-wave magnitude  $m_b \geq 2$  or higher. All other events are considered noise, because they are too small to be relevant for treaty monitoring purposes.

*Event Rate and Time.* Events are assumed to be generated by a time-homogeneous Poisson process with a rate parameter,  $\lambda_e$ . If  $e$  is a set of events (of size  $|e|$ ), and  $T$  is the time period under consideration (in seconds), the prior probability of the number of events,  $P_{\theta,n}(\cdot)$ , is given by

$$P_{\theta,n}(|e|) = \frac{(\lambda_e T)^{|e|} \exp(-\lambda_e T)}{|e|!}. \quad (1)$$

For each event,  $e^i$ , the event time,  $e_t^i$ , is uniformly distributed between 0 and  $T$ ; in other words, the prior probability density of the event time,  $P_{\theta,t}(\cdot)$ , is given by

$$P_{\theta,t}(e_t^i) = \frac{1}{T}. \quad (2)$$

The parameter,  $\lambda_e$ , is estimated from the average historical event frequencies as shown in Figure 2. The estimated value of  $\lambda_e$  is 0.001266 per second or 4.6 per hour.

*Event Location.* The longitude and latitude of the  $i$ th event,  $e_l^i$ , are drawn from an event location density,  $P_{\theta,l}(e_l)$ , on the surface of the Earth. This density is a mixture of a uniform density (to allow for explosions and *de novo* seismic activity anywhere on the surface) and a density estimate based on historical event locations. The historical density is modeled by a kernel density obtained by placing a kernel function  $K_{b,x}(\cdot)$  of some fixed width parameter  $b$  at every known event location, so the overall mixture has the form

$$P_{\theta,l}(e_l) = .001 \frac{1}{4\pi R^2} + .999 \frac{1}{H} \sum_{h=1}^H K_{b,g_l^h}(e_l),$$

where  $R$  is the Earth's radius,  $H$  is the number of historical events, and  $g_l^1, \dots, g_l^H$  are the locations of those events. We use an exponential kernel of the form

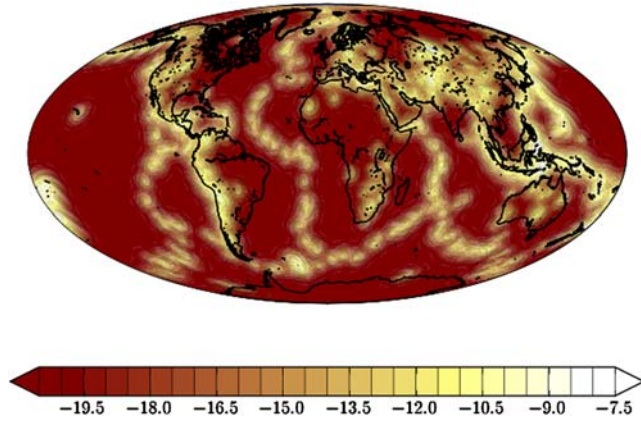
$$K_{b,x}(y) = \frac{1 + 1/b^2}{2\pi R^2} \frac{\exp(-\Delta_{xy}/b)}{1 + \exp(-\pi/b)},$$

where  $\Delta_{xy}$  is the great-circle distance between locations  $x$  and  $y$ , measured in radians. Notice that in the limit as  $b \rightarrow \infty$ , the kernel tends to the uniform density over the Earth's surface.

The optimal value of  $b$  is chosen by leave-one-out cross validation (LOOCV) over a random subset of 1000 event locations in the dataset. We use a grid search to find the value of  $b$  within 0 and 2 that maximizes the LOOCV value. We initially use a grid size of 0.2 and refine the search with a grid size of 0.05. Figure 3 plots the results of this search. The best value of  $b$  that is chosen is 0.7. In Figure 4 we show the event location prior  $\log[P_{\theta,l}(\cdot)]$ , using all of the training dataset and the optimal bandwidth.

The depth of the event,  $e_d^i$ , is uniformly distributed up to a maximum depth  $D$  (700 km in our experiments); in other words, the probability density of the event depth,  $P_{\theta,d}(\cdot)$ , is given by





**Figure 4.** Event location log density. The color version of this figure is available only in the electronic edition.

$$P_{\theta,d}(e_d^i) = \frac{1}{D}. \quad (3)$$

*Event Magnitude.* The Gutenberg–Richter law (Gutenberg and Richter, 1954) posits that the number of events with magnitude  $m$  or more is 10 times the number of events with magnitude  $m + 1$  or more. In terms of the event magnitude probability of an arbitrary event:  $P_{\theta,m}(e_m^i \geq m) = 10 \cdot P_{\theta,m}(e_m^i \geq m + 1)$ . We represent this prior knowledge in our model with an exponential distribution defined for  $e_m^i \geq 2$  with rate  $\lambda_m = \log(10)$ :

$$P_{\theta,m}(e_m^i) = \lambda_m \exp[-\lambda_m(e_m^i - 2)]. \quad (4)$$

Note that because we assume a minimum magnitude of 2, the exponential distribution is shifted forward by this amount. Although man-made events may not follow the Gutenberg–Richter law, their frequency is too low to affect the overall distribution significantly.

*Overall Event Prior.* Under the assumption that the event location, depth, time, and magnitude are independent of each other, the probability of a single event,  $e_i$ , is given by the product of the component terms:

$$P_{\theta}(e^i) = P_{\theta,l}(e_l^i)P_{\theta,d}(e_d^i)P_{\theta,t}(e_t^i)P_{\theta,m}(e_m^i).$$

Substituting from equations (2), (3), and (4), we get

$$P_{\theta}(e^i) = P_{\theta,l}(e_l^i) \frac{1}{D} \frac{1}{T} \lambda_m \exp[-\lambda_m(e_m^i - 2)]. \quad (5)$$

In our model, all the events are exchangeable and are generated independently, thus,

$$P_{\theta}(e) = P_{\theta,n}(|e|) \cdot |e|! \cdot \prod_{i=1}^{|e|} P_{\theta}(e^i).$$

(Note that we are overloading  $P_{\theta}(\cdot)$  to refer to the distribution over the set of events as well as the distribution of a single event.) Substituting from equations (1) and (5),

$$P_{\theta}(e) = \exp(-\lambda_e T) \prod_{i=1}^{|e|} P_{\theta,l}(e_l^i) \frac{1}{D} \lambda_e \lambda_m \exp[-\lambda_m(e_m^i - 2)]. \quad (6)$$

If we define the following single-event quantity, which is independent of the arbitrary interval  $T$ ,

$$\hat{P}_{\theta}(e^i) = P_{\theta,l}(e_l^i) \frac{1}{D} \lambda_e \lambda_m \exp[-\lambda_m(e_m^i - 2)].$$

then we can simplify equation (6) to

$$P_{\theta}(e) = \exp(-\lambda_e T) \prod_{i=1}^{|e|} \hat{P}_{\theta}(e^i). \quad (7)$$

As noted in the [Future Work](#) section, the time-homogeneity and independence assumptions are violated by aftershock phenomena.

#### True Arrivals

An event can generate at most one true arrival of each phase type at a station. Whether or not the arrival is generated depends on the detection probability.

*Detection Probability.* The probability that an event  $i$ 's  $j$ th phase,  $1 \leq j \leq J$ , is detected by a station  $k$ ,  $1 \leq k \leq K$ , depends on the phase, the station, and the event's magnitude, depth, and distance to the station ( $e_m^i$ ,  $e_d^i$ , and  $\Delta_{ik}$ ). As noted in the [Future Work](#) section, such a model ignores source location and path effects as well as anisotropic radiation patterns. Let  $P_{\phi,d}^{jk}(e_i)$  be the probability of this detection. The phase- and station-specific detection distributions,  $P_{\phi,d}^{jk}(\cdot)$ , were obtained using logistic regression models, which are a standard tool for modeling true/false random variables. In such models, a weighted linear expression is formed from the inputs (and possibly additional features computed from the inputs), and the probability that the output is true; in this case, the detection of the phase is given by applying a soft threshold logistic function to the value of the weighted linear expression. The net effect is that the log odds of detection are a linear function of the input features:

$$\log\left(\frac{P_{\phi,d}^{jk}(e^i)}{1 - P_{\phi,d}^{jk}(e^i)}\right) = \sum_{w \in \mathcal{F}_d} \mu_d^{wjk} \cdot w(e_m^i, e_d^i, \Delta_{ik}),$$

where  $\mathcal{F}_d$  is a set of feature functions such that for each  $w \in \mathcal{F}_d$ ,  $w: \mathbb{R}^3 \rightarrow \mathbb{R}$ . Also,  $\mu_d^{wjk}$  is the weight for the feature  $w$ . The complete set of features is defined in [Table 2](#). Because the event magnitude is one of the features, another way of thinking about the previous equation is that a unit increase

Table 2  
Detection Features\*

Feature	Value
(Intercept)	1
mag	$e_m^i$
depth	$e_d^i$
dist	$\Delta_{ik}$
dist0	$\mathcal{N}(\Delta_{ik}, 0, 5)$
dist35	$\mathcal{N}(\Delta_{ik}, 35, 20)$
dist40	$\mathcal{N}(\Delta_{ik}, 40, 20)$
dist12520	$\mathcal{N}(\Delta_{ik}, 125, 20)$
dist12540	$\mathcal{N}(\Delta_{ik}, 125, 40)$
dist 145	$\mathcal{N}(\Delta_{ik}, 145, 10)$
dist170	$\mathcal{N}(\Delta_{ik}, 170, 20)$
dist175	$\mathcal{N}(\Delta_{ik}, 175, 30)$
mag6	$\mathcal{N}(e_m^i, 6, 5.5)$
mag68	$\mathcal{N}(e_m^i, 6, 8)$
md	$(7 - e_m^i) \cdot \Delta_{ik}$

\*List of features used for computing the probability of detecting an arrival from an event  $i$  with magnitude  $e_m^i$ , depth  $e_d^i$ , and distance  $\Delta_{ik}$  from station  $k$ . Here  $\mathcal{N}(x, \mu, \sigma)$  is the standard Gaussian density with mean  $\mu$  and standard deviation  $\sigma$  measured at  $x$ .

in the event magnitude would result in the odds of detection increasing by a multiplicative constant. The exact constant is the exponentiation of the corresponding weight of the magnitude feature, and this is station dependent. In fact, if the weight was negative then the detections odds would decrease; this is, in fact, the case for the travel-time feature.

Directly estimating the feature weights  $\mu_d^{wjk}$  is not always possible because many of the station–phase combinations have very little data. To deal with this data sparsity we used a hierarchical Bayesian procedure (Gelman *et al.*, 2004), which posits that the weight for a station–phase is drawn from a global prior for that phase. This global prior is, in turn, drawn from a weakly informative prior, as follows:

$$\mu_d^{wjk} \sim \mathcal{N}(\mu_d^{wj}, \sigma_d^{wj}) \quad \mu_d^{wj} \sim \mathcal{N}(0, 100)$$

$$(\sigma_d^{wj})^{-2} \sim \Gamma(0.01, 100),$$

where  $\mathcal{N}$  and  $\Gamma$  are the Gaussian and the gamma distributions parameterized by their mean and standard deviation, and shape and scale, respectively. Estimation of parameters follows a coordinate ascent procedure. For each phase  $j$ , we initialize  $\mu_d^{wj} = 0$  and  $\sigma_d^{wj} = 1$ , and then alternately optimize  $\mu_d^{wjk} : \forall w, k$ ,  $\mu_d^{wj} : \forall w$ , and  $\sigma_d^{wj} : \forall w$  until convergence. In each maximization step, the optimal value of  $\mu_d^{wjk}$  is computed by second-order optimization, while the remaining values have a closed-form solution.

