Global seismic monitoring as probabilistic inference

Nimar S Arora
Stuart J. Russell
Paul Kidwell
Erik Sudderth

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2010-108
http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-108.html

July 7, 2010

Acknowledgement

We are very grateful to the help of various seismologists who very patiently explained to us the intricacies of seismology and provided us the data used in this paper. We would like to thank Ronan LeBras, Robert Engdahl, David Bowers, Bob Pearce, Dmitry Storchak, Istvan Bondar, and Barbara Romanowicz. We acknowledge the help provided by various undergraduate students in Berkeley---Matthew Cann, Hong Hu, Christopher Lin, and Andrew Lee. The third author's work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The other authors were partially supported by the Preparatory Commission for the CTBT.

Global seismic monitoring as probabilistic inference

Nimar S. Arora
Department of Computer Science
University of California, Berkeley
Berkeley, CA 94720
nimar@cs.berkeley.edu

Stuart Russell
Department of Computer Science
University of California, Berkeley
Berkeley, CA 94720
russell@cs.berkeley.edu

Paul Kidwell
Lawrence Livermore National Lab
Livermore, CA 94550
kidwell1@llnl.gov

Erik Sudderth
Department of Computer Science
Brown University
Providence, RI 02912
sudderth@cs.brown.edu

Abstract

The International Monitoring System (IMS) is a global network of sensors whose purpose is to identify potential violations of the Comprehensive Nuclear-Test-Ban Treaty (CTBT), primarily through detection and localization of seismic events. We report on the first stage of a project to improve on the current automated software system with a Bayesian inference system that computes the most likely global event history given the record of local sensor data. The new system, VISA (Vertically Integrated Seismological Analysis), is based on empirically calibrated, generative models of event occurrence, signal propagation, and signal detection. VISA exhibits significantly improved precision and recall compared to the current operational system and is able to detect events that are missed even by the human analysts who post-process the IMS output.

1 Introduction

The CTBT aims to prevent the proliferation and the advancement of nuclear weapon technology by banning all nuclear explosions. A global network of seismic, radionuclide, hydroacoustic, and infrasound sensors, the IMS, has been established to enforce the treaty. The IMS is the world's primary global-scale, continuous, real-time system for seismic event monitoring. Data from the IMS sensors are transmitted via satellite in real time to the International Data Center (IDC) in Vienna, where automatic event-bulletins are issued at predefined latency. 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 post-processes the automatic bulletins to improve their accuracy to acceptable levels.

Like most current systems, the IDC operates by *detection* of arriving signals at each sensor station (the *station processing* stage) and then grouping multiple detections together to form *events* (the *network processing* stage).[1] The time and location of each event are found by various search methods including grid search, [2], the double-difference algorithm [3], and the intersection method [4]. In the words of [5], "Seismic event location is—at its core—a minimization of the difference between observed and predicted arrival times." Although the mathematics of seismic event detection and

[1] Network processing is thus a *data association* problem similar to those arising in multitarget tracking [1].

localization has been studied for almost 100 years [6], the IDC results indicate that the problem is far from trivial.

There are three primary sources of difficulty: 1) the *travel time* between any two points on the 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) there are many thousands of detections per day, so the combinatorial 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 *non-detections* as negative evidence for a hypothesized event, something that classical methods cannot do.

In simple terms, let X be a random variable ranging over all possible collections of events, with each event defined by time, location, magnitude, and type (natural or man-made). Let Y range over all possible waveform signal recordings at all detection 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 observed recordings $Y = y$, we are interested in the posterior $P(X|Y = y)$, and perhaps in the value of X that maximizes it—i.e., the most likely explanation for all the sensor readings. We also learn the model parameters θ and ϕ from historical data.