For each phase, the previously described procedure ensures that if a station has a lot of data (both detections and nondetections), then its feature weights will be determined almost entirely by its own data, whereas feature weights for data-poor stations tend toward a global average obtained from all stations. Thus, if most data-rich stations

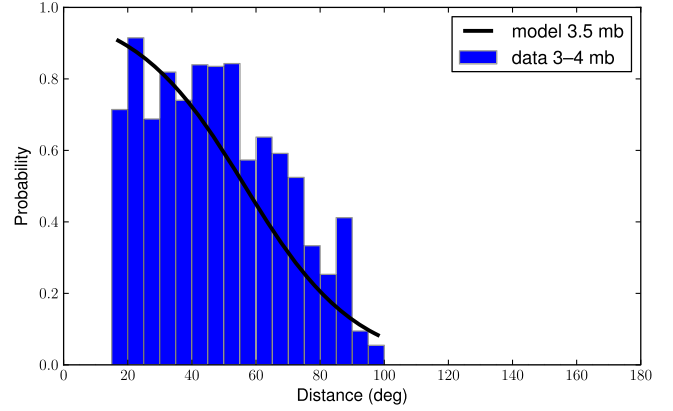


Figure 5. Conditional detection probabilities for the  $P$  phase of surface events between 3 and 4 mb at station ASAR. The color version of this figure is available only in the electronic edition.

have a positive weight for event magnitude for the  $P$  phase, then a station with very little data will also have a positive weight for this feature.

In Figure 5 we show the model prediction that is, the probability of detection, for one phase at a station.

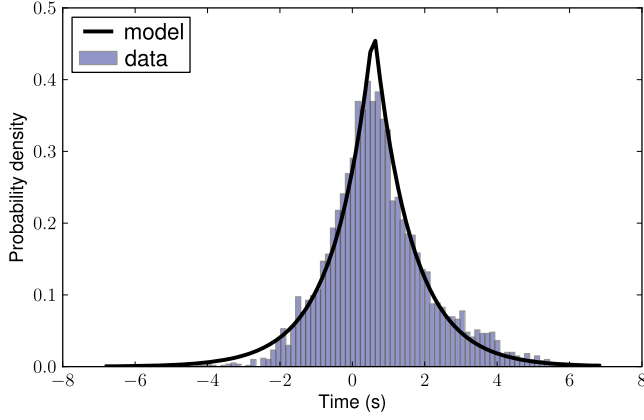
*Arrival Attributes.* If event  $i$ 's  $j$ th phase is detected by a station  $k$ , we define  $\Lambda^{ijk}$  as the corresponding arrival. For notational convenience, we define  $\Lambda^{ijk} = \zeta$  whenever this phase is not detected. Our model specifies a probability distribution for the measured attributes of this arrival: the onset time  $\Lambda_t^{ijk}$ , the azimuth  $\Lambda_z^{ijk}$ , the slowness  $\Lambda_s^{ijk}$ , the amplitude  $\Lambda_a^{ijk}$ , and the assigned phase label  $\Lambda_h^{ijk}$ . The arrival time is

$$\Lambda_t^{ijk} = e_t^i + I_T^j(e_d^i, \Delta_{ik}) + r_t^{ijk},$$

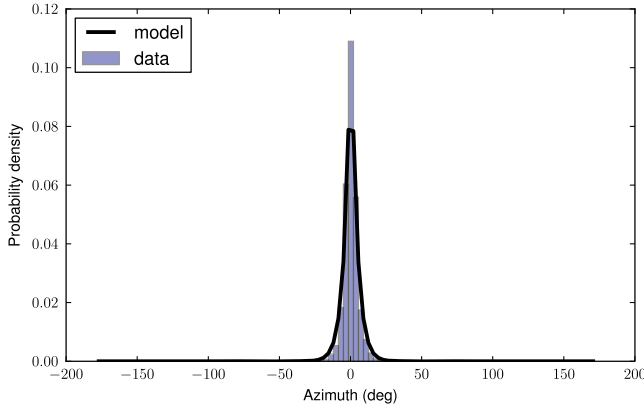
where  $I_T^j$  is the prediction from the IASPEI travel-time model (a function of the event depth and distance to the station), and  $r_t^{ijk}$  is a random residual. The residual distribution accounts for the inhomogeneities in the Earth's crust, which allow seismic waves to travel faster or slower than the IASPEI prediction. This distribution also accounts for any systematic biases in picking seismic onsets from waveforms. Whereas most authors assume Gaussian residuals, an assumption implicit in the use of quadratic cost functions in the GA system, we find experimentally that travel time and indeed most other residuals are distributed according to a Laplacian method. The parameters of a Laplacian distribution are the mean  $\mu_t^{ijk}$  and scale  $b_t^{ijk}$ , and the distribution is given by

$$P_{\phi,t}^{ijk}(\Lambda_t^{ijk} | e^i) = \frac{1}{2b_t^{ijk}} \exp\left(-\frac{|\Lambda_t^{ijk} - e_t^i - I_T^j(e_d^i, \Delta_{ik}) - \mu_t^{ijk}|}{b_t^{ijk}}\right).$$

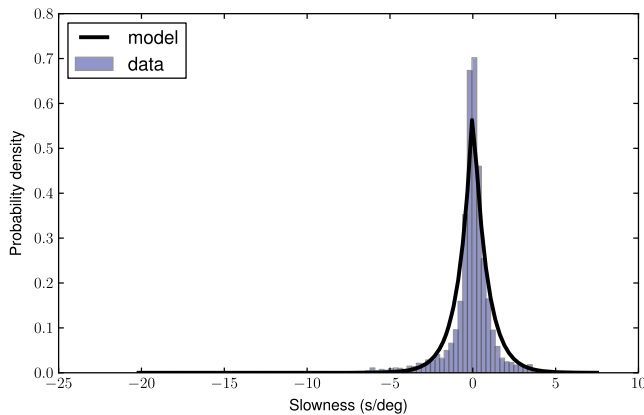
Similarly, the arrival azimuth and slowness follow a Laplacian distribution:



**Figure 6.** Arrival time distribution around the IASPEI prediction for  $P$ -phase arrivals at station ASAR. The color version of this figure is available only in the electronic edition.



**Figure 7.** Arrival azimuth distribution around the geographical prediction (plus SASC correction) for  $P$ -phase arrivals at station ASAR. The color version of this figure is available only in the electronic edition.



**Figure 8.** Arrival slowness distribution around the IASPEI prediction (plus SASC correction) for  $P$ -phase arrivals at station ASAR. The color version of this figure is available only in the electronic edition.

$$P_{\phi,z}^{jk}(\Lambda_z^{ijk}|e^i) = \frac{1}{2b_z^{jk}} \exp\left(-\frac{|\psi[\Lambda_z^{ijk}, G_z(s_l^k, e_l^i)] - \mu_z^{jk}|}{b_z^{jk}}\right),$$

$$P_{\phi,s}^{jk}(\Lambda_s^{ijk}|e^i) = \frac{1}{2b_s^{jk}} \exp\left(-\frac{|\Lambda_s^{ijk} - I_s^j(e_d^i, \Delta_{ik}) - \mu_s^{jk}|}{b_s^{jk}}\right).$$

Here the function  $\psi$  computes the difference in the observed angle  $\Lambda_z^{ijk}$  and the angle computed from the geographical function  $G_z$ , which depends on the station location,  $s_l^k$ , and the event location,  $e_l^i$ . Also,  $I_s^j$  is the slowness value computed from the IASPEI model for phase  $j$ . Technically, the fact that the domain of Laplace distributions is the entire real line makes them inappropriate for azimuth and slowness variables; but the model is reasonable given the small scale of the residuals. It should be noted that the observed values of azimuth and slowness referred to already include certain CTBTO-defined, station-specific corrections called slowness azimuth site corrections (SASC).

The estimation of all the station-and-phase-specific means and scales, for example  $\mu_t^{jk}$  and  $b_t^{jk}$ , is based on a hierarchical model. In this model, each mean and scale is generated from a phase-global prior, which we describe for the arrival time in the following equation:

$$r_t^{ijk} \sim \text{Laplace}(\mu_t^{jk}, b_t^{jk}) \quad \mu_t^{jk} \sim \text{Laplace}(\mu_t^j, b_t^j)$$

$$(b_t^{jk})^{-1} \sim \Gamma(1, \beta_t^j) \quad \mu_t^j \sim \text{Laplace}(0, 100)$$

$$(b_t^j)^{-1} \sim \Gamma(.01, 100) \quad (\beta_t^j)^{-1} \sim \Gamma(.01, 100).$$

For each phase  $j$ , we start by initializing  $\mu_t^{jk} = 0$ ,  $b_t^{jk} = 1$  for all stations  $k$ ,  $\mu_t^j = 0$ ,  $b_t^j = 1$ , and  $\beta_t^j = 1$ . Next, we iteratively optimize the values of  $\mu_t^{jk}$ ,  $b_t^{jk}$  for each station, and the global values  $\mu_t^j$ ,  $b_t^j$ , and  $\beta_t^j$ . This is repeated until convergence. Each of these optimization steps has a simple closed-form solution. A similar procedure is adopted for estimating the azimuth parameters,  $\mu_z^{jk}$  and  $b_z^{jk}$ , and the slowness parameters,  $\mu_s^{jk}$ , and  $b_s^{jk}$ . Examples for these three types of distributions are shown in Figures 6, 7, and 8.

The effect of the hierarchical model is similar in spirit to the model for detection probabilities. For each phase, stations with a lot of data will compute the mean and scale almost exclusively from their own data, whereas parameters for data-poor stations will tend to reflect the global average.

The arrival amplitude  $\Lambda_a^{ijk}$  is similar to the detection probability because it depends only on the event magnitude, depth, and distance to the station. We model the log of the amplitude via a linear regression model with Gaussian noise,

$$r_a^{ijk} = \log(\Lambda_a^{ijk}) - \sum_{w \in \mathcal{F}_a} \mu_a^{wjk} \cdot w[e_m^i, e_d^i, I_T^j(e_d^i, \Delta_{ik})],$$

$$r_a^{ijk} \sim \mathcal{N}(0, \sigma_a^{jk}),$$

where  $\mathcal{F}_a$  is a set of feature functions (see Table 3), and  $\mu_a^{wjk}$  is the weight for feature  $w$ . This implies that

$$P_{\phi,a}^{jk}(\Lambda_a^{ijk}|e^i) = \frac{1}{\sqrt{2\pi\sigma_a^{jk}}} \exp\left[-\frac{(r_a^{ijk})^2}{2\sigma_a^{jk2}}\right] \frac{1}{\Lambda_a^{ijk}}.$$

Table 3  
Amplitude Features\*

Feature	Value
(Intercept)	1
mag	$e_m^i$
depth	$e_d^i$
$t$ time	$I_T^j(e_d^i, \Delta_{ik})$
$t$ time0	$\mathcal{N}[I_T^j(e_d^i, \Delta_{ik}), 0, 50]$

\*List of features used for predicting the amplitude of a phase,  $j$ , arrival from an event,  $i$ , at a station,  $k$ .

In order to estimate the feature weights, we use a hierarchical model, as before, which assumes that for each phase the feature weights at a station are drawn from a global prior:

$$\begin{aligned} \mu_a^{wjk} &\sim \mathcal{N}(\mu_a^{wj}, \sigma_a^{wj}) & (\sigma_a^{jk})^{-2} &\sim \Gamma(100, \beta_a^j) \\ \mu_a^{wj} &\sim \mathcal{N}(0, 100) & (\sigma_a^{wj})^{-2} &\sim \Gamma(.01, 100) \\ (\beta_a^j)^{-1} &\sim \Gamma(.01, 100). \end{aligned}$$

Maximum *a posteriori* (MAP) inference of these weights is along the lines of the method already described for other parameters. Figure 9 has an example of a learned model.

Finally, the distribution for the assigned phase label  $\Lambda_h^{ijk}$  reflects the phase classification errors made by station processing. We model this with a categorical distribution whose parameters depends only on the true phase,  $j$ :

$$P_{\phi,h}^{jk}(\Lambda_h^{ijk} | e^i) = P_h^j(a_h^{ijk}).$$

Learning the categorical distributions  $P_h^j$  is a simple matter of counting with add-one smoothing. The learned distribution is plotted as a heat map in Figure 10. The results indicate that many phases are falsely classified as  $P$  by the station-processing software.

*Overall True Arrivals.* We assume that all the measured attributes of an arrival are conditionally independent of each other, given the true event. Thus,

$$\begin{aligned} P_{\phi}^{jk}(\Lambda^{ijk} | e^i) &= P_{\phi,t}^{jk}(\Lambda_r^{ijk} | e^i) P_{\phi,z}^{jk}(\Lambda_z^{ijk} | e^i) P_{\phi,s}^{jk}(\Lambda_s^{ijk} | e^i) \\ &\times P_{\phi,a}^{jk}(\Lambda_a^{ijk} | e^i) P_{\phi,h}^{jk}(\Lambda_h^{ijk} | e^i). \end{aligned}$$

Further, assuming that the arrivals are all conditionally independent, and keeping in mind that the observed arrivals  $\Lambda$  include by implication all the nondetections for every missing station–phase combination, we have

$$\begin{aligned} P_{\phi}(\Lambda | e) &= \prod_{i=1}^{|e|} \prod_{j=1}^J \prod_{k=1}^K [\mathbb{1}(\Lambda^{ijk} = \zeta) [1 - P_{\phi,d}^{jk}(e^i)] \\ &+ \mathbb{1}(\Lambda^{ijk} \\ &\neq \zeta) P_{\phi,d}^{jk}(e^i) P_{\phi}^{jk}(\Lambda^{ijk} | e^i)]. \end{aligned} \quad (8)$$

The term inside the product is either a nondetection probability or a detection probability times the arrival probability.

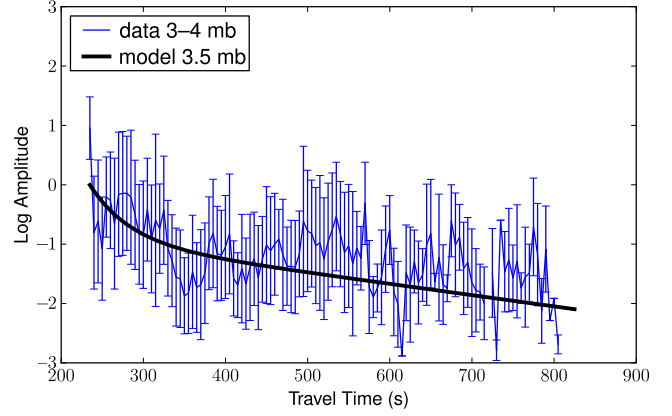


Figure 9. Arrival log amplitude distribution for the  $P$  phase of a surface event of 3–4 mb at station ASAR. The color version of this figure is available only in the electronic edition.

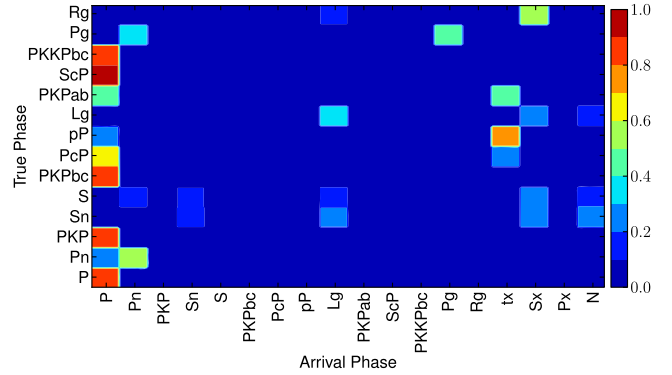


Figure 10. Arrival phase probability as function of the true phase. The color version of this figure is available only in the electronic edition.

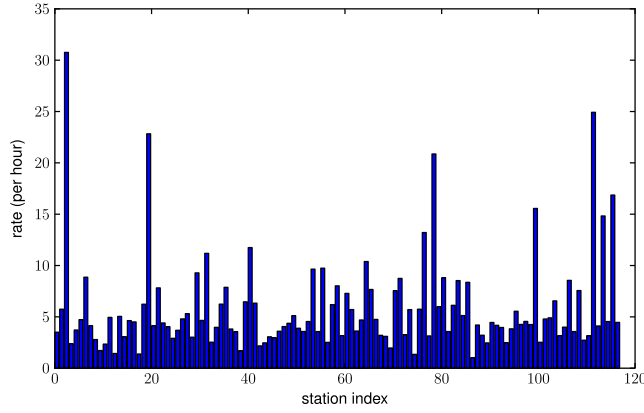
The indicator variable on the condition  $\Lambda^{ijk} = \zeta$  selects the appropriate term.

These conditional independence assumptions are important in reducing the number of model parameters, but they are certainly open to question. Considering arrivals at two different stations, the assumption implies that the residuals are uncorrelated; this is reasonable because IMS stations are widely separated. For arrivals of two different phases at the same station, the assumption of uncorrelated residuals is more problematic. Our analysis of travel-time residuals for  $P$  and  $S$  phases does show some correlation, but it is weak enough that ignoring it does not seriously compromise performance.

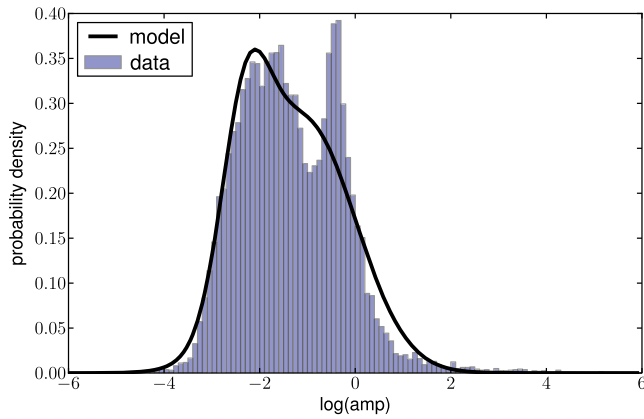
Each station  $k$  also generates a set  $\xi^k$  of false arrivals, that is, arrivals not associated with any event phase. These are described by an overall distribution  $P_{\omega}^k$ . The number of false arrivals is distributed according to a time-homogeneous Poisson process with rate  $\lambda_f^k$ :

$$P_{\omega,n}^k(|\xi^k|) = \frac{(\lambda_f^k T)^{|\xi^k|} \exp(-\lambda_f^k T)}{|\xi^k|!}.$$

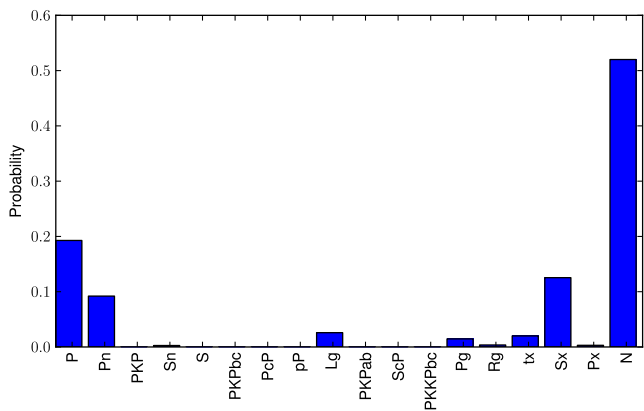




**Figure 11.** The average false arrival rate per hour at all the stations. The color version of this figure is available only in the electronic edition.



**Figure 12.** Amplitude distribution for false arrivals at station ASAR. The color version of this figure is available only in the electronic edition.



**Figure 13.** Phase distribution for all false arrivals. The color version of this figure is available only in the electronic edition.

The values of  $\lambda_f^k$  are displayed in Figure 11. If  $\xi^{kl}$  is one of this set of false arrivals, its time  $\xi_t^{kl}$ , azimuth  $\xi_z^{kl}$ , and slowness  $\xi_s^{kl}$  are generated uniformly over their respective ranges:

$$P_{\omega,t}^k(\xi_t^{kl}) = \frac{1}{T}, \quad P_{\omega,z}^k(\xi_z^{kl}) = \frac{1}{M_z}, \quad P_{\omega,s}^k(\xi_s^{kl}) = \frac{1}{M_s},$$

where  $M_z$  and  $M_s$  are the range of values for azimuth and slowness, respectively. The log amplitude  $\log(\xi_a^{kl})$  of the false arrival is generated from either a uniform distribution with probability 0.1 or a mixture of two Gaussians, which is estimated from the data with a standard expectation maximization procedure. This distribution,  $f_{\omega,a}^k(\cdot)$ , at one station is displayed in Figure 12.

### False Arrivals

The resulting distribution on the false arrival amplitude is given by

$$P_{\omega,a}^k(\xi_a^{kl}) = f_{\omega,a}^k[\log(\xi_a^{kl})] \frac{1}{\xi_a^{kl}}.$$

Finally, the phase label  $\xi_h^{kl}$  assigned to the false arrival follows a categorical distribution,  $P_{\omega,h}^k(\xi_h^{kl})$ . This is estimated via the posterior mean under a uniform Dirichlet prior (see Fig. 13).

Overall, assuming the false arrival attributes are independently generated, we have

$$P_{\omega}^k(\xi^{kl}) = P_{\omega,t}^k(\xi_t^{kl}) P_{\omega,z}^k(\xi_z^{kl}) P_{\omega,s}^k(\xi_s^{kl}) P_{\omega,a}^k(\xi_a^{kl}) P_{\omega,h}^k(\xi_h^{kl}).$$

Because the false arrivals at a station are exchangeable, the probability for the set  $\xi^k$  is

$$\begin{aligned} P_{\omega}^k(\xi^k) &= P_{\omega,n}^k(|\xi^k|) \cdot |\xi^k|! \prod_{l=1}^{|\xi^k|} P_{\omega}^k(\xi^{kl}) \\ &= \exp(-\lambda_f^k T) \prod_{l=1}^{|\xi^k|} \frac{\lambda_f^k}{M_z M_s} P_{\omega,a}^k(\xi_a^{kl}) P_{\omega,h}^k(\xi_h^{kl}). \end{aligned}$$

As before, we have overloaded  $P_{\omega}^k$  to refer to a distribution over a set of arrivals as well as a single arrival. As before, we can simplify the equations by defining

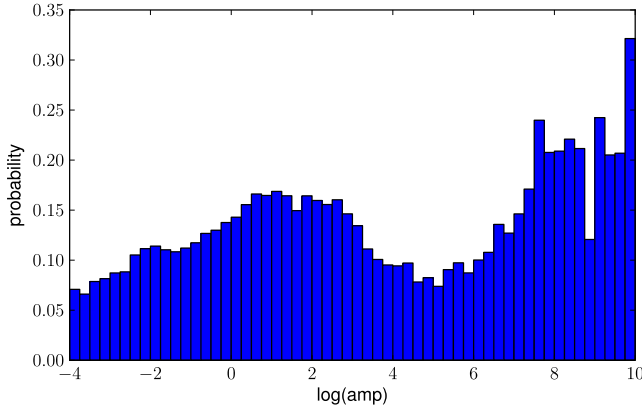
$$\hat{P}_{\omega}^k(\xi^{kl}) = \frac{\lambda_f^k}{M_z M_s} P_{\omega,a}^k(\xi_a^{kl}) P_{\omega,h}^k(\xi_h^{kl}).$$

Now, assuming that the false arrivals at different stations are independent of each other, the probability for the complete set  $\xi$  of all false arrivals at all stations is

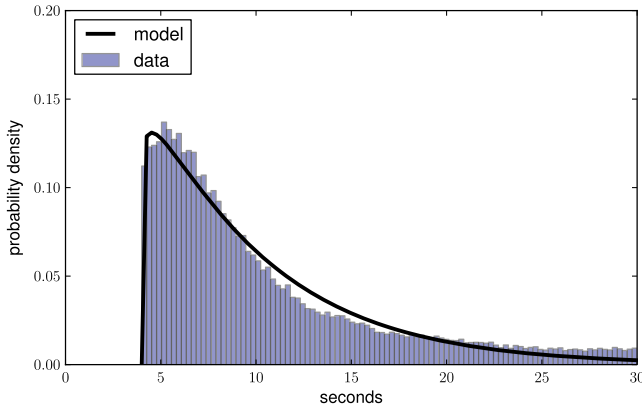
$$P_{\omega}(\xi) = \exp\left(-\sum_{k=1}^K \lambda_f^k T\right) \prod_{k=1}^K \prod_{l=1}^{|\xi^k|} \hat{P}_{\omega}^k(\xi^{kl}). \quad (9)$$

### Coda Arrivals

Following the initial onset and peak, the arriving energy from a seismic phase does not decay smoothly, but exhibits many subsidiary peaks, some of which may even be larger than the initial peak. These subsidiary peaks may fool the STA/LTA detector into announcing additional arrivals, called coda arrivals; in the IMS data, up to half of all arrivals are of



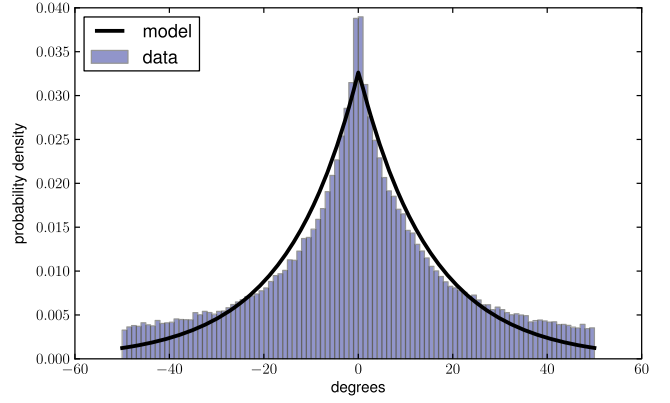
**Figure 14.** Coda detection probability as a function of the log amplitude of the triggering arrival. The color version of this figure is available only in the electronic edition.