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 algorithms. (The second stage, SIG-VISA, will incorporate a signal waveform model and thereby subsume the detection function.) NET-VISA computes a single most-likely explanation: a set of hypothesized events with their associated detections, marking all other detections as noise. This input–output specification, while not fully Bayesian in spirit, enables direct comparison to the current automated system bulletin, SEL3. Using the final expert-generated bulletin, LEB, as ground truth, we compared the two systems on 7 days of held-out data. NET-VISA has 13% more recall at the same precision as SEL3, and 24% more precision at the same recall as SEL3. Furthermore, taking data from the more comprehensive NEIC (National Event Information Center) database 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; thus, NET-VISA's true performance may be higher than the LEB-based calculation would suggest.

The rest of the paper is structured as follows. Section 2 describes the problem in detail and covers some elementary seismology. Sections 3 and 4 describe the probability model and inference algorithm. Section 5 presents the results of our evaluation, and Section 6 concludes.

2 Detailed problem description

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—waves breaking, trees falling, ice falling, etc.—may generate seismic waves too. All such waves occur in a variety of types [7]—*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 crust and these are labeled distinctly by seismologists. Each particular wave type generated by a given event is called a *phase*; These waves are picked up in seismic stations as ground vibrations. Typically, seismic stations have either a single 3-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 so are quite susceptible to noise.

Raw seismometer measurements are run through standard signal processing software that filters out non-seismic frequencies and computes short-term and long-term averages of the signal amplitude. When the ratio of these averages exceeds a fixed threshold, a *detection* is announced. Various parameters of the detection 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*, etc.

Based on these parameters, a phase label may be assigned to the detection based on the standard IASPEI phase catalog [7]. All of these detection attributes may be erroneous.

The problem that we attempt to solve in this paper is to take a continuous stream of detections (with onset time, azimuth, slowness, amplitude, and phase label) from the roughly 120 IMS seismic stations as input and produce a continuous stream of events and associations between events and detections. The parameters of an event are its longitude, latitude, depth, time, and magnitude (m_b or body-wave magnitude). A 3-month dataset (660 GB) has been made available by the IDC for the purposes of this research. We have divided the dataset into 7 days of validation, 7 days of test, and the rest as training data. 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 are at most 5 degrees in distance and 50 seconds in time apart. The weight of the edge is the distance between the two events. Finally, a max-cardinality min-weight matching is computed on the graph. We report 3 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).

3 Generative Model

Our generative model for seismic events and detections follows along the lines of the aircraft detection model in [8, Figure 3]. In our model, there is an unknown number of seismic events with unknown parameters (location, time, etc.). These events produce 14 different types of seismic waves or phases. A phase from an event may or may not be detected by a station. If a phase is detected at a station, a corresponding detection is generated. However, the parameters of the detection may be imprecise. Additionally, an unknown number of noise detections are generated at each station. Overall, the only thing observed is each station's set of detections and their parameters.

3.1 Events

The events are generated by a time-homogeneous Poisson process. If e is the set of events, λ_e is the rate of event generation, and T is the time-period under consideration.

$$P_\theta(|e|) = \frac{(\lambda_e \cdot T)^{|e|} exp\left(-\lambda_e \cdot T\right)}{|e|!} \tag{1}$$

The longitude and latitude of the ith event, e_l^i are given by an event location density, p_l on the surface of the earth. The depth of the event, e_d^i is uniformly distributed up to a max-depth D (700 km). Similarly, the time of the event e_t^i is uniformly distributed between 0 and T. The magnitude of the event, e_m^i, is given by what seismologists refer to as a Gutenberg-Richter distribution, which is in fact an exponential distribution with rate of $log(10)$, λ_m.

$$P_\theta(e^i) = p_l(e_l^i)\frac{1}{D}\frac{1}{T}\lambda_m exp\left(-\lambda_m e_m^i\right) \tag{2}$$

Since all the events are exchangeable.

$$\begin{aligned} P_\theta(e) &= P_\theta(|e|) \cdot |e|! \cdot \prod_{i=1}^{|e|} P_\theta(e^i) && (3) \\ &= exp\left(-\lambda_e \cdot T\right) \prod_{i=1}^{|e|} p_l(e_l^i)\frac{1}{D}\lambda_e\lambda_m exp\left(-\lambda_m e_m^i\right) && (4) \end{aligned}$$

λ_e is easily estimated from the data, by using the maximum likelihood estimate. For measuring p_l, we divide the earth into 2 degree buckets of equal surface area and count the number of events in these buckets. The counts are then smoothed by absolute discount smoothing [9] with discount=.5 and normalized. See Figure 1.