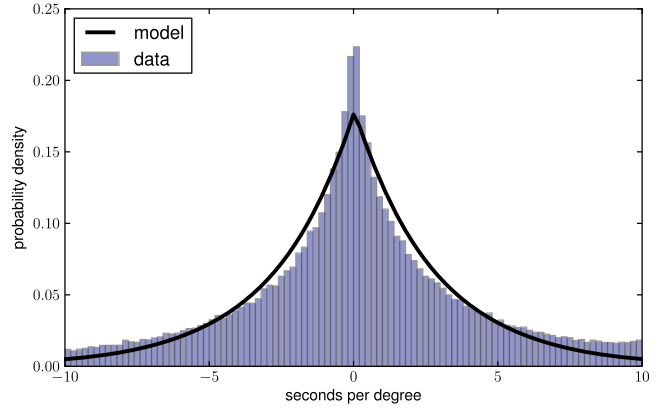
**Figure 15.** Time delay for coda arrival after the triggering arrival. The color version of this figure is available only in the electronic edition.

this type. An example of coda arrivals can be seen in Figure 1 around 20 s after the main arrival, which is at offset 100 s. One might imagine that coda arrivals can be lumped in with false arrivals, but it turns out that the attributes of coda arrivals are strongly correlated with those of the triggering arrival. If the coda arrivals are not modeled explicitly, then our inference will end up hypothesizing additional spurious events as the most likely explanation for many of the coda arrivals.

We model the parameters of the coda arrival with a distribution  $P_\gamma$ , which is a function of the parameters of the triggering arrival. Whether the triggering arrival was a false arrival, or caused by an event, or itself triggered by another arrival, is immaterial. We define  $\eta^a$  as the arrival triggered by arrival  $a$  (at the same station), or  $\zeta$  if there is no such triggered arrival. Now, the probability that an arrival  $a$  triggers a coda arrival is given by  $P_{\gamma,d}(a)$ , which is a function of the amplitude of the arrival  $a$ . We estimate  $P_{\gamma,d}$  with a nonparametric model by discretizing the log amplitude of the triggering arrival into buckets of size 0.25 between  $-4$  and  $10$ . This distribution is displayed in Figure 14. Any points outside these extreme values are mapped to the nearest bucket.



**Figure 16.** Coda azimuth difference versus triggering arrival. The color version of this figure is available only in the electronic edition.



**Figure 17.** Coda slowness difference versus triggering arrival. The color version of this figure is available only in the electronic edition.

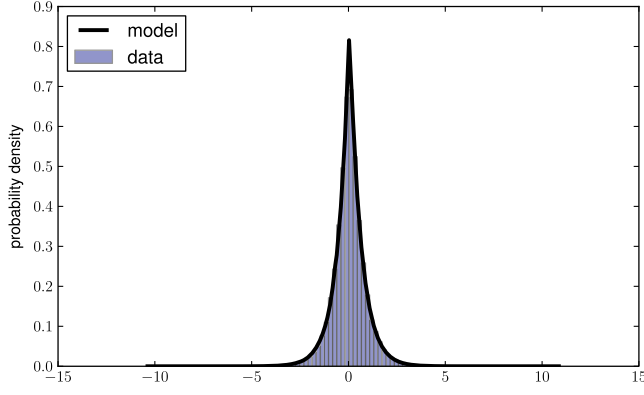
One problem that arises while training the model for coda arrivals is that the IDC analysts do not annotate coda arrivals in the LEB bulletin, which makes it hard to estimate the parameters of  $P_\gamma$ . Our solution is to heuristically annotate some of the unassociated arrivals as coda arrivals and use this annotation to learn  $P_\gamma$  and also  $P_\omega$ . Our procedure searches the training data for any unassociated arrivals within 30 s of a prior arrival at the same station and with an azimuth and slowness within  $50^\circ$  and 10 s per degree, respectively, of the prior arrival's values, and to mark such arrivals as coda.

We model the distribution of the time delay between the coda arrival and the triggering arrival as a gamma distribution (see Fig. 15):

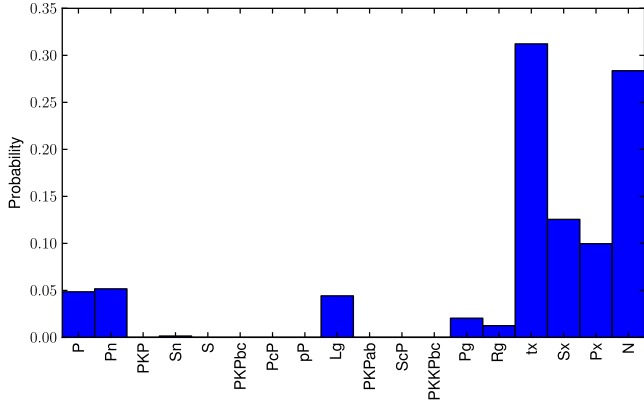
$$\eta_t^a - a_t \sim \Gamma(\rho_t, \nu_t), \quad \text{i.e.}$$

$$P_{\gamma,t}(\eta_t^a | a) = \frac{1}{\Gamma(\rho_t) \nu_t^{\rho_t}} (\eta_t^a - a_t)^{\rho_t - 1} \exp\left(-\frac{\eta_t^a - a_t}{\nu_t}\right).$$

The differences in azimuth, slowness, and log amplitude of the coda versus the triggering arrival are all modeled as Laplace distributions (Figs. 16, 17, and 18):



**Figure 18.** Coda log amplitude difference versus triggering arrival. The color version of this figure is available only in the electronic edition.



**Figure 19.** Coda phase. The color version of this figure is available only in the electronic edition.

$$P_{\gamma,z}(\eta_z^a|a) = \frac{1}{2\nu_z} \exp\left(-\frac{|\psi(\eta_z^a, a_z) - \rho_z|}{\nu_z}\right)$$

$$P_{\gamma,s}(\eta_s^a|a) = \frac{1}{2\nu_s} \exp\left(-\frac{|\eta_s^a - a_s - \rho_s|}{\nu_s}\right)$$

$$P_{\gamma,a}(\eta_a^a|a) = \frac{1}{2\nu_a} \exp\left(-\frac{|\log(\eta_a^a) - \log(a_a) - \rho_a|}{\nu_a}\right) \frac{1}{\eta_a^a}.$$

Finally, the coda arrival's assigned phase label is a categorical distribution,  $P_{\gamma,h}$ , that does not depend on the previous arrival's phase (see Fig. 19). The phase of the coda arrival is uncorrelated with the phase of the triggering arrival because these are mostly labeled using a slightly different algorithm by station processing. Finally, it is worth noting that the coda model is not station specific, hence data sparsity is not a concern.

Overall, assuming that all the parameters of the coda arrival are independent,

$$P_{\gamma}(\eta^a|a) = P_{\gamma,i}(\eta_i^a|a_i)P_{\gamma,z}(\eta_z^a|a_z)P_{\gamma,s}(\eta_s^a|a_s) \\ \times P_{\gamma,a}(\eta_a^a|a_a)P_{\gamma,h}(\eta_h^a).$$

We will denote by  $\eta$  the generative process of all the coda arrivals.  $\eta$  encapsulates the relationship between each arrival and its coda arrival (if any), as well as the set of coda arrivals. If  $\Lambda$  is the set of true arrivals, and  $\xi$  is the set of false arrivals, then the complete set of arrivals  $A$  is given by

$$A = \Lambda \cup \xi \cup \eta.$$

Note that some of the coda arrivals are triggered by coda arrivals and others by noncoda arrivals. However, for each arrival a decision is made whether or not to generate a coda. Assuming that the coda arrivals are independent of each other,

$$P_{\gamma}(\eta|\Lambda \cup \xi) = \prod_{a \in A \wedge \eta^a \neq \zeta} P_{\gamma,d}(a)P_{\gamma}(\eta^a|a) \\ \times \prod_{a \in A \wedge \eta^a = \zeta} [1 - P_{\gamma,d}(a)]. \quad (10)$$

The first product term in equation (10) accounts for the generation of all the coda arrivals, while the second term accounts for all arrivals which do not generate any coda.

### Summary of Model

Combining the model components developed in the preceding sections, the overall probability of any hypothesized sequence of events  $e$ , true arrivals  $\Lambda$ , false arrivals  $\xi$ , coda arrivals  $\eta$ , and the complete set of arrivals  $A$ , where  $A^k$  is the set of arrivals at station  $k$ , is

$$P(e, \Lambda, \xi, \eta, A) = P_{\theta}(e)P_{\phi}(\Lambda|e)P_{\omega}(\xi)P(\eta|\Lambda \cup \xi)\mathbb{1}(A \\ = \Lambda \oplus \xi \oplus \eta). \quad (11)$$

Here the last term is required to ensure that any arrival is in exactly one of the three sets  $\Lambda$ ,  $\xi$ , and  $\eta$ . Other obvious consistency requirements, such as requiring  $\xi^k \subseteq A^k$ , are left out for brevity.

### Inference

As noted in the [Introduction](#), the goal of NET-VISA's inference algorithm is to find the most likely explanation consistent with the observations,

$$\arg \max_{e, \Lambda, \xi, \eta} P(e, \Lambda, \xi, \eta|A) = \arg \max_{e, \Lambda, \xi, \eta} P(e, \Lambda, \xi, \eta, A),$$

where  $P(e, \Lambda, \xi, \eta, A)$  is given by equation (11). Because arrivals from real seismic sensors are observed incrementally and roughly in time-ascending order, our inference algorithm also produces an incremental hypothesis that advances with time. Our algorithm operates by a series of local moves, modifications to the current hypothesis, that (with some exceptions) improve its probability score. It can be seen as a form of optimization by gradient ascent, also known as greedy local search (Cormen *et al.*, 2009). Further we will assume that the coda arrivals form a contiguous chain, or in other words  $\eta^a$  can be either  $\zeta$  or the immediately following arrival.

Let  $M_T$  denote the maximum travel time for any phase. Initially, we start with an event window of size  $W$  from  $t_0 = 0$  to  $t_1 = W$ , and an arrival window of size  $W + M_T$  from  $t_0 = 0$  to  $t_2 = W + M_T$ . Then we perform a series of local moves that add or update events in the event window, delete existing events, or classify (as true arrival, false arrival, or coda arrival) the arrivals in the arrival window. Next, the windows are moved forward by a step size,  $S$ . At this point, events older than  $t_0 - M_T$  become stable: none of the moves modify either the events or arrivals associated with them. These events are then output. While in theory this algorithm never needs to terminate, our experiments continue until the test dataset is fully consumed.

The algorithm's initial hypothesis is that all new arrivals added to the arrival window are false arrivals. This is refined by classifying any arrival  $a$  (at station  $k$ ), with the immediately preceding arrival  $a^-$ , as a coda arrival if

$$P_{\gamma,d}(a^-)P_{\gamma}(a|a^-) > [1 - P_{\gamma,d}(a^-)]P_{\omega}^k(a).$$

In simple terms, the condition expressed in the previous equation states that it is more likely that the arrival  $a^-$  generated a coda and this coda was  $a$  than that  $a^-$  did not generate a coda and  $a$  was a false arrival. This default classification for an arrival is retained whenever it is no longer associated with an event. For convenience we define

$$Y^k(a) = \max\{P_{\gamma,d}(a^-)P_{\gamma}(a|a^-), [1 - P_{\gamma,d}(a^-)]P_{\omega}^k(a)\}. \quad (12)$$

Next, the birth move generates new events in the event window. These events are added to the hypothesis with  $\Lambda^{ijk} = \zeta$  for each new event  $i$ . Subsequently, we repeat the following  $N$  times: one improve-arrival move for each arrival in the arrival window, and one improve-event move for each event in the event window. Finally, the death move kills some of the events, and we repeat one round of improve-arrival and improve-event moves. We describe these steps algorithmically next. The individual moves will be described in more detail later.

1.  $t_0 = 0$ ;  $t_1 = W$ ;  $t_2 = W + M_T$ .
2. Repeat while  $t_0 < \text{max time}$ .
  - a. Give a default classification to arrivals in  $t_0$  to  $t_2$ .
  - b. Add events from birth move ( $t_0, t_1, \{a: t_0 \leq a_t \leq t_2\}$ ).
  - c. Repeat  $N$  times.
    - i. For each arrival  $a$ , such that  $t_0 \leq a_t \leq t_2$ , improve-arrival ( $a$ ).
    - ii. For each event  $e^i$ , such that  $t_0 \leq e_t^i \leq t_1$ , improve-event ( $e^i$ ).
  - d. For all events  $e^i$ , death move ( $e^i$ ).
  - e. For each arrival  $a$ , such that  $t_0 \leq a_t \leq t_2$ , improve-arrival ( $a$ ).
  - f. For each event  $e^i$ , such that  $t_0 \leq e_t^i \leq t_1$ , improve-event ( $e^i$ ).
  - g.  $t_0 + = S, t_1 + = S, t_2 + = S$ .
  - h. Output  $e^i, \Lambda^{ijk}$  for all  $e^i$  such that  $e_t^i < t_0 - M_T$ .
3. Output any remaining  $e^i$ .

In order to simplify the computations needed to compare alternate hypotheses, we decompose the overall probability of equation (11) into the contribution from each event. We define the score  $S_e$  of an event as the probability ratio of two hypotheses: one in which the event exists, and another in which the event does not exist and all of its associated arrivals have the default classification (false or coda). If an event has score less than 1, an alternative hypothesis in which the event is deleted clearly has higher probability. Critically, this event score is unaffected by other events in the current hypothesis. From equations (7), (8), (9), (10), and (12) we have

$$S_e(e^i) = P_{\theta}(e^i) \prod_{j=1}^J \prod_{k=1}^K \left[ \mathbb{1}(\Lambda^{ijk} = \zeta) [1 - P_{\phi,d}^{jk}(e^i)] + \frac{\mathbb{1}(\Lambda^{ijk} \neq \zeta) P_{\phi,d}^{jk}(e^i) P_{\phi}^{jk}(\Lambda^{ijk}|e^i)}{Y^k(\Lambda^{ijk})} \right].$$

Note that the final fraction in the previous equation is a likelihood ratio comparing interpretations of the same arrival as either the arrival of event  $i$ 's  $j$ th phase at station  $k$ , or as a false arrival or a coda arrival. We can further decompose the score into scores  $S_d$  for each arrival. The score of  $\Lambda^{ijk}$ , defined when  $\Lambda^{ijk} \neq \zeta$ , is the ratio of the probabilities of the hypothesis, where the arrival is associated with phase  $j$  of event  $i$  at station  $k$  versus the default classification.

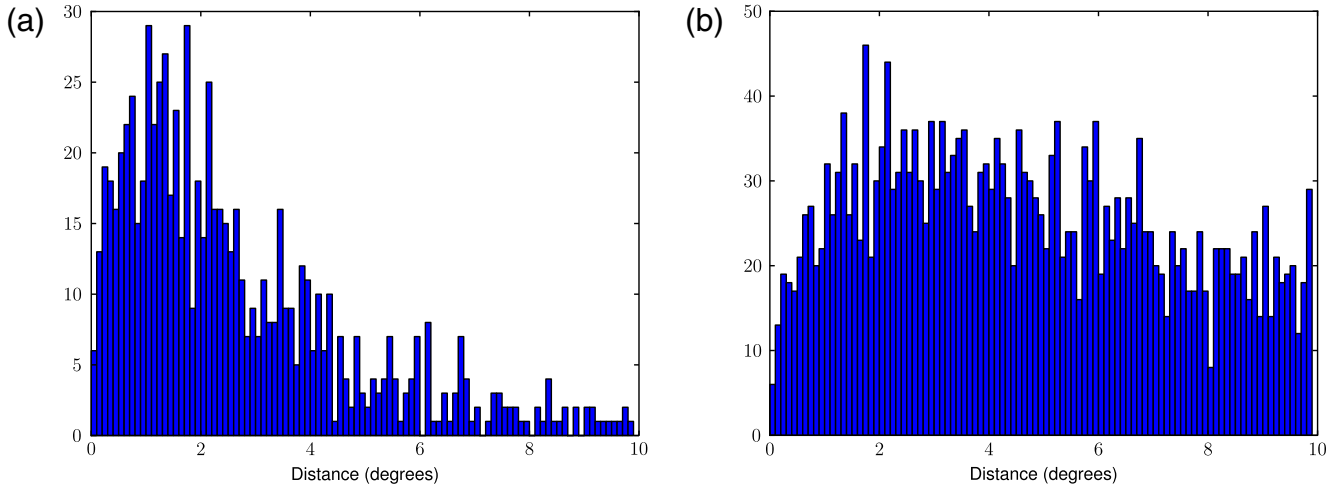
$$S_d^{jk}(\Lambda^{ijk}|e^i) = \frac{P_{\phi,d}^{jk}(e^i) P_{\phi}^{jk}(\Lambda^{ijk}|e^i)}{1 - P_{\phi,d}^{jk}(e^i) Y^k(\Lambda^{ijk})}.$$

By definition, any arrival with a score less than 1 is more likely to be a false or coda arrival. Also, the score of an individual arrival is independent of other arrivals and events in the hypothesis. These scores play a key role in the following local search moves.

### Birth Move

The birth move proposes events within a given time range, based on a list of arrivals. It starts off by inverting each of these arrivals to obtain an initial candidate list of event locations and times. The ability to invert an event follows from the fact that the slowness of an arrival is a monotonic function of distance (with fixed depth). If one assumes that an arrival is the  $P$  phase of a surface event, one can obtain a distance estimate from the slowness, which, combined with the arrival azimuth and time, gives an estimate for an event location and time. In Figure 20 we show the statistics of the distance between the inverted locations obtained from all arrivals in a one-week period and the corresponding ground-truth events during the same time period. The statistics in the figure show that for the great majority of events there is some arrival with an inverted location within  $5^\circ$ .

Having proposed a location and time by inverting one arrival, the birth move next attempts to construct the best possible event in a  $5^\circ$ -ball around the candidate location in steps of  $2.5^\circ$  using all available detections. Event depth



**Figure 20.** Distance between events and inverted locations within  $10^\circ$  and 100 s. (a) The figure shows the distance from a true event and the nearest inverted location. (b) The figure shows the converse, that is, the distance from an inverted location to the nearest true event. The color version of this figure is available only in the electronic edition.

is fixed to the surface, and two possible magnitude values are used (3 and 4). The event time is computed from the IASPEI model using the current event location and the arrival time. The best such event is further optimized using the improve-arrival and improve-event moves and then set aside. This process is repeated as long as the best event has a score greater than 1. An event is not allowed to use arrivals associated with events found earlier in this process. Finally, all the events generated in this process are returned (without their associated arrivals). In algorithmic form, the process is as follows:

1. Given  $t_0$ ,  $t_1$ , and arrivals  $A$ .
2. Repeat for each  $a$  in  $A$ .
  - a. Invert  $a$  to obtain a candidate location  $\iota_a$ .
  - b. Repeat for each location  $e$  in a ball around  $\iota_a$ .
    - i. Initialize  $\Lambda^e$ .
    - ii. Repeat for each arrival  $a$  in  $A$  (let  $k$  be  $a$ 's station).
      - A. Determine the phase  $j$  with the maximum score  $S_d^{jk}(a|e)$ .
      - B. If  $S_d^{jk}(a|e) > S_d^{jk}(\Lambda^{jk}|e)$  or if  $\Lambda^{jk} = \zeta$  and  $S_d^{jk}(a|e) > 1$ , then set  $\Lambda^{jk} = a$ .
3. Let  $e$  be the event with the maximum score  $S_e(e)$  in step 2.
4. Repeat 100 times.
  - a. Invoke improve-event ( $e$ ).
  - b. Invoke improve-arrival (a) for all  $a$  in  $A$  with  $e$  as the only potential event.
5. If  $S_e(e) > 1$ , then set aside event  $e$ , and remove arrivals in  $\Lambda^{e, jk}$  from  $A$ , then go to step 2.
6. Return set-aside events.

Note that the design of the birth move allows for easy parallelization using threads. On a machine with multiple CPU cores, a simple variant of the previously described birth

move is employed. In step 2, the arrivals are divided equally among the threads. Each thread uses its arrivals to compute its candidate events that are then evaluated against all the arrivals to compute the best event. Finally, in step 3 the overall best event from among all the threads is computed in serial.

#### Improve-Arrival Move

For each arrival in the arrival window, we consider all possible phases  $j$  of all events  $i$  up to  $M_T$  seconds earlier. We then associate the best event–phase pair for this arrival that is not already assigned to an arrival with higher score at the same station  $k$ . If this best event–phase pair has score  $S_d^{jk}(\Lambda^{ijk}|e^i) < 1$ , the arrival is changed to its default status (one of false or coda). In more precise terms:

1. Given arrival  $a$  at station  $k$ .
2. Repeat for each event  $e$ .
  - a. Determine the phase  $j$  with the maximum score  $S_d^{jk}(\Lambda^{e, jk}|e)$ .
3. Let  $e$  be the event with the maximum score  $S_d^{jk}(\Lambda^{e, jk}|e)$ .
4. If  $S_d^{jk}(a|e) > S_d^{jk}(\Lambda^{e, jk}|e)$  or if  $\Lambda^{e, jk} = \zeta$  and  $S_d^{jk}(a|e) > 1$ , then set  $\Lambda^{e, jk} = a$ .

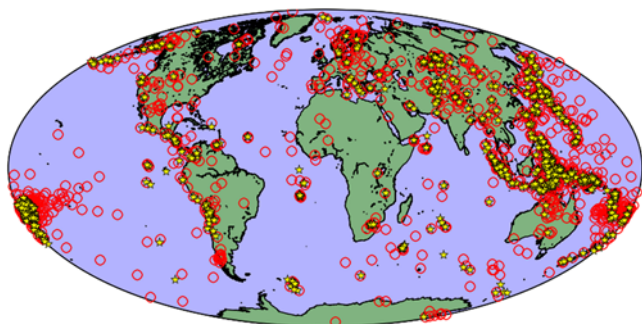
#### Improve-Event Move

For each event  $e^i$ , we consider 100 points chosen uniformly at random in a small ball around the event ( $2^\circ$  in longitude and latitude, 100 km in depth, 5 s in time, and 2 units of magnitude), and choose those attributes with the highest score  $S_e(e^i)$ .

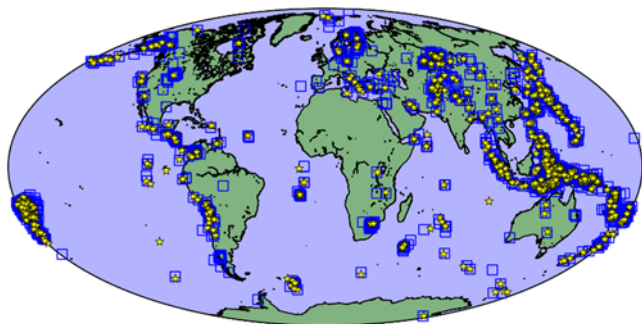
#### Death Move

Any event  $e^i$  with score  $S_e(e^i) < 1$  is deleted, and all of its currently associated arrivals are marked as false alarms.

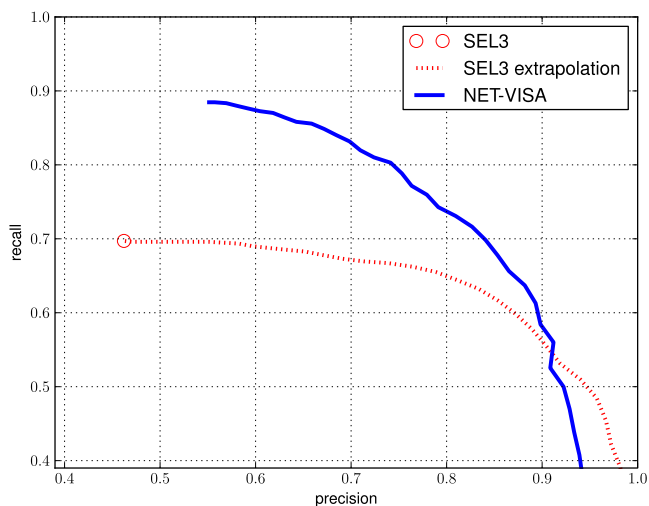




**Figure 21.** Distribution of all the events predicted by the SEL3 (circles) and the LEB (stars). The color version of this figure is available only in the electronic edition.



**Figure 22.** Distribution of all the events predicted by NET-VISA (squares) and the LEB (stars). The color version of this figure is available only in the electronic edition.



**Figure 23.** Precision-recall performance of the proposed NET-VISA and deployed SEL3 algorithms, treating the analyst-generated LEB as ground truth. The color version of this figure is available only in the electronic edition.

Note that the birth move is not a greedy move: the proposed event will almost always have a score  $S_e(e^i) < 1$  until some number of arrivals are assigned in subsequent moves. The overall structure of these moves could be easily converted to a Markov chain Monte-Carlo (MCMC) method or

**Table 4**  
Recall ( $m_b$ )

$m_b$	Count	SEL3		NET-VISA	
		Recall	Error	Recall	Error
Unknown	74	64.9	101	83.8	76
2–3	36	50.0	186	88.9	130
3–4	558	66.5	104	88.0	107
> 4	164	86.6	70	92.1	66
All	832	69.7	99	88.5	97

Recall and average error (km) subdivided by the LEB event magnitude ( $m_b$ ).

**Table 5**  
Recall (Azimuth Gap)

Azimuth Gap	Count	SEL3		NET-VISA	
		Recall	Error	Recall	Error
0–90	55	100.0	22	100.0	33
90–180	260	93.5	66	96.5	55
180–270	273	59.7	120	88.3	105
270–360	244	48.8	173	77.0	161
All	832	69.7	99	88.5	97

Recall and average error (km) subdivided by the LEB event azimuth gap.

simulated annealing algorithm. However, in our experiments this search outperformed simple MCMC methods in terms of speed and accuracy.

## Experimental Results

### Metrics

We compute the accuracy of an event history hypothesis by comparison to a chosen ground-truth history. A bipartite graph is created between predicted and true events. An edge is added between a predicted and a true event that is at most  $5^\circ$  in distance and 50 s in time apart. The weight of the edge is the distance between the two events. Finally, a minimum weight–maximum cardinality matching is computed on the graph. We report three quantities from this matching, precision (percentage of predicted events that are matched), recall (percentage of true events that are matched), and average error (average distance in kilometers between matched events).

### Summary of Results

A visual analysis of events appearing in the NET-VISA and SEL3 bulletins highlights the obvious differences between them. In Figure 21 we show all the SEL3 and LEB events for a week, and in Figure 22 we show all the NET-VISA and LEB events for the same week. The figures show that the NET-VISA events are much more closely aligned to the LEB events.

For a more formal comparison, we quantify our results by measuring the precision, recall, and average error of our predictions with respect to an assumed ground truth. We first

Table 6  
LEB Origin-ID 5287957 Locations

Bulletin	Longitude (°)	Latitude (°)	Depth (km)	$m_b$	Time	$S_e(\cdot)$	$P_{\theta,l}(\cdot)$
LEB	177.67	-33.29	0.0	3.5	1237691617.7	13.3	-10.9
NET-VISA	-179.61	-33.50	0.7	3.3	1237691599.7	16.0	-9.9
SEL3	170.48	-32.42	0.0	3.8	1237691669.5	-1.8	-17.2

The locations of various bulletin events around the LEB origin-ID 5287957.

Table 7  
LEB Origin-ID 5287957 Associations

Bulletin	Station: Phase	Station: Phase	Station: Phase
LEB	URZ: $S_n$	ASAR: $P$	WRA: $P$
$S_d(\cdot)$	7.8	13.1	15.0
NET-VISA	URZ: $S_n$	ASAR: $P$	WRA: $P$
$S_d(\cdot)$	8.8	12.6	14.8
SEL3	PLCA: $P$	ASAR: $P$	WRA: $P$
$S_d(\cdot)$	1.2	13.0	15.1

The station–phase associations of various bulletin events around the LEB origin-ID 5287957.

treat the IDC analyst-generated LEB as the ground truth, and compare the performance of our NET-VISA algorithm to the currently deployed GA system and the SEL3 bulletin it produces. Because there is a natural trade-off between precision and recall, for example, perfect recall can be achieved at the expense of precision by reporting events in all locations at all times, it is common in statistics and machine learning to compute a precision–recall curve showing the best recall that can be achieved for each possible level of precision (or vice versa). In NET-VISA, the curve is generated by adjusting the score threshold for including events in the bulletin: a higher threshold increases precision but lowers recall (see Fig. 23). Because we are not able to adjust the GA software, we have marked SEL3 on the graph as a single point. NET-VISA has at least 18% more recall at the same precision as the SEL3, and at least 38% more precision at the same recall as the SEL3. Put another way, NET-VISA reduces the number of missed events by about 60%.

Also in this figure, we show an extrapolated precision–recall curve for the SEL3 using scores from a Support Vector Machine trained to classify true and false SEL3 events from the work of Mackey *et al.* (2009). Next, we discuss further the right (high-precision) end of this curve.

For completeness, we note the following run-time statistics: The NET-VISA inference algorithm used a window size,  $W$ , of 30 minutes; a step size,  $S$ , of 15 minutes; and  $N = 1000$  iterations. The inference for 1 week of data took about 4.5 days on a single CPU core running at 2.5 GHz. Estimating model parameters from 2.5 months of training data took about 1 hour.

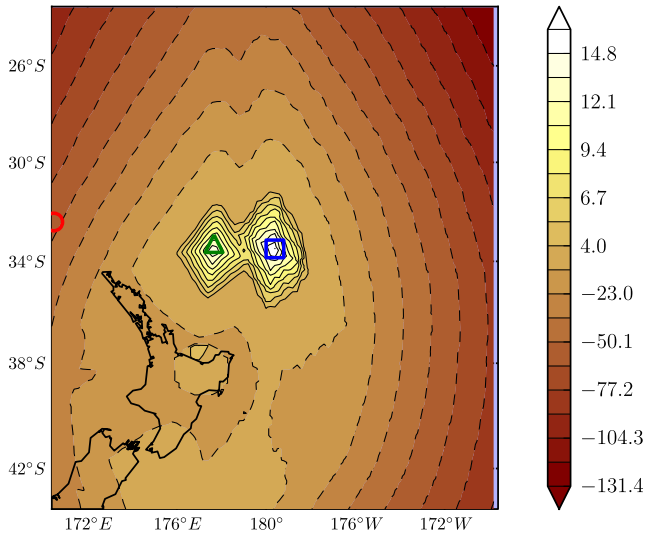
The arrivals include those from both the primary and auxiliary stations in the IMS network. Although both the NET-VISA bulletin and the SEL3 were produced using the same set of arrivals, the GA algorithm treats arrivals from auxiliary IMS stations differently. These arrivals are not al-

lowed to drive event formation. NET-VISA makes no such distinction, and thus enjoys a slight advantage.

### Underconstrained Events

To further understand why NET-VISA is able to find events missed by the SEL3, we subdivide the NET-VISA and the SEL3 recall and average error by two different criteria. In Table 4 we subdivide by the LEB event magnitude. For magnitudes up to 4, NET-VISA has nearly 20% higher recall with similar error. In Table 5 we subdivide by the LEB event azimuth gap. The azimuth gap of an event is the largest difference between successive event-to-station azimuths for stations where the automated processing detected an arrival for the event. Large gaps indicate that the event location is underconstrained. For example, if all stations are to the southwest of an event, the gap is greater than  $270^\circ$ , and the event will be poorly localized along a line running from southwest to northeast. The results in these two tables highlight a common theme: NET-VISA performs significantly better than the SEL3 whenever there are less data available. Under these circumstances the additional information in the NET-VISA model, location prior, amplitude, nondetections, etc., plays a critical role in determining a better location for the events.

In Table 6 we show the location of one such underconstrained LEB event, origin-ID 5287957, the nearest NET-VISA event, which is  $2^\circ$  and 18 s away, and the nearest SEL3 event, which is  $8^\circ$  and 52 s away. Although there are very few arrivals to help locate this event, and the azimuth gap is more than  $270^\circ$ , it is worth noting that NET-VISA is able to use the event location prior, shown in the last column of Table 6, to choose a more likely location for the event. Further, NET-VISA associates this event with an  $S_n$  phase at station URZ (Urewera, New Zealand) without a corresponding  $P_n$  phase at the same station, something SEL3 would avoid. The associations of the events to the automatically detected arrivals are displayed in Table 7. In fact, the human analysts only associated the  $S_n$  phase because they manually added a new arrival, which they associated as the  $P_n$  phase. Of course, the additional  $P_n$  arrival was not available to either NET-VISA or the SEL3. This particular event also demonstrates the perils of choosing the MAP event location. In Figure 24 we plot the contours of  $\max[S_e(e_l = \cdot, e_{m,t} = e_{m,t}^{\text{LEB}})]$ ,  $S_e(e_l = \cdot, e_{m,t} = e_{m,t}^{\text{NET-VISA}})$ . Or, in other words, at each point we plot the score of an event using either the LEB event magnitude, time, depth, and associations or the corresponding values for the NET-VISA event, whichever is better for that



**Figure 24.** SEL3 (circle), LEB (triangle), and NET-VISA (square) locations and posterior log odds ratio around the LEB origin-ID 5287957. The color version of this figure is available only in the electronic edition.

location. As this figure suggests, the posterior event location density is multimodal. The mode picked by NET-VISA happens to be  $2^\circ$  away from the mode near the LEB location.

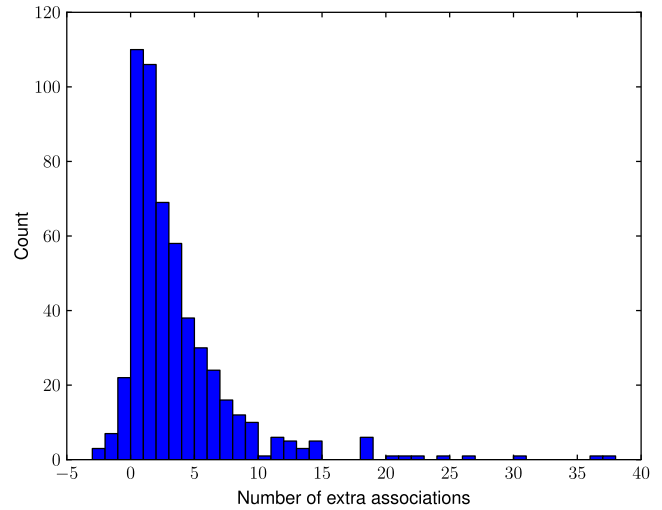
#### Phase Association

Another factor contributing to the better results of NET-VISA is that it associates many more phases per event than GA. Figure 25 shows a histogram of the number of additional associations for events located by both. We have already alluded to one reason for this. Global association follows certain rules for its associations; for example, an  $Sn$  phase can only be associated if a  $Pn$  phase is associated. NET-VISA, on the other hand, allows greater flexibility in associations. For example, it allows for the possibility that the  $Pn$  phase might be missed, but the  $Sn$  phase is detected. Another reason for the additional phases is that GA is unable to change a phase label once this is determined early in the processing pipeline. In contrast, NET-VISA distinguishes between true phase and the phase label assigned by station processing and can revise its phase hypotheses to optimize the bulletin.

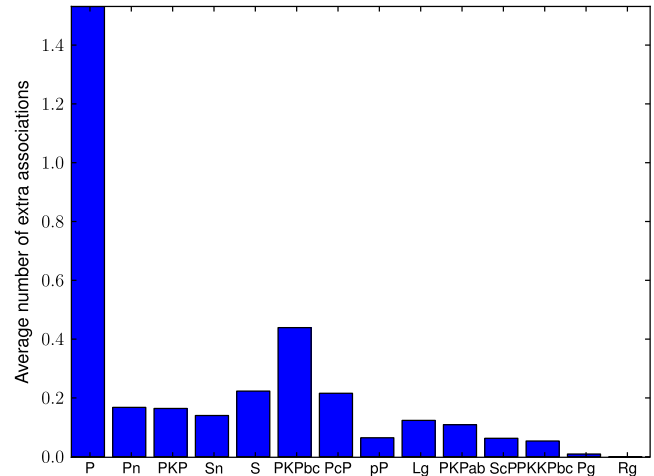
The distribution of additional phases is shown in Figure 26. This graph shows that NET-VISA is getting many more of the first-arriving  $P$  phases, which are critical for event location.

#### Depth

As described in equation (3), NET-VISA assumes *a priori* that seismic events are equally likely at all depths. This assumption gives it a slight disadvantage in areas of known deep seismicity. In such areas, prior knowledge of likely event depths could have helped constrain the event location. Table 8 summarizes the relative performance of the SEL3 and NET-VISA broken down by event depth. The results



**Figure 25.** The number of additional phases associated by NET-VISA versus the SEL3 for common events. The color version of this figure is available only in the electronic edition.



**Figure 26.** The average number of additional phases associated per event by NET-VISA versus the SEL3 for common events. The color version of this figure is available only in the electronic edition.

**Table 8**  
Recall (Depth)

Depth	Count	SEL3		NET-VISA	
		Recall	Error	Recall	Error
0	595	66.2	105	86.4	101
0–100	93	86.0	64	100.0	52
100–200	79	75.9	87	98.7	98
200–300	21	81.0	75	100.0	166
300–400	13	69.2	55	92.3	102
400–500	15	73.3	211	66.7	121
500–600	12	58.3	131	41.7	108
600–700	4	50.0	164	50.0	182
All	832	69.7	99	88.5	97

Recall and average error (km) subdivided by the LEB event depth.

Table 9  
Regional Results

Region	Count	LEB			NET-VISA		
		Precision	Recall	Error	Precision	Recall	Error
Japan	1565	100.0	1.9	38	83.3	2.2	104
U.S.	132	100.0	3.0	33	90.5	14.4	93
Italy	96	50.0	1.0	49	66.7	2.1	55
Kazakhstan	65	73.3	16.9	57	70.4	29.2	63

Precision, recall, and average error (km) of the LEB and NET-VISA measured against various regional bulletins.

in the table suggest that NET-VISA has slightly poorer results for events at depth more than 400 km, although there are too few deep events in this dataset to support a definite finding.

### Comparison with Regional Bulletins

Returning to Figure 23, the gap between the SEL3 extrapolation and NET-VISA on the lower right end of the figure suggests that NET-VISA is predicting spurious events with extremely high confidence. In reality, many of these events are indeed real events that are missed by the human analysts and hence absent from the LEB. To understand the true scope of the LEB, we have compared it against various regional bulletins, which are based on data from many more stations than in the IMS. In Table 9 we consider both the LEB and NET-VISA bulletins restricted to specific regions of the earth and evaluate them against the corresponding regional bulletins. This table shows that in all regions NET-VISA finds new events not appearing in the LEB, most notably in the United States and Kazakhstan. The results in the continental United States are further subdivided by the local magnitude ( $M_L$ ) in Table 10. These results indicate that NET-VISA is able to find roughly half of the events with  $M_L$  between 3 and 4.5, as well as some weaker events, while the LEB finds only 7% of events with  $M_L$  between 3 and 4.5 and none below this.

The exact demarcations of these regions and the bulletins used are described in Table 11. The NEIC bulletin was used as the ground truth for the continental United States. This bulletin was downloaded from the Incorporated Research Institutions for Seismology (IRIS). For the other regions we relied upon the raw bulletins from the International Seismological Center (ISC; see [Data and Resources](#)). The relevant agency codes are displayed in the table.

*ISC Event 13437052.* As a concrete example we display the waveforms of one of the events in central Asia that is absent in the SEL3 and the LEB, but is found by NET-VISA and confirmed by a regional bulletin. This event has been given the event identifier 13437052 by ISC. (See Table 12 for the event location.) NET-VISA forms the events using regional phases ( $P_n$ ,  $P_g$ ,  $S_n$ ,  $L_g$ ) at three IMS stations (MKAR, AAK, and KURK). The waveforms for the three stations are displayed

Table 10  
Continental U.S. Results ( $M_L$ )

$M_L$	Count	LEB		NET-VISA	
		Recall	Error	Recall	Error
Unknown	13	23.1	35	23.1	60
1.0–2.0	23	0.0	–	0	–
2.0–2.5	48	0.0	–	4.2	112
2.5–3.0	35	0.0	–	20.0	118
3.0–4.5	13	7.1	28	53.8	78
All	132	3.0	33	14.4	93

Recall and average error (km) subdivided by event magnitude ( $M_L$ ) in the continental United States.

Table 11  
Regional Boundaries

Region	Longitude (°)	Latitude (°)	Ground Truth
Japan	130 to 145	30 to 45	JMA bulletin (ISC)
U.S.	–125 to –70	25 to 50	NEIC bulletin (IRIS)
Italy	6 to 19	36 to 48	ROM bulletin (ISC)
Kazakhstan	46 to 86	40 to 55	NNC bulletin (ISC)

The definition of the various regions used for the regional evaluation and the corresponding ground-truth bulletin.

in Figures 27, 28, and 29, respectively. In all three figures the first arrival comes in around 12 s into the waveform. The figures show the raw waveforms, two filtered versions, and the STA/LTA for the higher frequency filtered waveform. The automatic arrivals are marked with the STA/LTA as dotted vertical lines. Of these three, MKAR is a primary IMS seismic station, while the other two are auxiliary seismic stations. While the event satisfies the criteria for inclusion in the LEB, it lacks the three primary stations required for inclusion in the Reviewed Event Bulletin (REB); had the event been found by NET-VISA and presented to an analyst, it is possible that further primary arrivals could have been added manually.

### IDC Evaluation

Finally, we present results of an independent evaluation of NET-VISA by the IDC. The results were broadly similar to



Table 12  
ISC Event 13437052 Location

Bulletin	Longitude (°)	Latitude (°)	Depth (km)	$m_b$	Time	$S_r(-)$
NNC	81.53	42.37	3.9	3.2	1237760634.43	
NET-VISA	82.7	42.5	0.4	2.2	1237760637.8	47.2

The location of ISC event 13437052 in the regional bulletin (NNC) and NET-VISA.

those reported in the preceding paragraphs, but the evaluation criteria used by IDC differed from ours in two aspects. First, the Reviewed Event Bulletin was used as ground truth instead of the LEB. The REB is a subset of the LEB events that satisfy certain rules, the most critical of these rules being that three primary IMS stations must detect each event. Second, an REB event was considered matched if it was within  $18^\circ$  of epicentral distance and 120 s of time of the predicted event and if it shared at least two time-defining arrivals with a predicted event (where time-defining arrivals are those satisfying a set of rules designed to ensure accuracy).

The IDC evaluation was conducted for two different time periods. First, a three-month interval with normal seismic activity was considered. During this time period, NET-VISA and GA (SEL3) were running concurrently on the same set of arrivals in near real time. For the second time period, NET-VISA was evaluated on historical data from the large aftershock sequence following the March 2011 Tohoku earthquake. These results are summarized in Table 13. The table also reports under the unmatchable column the subset of REB events that lacked two automatically picked time-defining arrivals. These events are effectively unmatchable by both the SEL3 and NET-VISA, which are built exclusively from automatically picked arrivals. The most striking result from the Tohoku evaluation is that nearly all (98.4%, to be precise) of the NET-VISA events matched

an REB event, while at the same time matching 15% more events than the SEL3. A possible explanation for this unusually small false positive rate from NET-VISA is that most of the events in the Tohoku aftershock sequence are fairly large events that are detected across many stations in the IMS

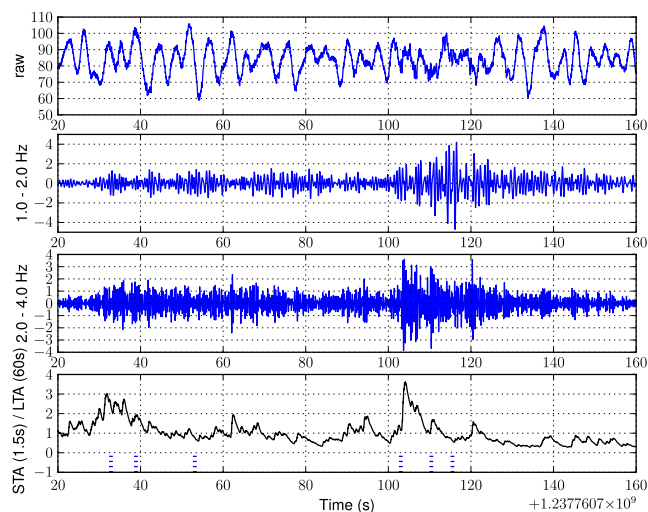


Figure 28. Waveforms at station AAK (channel BHZ) for ISC event 13437052. The color version of this figure is available only in the electronic edition.

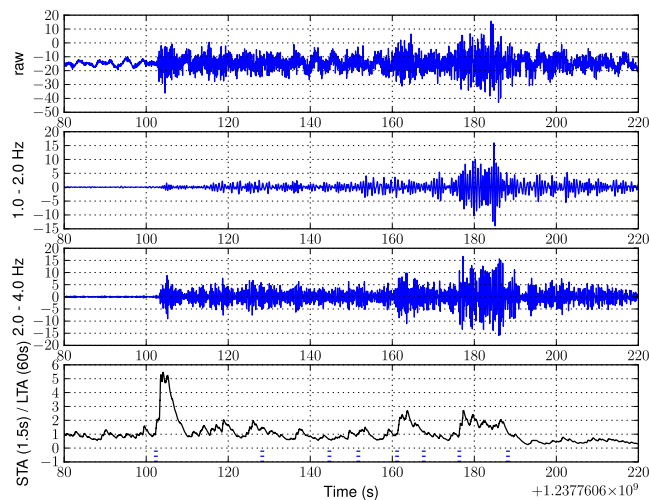


Figure 27. Waveforms at station MKAR (channel BHZ) for ISC event 13437052. The color version of this figure is available only in the electronic edition.

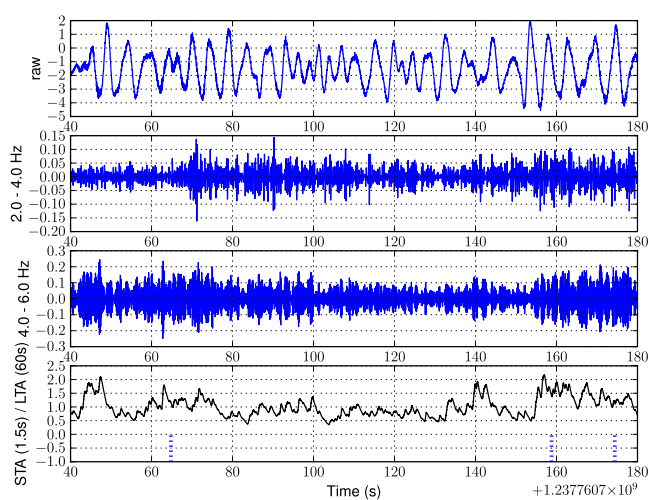


Figure 29. Waveforms at station KURK (channel BHZ) for ISC event 13437052. The color version of this figure is available only in the electronic edition.

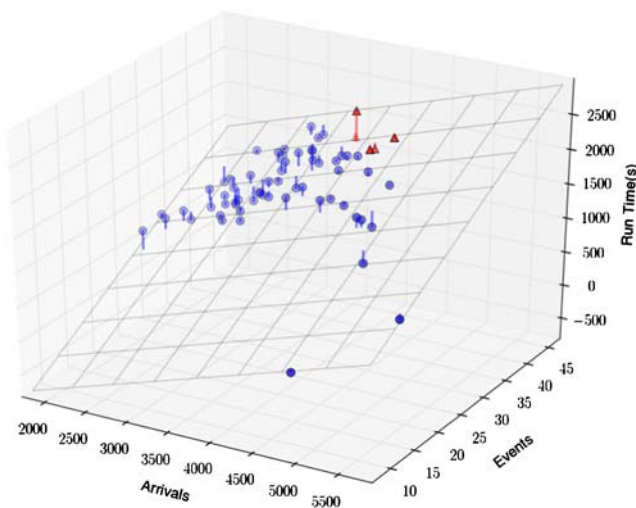


Table 13  
Results of IDC Evaluation

Time Period		REB		SEL3		NET-VISA	
Start	Duration	Total	Unmatchable	Total	Matched	Total	Matched
11 June 2012	3 months	7898	340	12,458	5921	13,575	6814
05:30 11 March 2011	27.5 hrs	1078	66	1208	747	876	862

network, and in particular many primary stations, thus satisfying the criteria for REB inclusion.

The other important metric measured in this evaluation was the running time of NET-VISA. For a normal period, this would not be a concern because NET-VISA easily keeps up in real time on a machine with six CPU cores. For a large aftershock sequence, on the other hand, the running time can be considerably higher. To obtain near-real-time performance for the Tohoku sequence, NET-VISA was run on a machine with 160 CPU cores. The running time of NET-VISA on each 30-minute interval of the Tohoku workload on this large machine is shown as a scatter plot in Figure 30. The  $x$  axis is the number of arrivals considered by the birth move (i.e., the number of arrivals in the 30-minute birth window plus the subsequent 2000 seconds), and the  $y$  axis is the number of events produced by the birth move. In this graph, all running times above 30 minutes are marked with triangles. For visualization purposes, a best-fit hyperplane and the residuals are also drawn in this figure. The graph shows that the running time is very nearly linear in the number of birth events and arrivals. Most importantly, it shows that given a sufficiently powerful machine, NET-VISA can keep up in near real time on the most punishing IDC workload to date.



**Figure 30.** Running times on Tohoku workload versus the number of birth events and arrivals. The triangles mark run times in excess of 30 minutes. The graph includes a best-fit hyperplane and the residuals. The color version of this figure is available only in the electronic edition.

For a small subset (26 hours) of the three-month workload, IDC analysts also manually reviewed the NET-VISA bulletin confirming the salient points of the evaluation presented in this paper. Full details of that analysis are provided in [Arora et al. \(2012\)](#).

### Future Work

Our current model makes a number of simplifying assumptions that we plan to address in our future work. We outline a few of the important ones.

Our assumption that seismic events occur independently of each other breaks down in the aftershock sequence of a large earthquake. For example, the rate of aftershocks in the immediate vicinity of the 2011 Tohoku earthquake was five times the rate of seismic events for the whole earth. One solution that we are pursuing is to continuously update our model parameters after every time interval. For example, the expected rate of seismic events at each spot on the earth can be adjusted upward or downward based on the actual number and magnitude of events in the last interval. [Omori \(1894\)](#) has proposed a specific model for how such adjustments should be made.

Another aspect of our model that seems to be untenable after a large event is the assumption that the event detection probabilities and the noise arrival distributions remain constant. Because a large earthquake can considerably increase the background noise levels for many days, it can dramatically alter the detection probabilities; for example, some aftershocks with magnitude 6 can become almost undetectable by most stations. Continuously relearning the models is one approach to address this shortcoming. Another approach is to incorporate the current background noise level directly into the model, for example, as a detection probability feature in [Table 2](#).

Our model also includes many conditional independence assumptions that may be questionable. For example, the detection probabilities in our model do not depend on azimuth. In fact, due to the local geology around a station, it is quite conceivable that a station might have widely varying detection probabilities for similar-sized events coming in from different directions. Moreover, earthquakes do not radiate energy uniformly in all directions, and hence might be preferentially detected by stations in certain event-to-station azimuths over others. Similar considerations hold for the arrival amplitude model as well. A comprehensive solution would be to build an azimuth-specific detection model for

each station and to add the moment tensor as an event attribute in the model. The energy radiation pattern for a seismic event is clearly very important for the CTBTO because it could be used as a discriminant for explosions and earthquakes. The radiation pattern can also account for the correlations between the detection probabilities of nearby stations, as well as the correlations between different phases at the same station.

Elaborating the model by adding dependencies and new hidden variables (such as the moment tensor) is certainly feasible, but may require additional historical data to estimate the necessary parameters. Whether such steps improve monitoring performance remains to be seen. As noted in the [Introduction](#), we are also extending the generative model downward to include waveform characteristics. In this way, detection becomes part of a globally integrated inference process and hence susceptible to top-down influences, rather than being a purely local, bottom-up, hard-threshold decision.

## Conclusions

Our results demonstrate that a Bayesian approach to seismic monitoring can improve significantly on the performance of classical systems. The NET-VISA system cannot only reduce the human analyst effort required to achieve a given level of accuracy, but it can also lower the magnitude threshold for reliable detection. Given that the difficulty of seismic monitoring was cited as one of the principal reasons for non-ratification of the CTBT by the United States Senate in 1999, one hopes that improvements in monitoring may increase the chances of final ratification and entry into force.

Putting monitoring onto a sound probabilistic footing also facilitates further improvements such as continuous estimation of local noise conditions, travel time, and attenuation models without the need for ground-truth calibration experiments (controlled explosions). Moreover, it facilitates an open-source approach, whereby various expert groups can devise and test more refined and accurate model components and contribute them as modules in an open probabilistic architecture.

## Data and Resources

A three-month dataset of IMS arrivals, covering the period 22 March to 20 June 2009, was made available by the CTBTO through the Virtual Data Exploitation Center, or vDEC ([Vaidya et al., 2009](#)). All of the results described here except for the IDC evaluation were produced using seven days of data from the validation set (22 March to 29 March). There were a total of 832 LEB events during this period and roughly 120,000 arrivals. The training set was 2.5 months (5 April to 20 June), including 8313 LEB events and 1,163,848 arrivals.

Additional data were obtained for regional bulletin comparisons from IRIS at <http://www.iris.edu/SeismiQuery/sq-events.htm> (last accessed January 2012) and from ISC at <http://www.isc.ac.uk/iscbulletin/search/catalogue/> (last ac-

cessed July 2011). International Seismological Center is available at <http://www.isc.ac.uk> (last accessed July 2011).

## Acknowledgments

We would like to thank the many seismologists who patiently explained to us the intricacies of their field, among them Ronan LeBras, Robert Engdahl, David Bowers, Bob Pearce, Stephen Myers, Dmitry Storchak, Istvan Bondar, and Barbara Romanowicz. Paul Kidwell of the Lawrence Livermore National Laboratory designed the features for the event detection probability and wrote the initial parameter estimation code for the corresponding weights. We also received assistance from several Berkeley undergraduates, including Matthew Cann, Hong Hu, Christopher Lin, Andrew Lee, Tony Dear, and Ron Sun.

We are grateful to Alexander Poplavskiy for providing the results of the IDC evaluation. The specific evaluation criteria was devised by Dmitry Bobrov.

All the results in this paper except for the IDC evaluation were obtained on the Amazon EC2 cluster, which was made possible by an AWS in Education grant award.

The facilities of the IRIS Data Management System, and specifically the IRIS Data Management Center, were used for access to waveform and metadata required in this study. The IRIS DMS is funded through the National Science Foundation and specifically the GEO Directorate through the Instrumentation and Facilities Program of the National Science Foundation under Cooperative Agreement EAR-0552316.

All the authors were partially supported by the Preparatory Commission for the CTBT.

## References

- Allen, R. (1978). Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.* **68**, 1521–1532.
- Arora, N. S., J. Given, E. Tomuta, S. Russell, and S. Spiliopoulos (2012). Analyst evaluation of model-based Bayesian seismic monitoring at the CTBTO, in *The 34th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, Albuquerque, New Mexico.
- Bar-Shalom, Y., and T. Fortmann (1988). *Tracking and Data Association*, Academic Press, Orlando, Florida.
- Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein (2009). *Introduction to Algorithms*, Third Ed., MIT Press and McGraw-Hill, Cambridge, Massachusetts.
- Duijndam, A. J. W. (1988a). Bayesian-estimation in seismic inversion. Part 2. Uncertainty Analysis, *Geophys. Prospect.* **36**, no. 8, 899.
- Duijndam, A. J. W. (1988b). Bayesian estimation in seismic inversion. Part I: Principles, *Geophys. Prospect.* **36**, no. 8, 878–898.
- Geiger, L. (1910). Herdbestimmung bei erdbeben ans den ankunftszeiten, *K. Gessel. Wiss. Goett.* **4**, 331–349 (in German).
- Geiger, L. (1912). Probability method for the determination of earthquake epicenters from the arrival time only, *Bull. St. Louis Univ.* **8**, 60–71.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis*, Chapman & Hall, Boca Raton, Florida.
- Gutenberg, B., and C. Richter (1954). *Seismicity of the Earth and Associated Phenomena*, Second Ed., Princeton University Press, Princeton, New Jersey.
- Lay, T., and T. C. Wallace (1995). *Modern Global Seismology*, Academic Press, San Diego, California.
- Le Bras, R., H. Swanger, T. Sereno, G. Beall, and R. Jenkins (1994). Global association, *Science Applications International Corp. Tech. Rept. ADA304805*, San Diego, California.
- Mackey, L., A. Kleiner, and M. I. Jordan (2009). Improved automated seismic event extraction using machine learning (abstract S31B-1714), *Eos Trans. AGU* **90**, no. 52, Fall Meet. Suppl.

- Magotra, N., N. Ahmed, and E. Chael (1987). Seismic event detection and source location using single-station (three-component) data, *Bull. Seismol. Soc. Am.* **77**, 958–971.
- Menke, W. (1989). *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, San Diego, California.
- Myers, S. C., G. Johannesson, and W. Hanley (2007). A Bayesian hierarchical method for multiple-event seismic location, *Geophys. J. Int.* **171**, 1049–1063.
- Omori, F. (1894). On the aftershocks of earthquakes, *Journal of the College of Science, Imperial University of Tokyo* **7**, 111–200.
- Roberts, R. G., A. Christofferson, and F. Cassidy (1989). Real-time event detection, phase identification and source location estimation using single station three-component seismic data, *Geophys. J.* **97**, 471–480.
- Storchak, D. A., J. Schweitzer, and P. Bormann (2003). The IASPEI standard seismic phase list, *Seismol. Res. Lett.* **74**, no. 6, 761–772.
- Vaidya, S., R. Engdahl, R. Le Bras, K. Koch, and O. Dahlman (2009). Strategic initiative in support of CTBT data processing: vDEC (virtual Data Exploitation Centre) (abstract DM-01/A), in *CTBTO International Scientific Studies*, Hofburg, Vienna, Austria, 10–12 June 2009.
- Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California, *Bull. Seismol. Soc. Am.* **90**, 1353–1368.
- Bayesian Logic, Inc.  
32432 Lighthouse Way  
Union City, California 94587  
nimar.arora@gmail.com  
(N.S.A.)
- Computer Science Division,  
387 Soda Hall  
University of California  
Berkeley, California 94720  
russell@cs.berkeley.edu  
(S.R.)
- Department of Computer Science  
Brown University  
115 Waterman Street, Box 1910  
Providence, Rhode Island 02912  
sudderth@cs.brown.edu  
(E.S.)

Manuscript received 26 March 2012