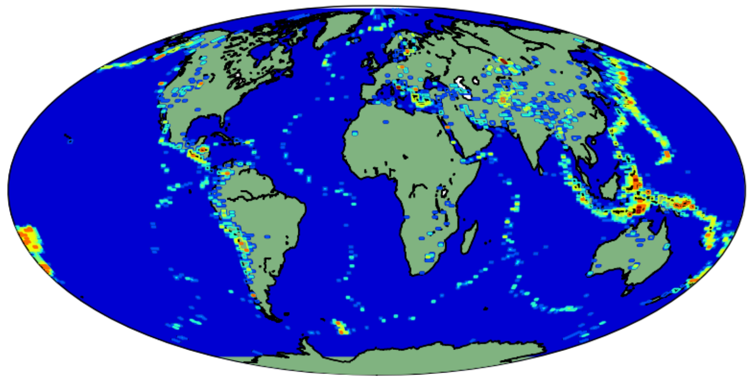

Figure 1: Heat map of the log of the location density, p_l.

3.2 Detections

The probability, p_d^{jk}, that an event i's jth phase (1 to J) gets detected by a station k (1 to K) depends on the phase and the station and is a function of the event magnitude, depth, and great-circle distance to the station. Let d^{ijk} be an indicator variable for this detection and Δ_{ik} the distance between the event i and station k.

$$P_\phi(d^{ijk} = 1|e^i) = p_d^{jk}(e^i_m, e^i_d, \Delta_{ik}) \tag{5}$$

If an event-phase is detected at a station, i.e. $d^{ijk} = 1$, our model specifies the generation of the attributes of that detection, a^{ijk}. The arrival time, a_t^{ijk}, has a Laplacian distribution. The mean of this Laplacian consists of two parts. One is the IASPEI prediction for the phase which depends only on the event depth and the distance between the event and the station. The second is a station-specific correction which accounts for the inhomogeneities in the earth's crust. These inhomogeneities allow seismic waves to travel faster or slower than the IASPEI prediction. The station-specific correction also includes the error in picking seismic onsets from waveforms. Let μ_t^{jk} be the location of this Laplacian which depends on the event time, depth, and distance to the station and let b_t^{jk} be the scale. Since the Laplacian distribution extends to infinity while our arrival time is bounded we need a normalization constant, Z_t, to account for the truncation.

$$P_\phi(a_t^{ijk}|d^{ijk} = 1, e^i) = \frac{1}{Z_t^{jk}} exp\left(-\frac{|a_t^{ijk} - \mu_t^{jk}(e^i_t, e^i_d, \Delta_{ik})|}{b_t^{jk}}\right) \tag{6}$$

Similarly, the arrival azimuth and slowness follow a Laplacian distribution. The location of the arrival azimuth, a_z^{ijk}, depends only on the location of the event. While the location of the arrival slowness, a_s^{ijk}, depends only on the event depth and distance to the station. The scales of all the Laplacians are fixed for a given phase and station.

$$P_\phi(a_z^{ijk}|d^{ijk} = 1, e^i) = \frac{1}{Z_z^{jk}} exp\left(-\frac{|a_z^{ijk} - \mu_z^{jk}(e^i_l)|}{b_z^{jk}}\right) \tag{7}$$

$$P_\phi(a_s^{ijk}|d^{ijk} = 1, e^i) = \frac{1}{Z_s^{jk}} exp\left(-\frac{|a_s^{ijk} - \mu_s^{jk}(e^i_d, \Delta_{ik})|}{b_s^{jk}}\right) \tag{8}$$

The arrival amplitude, a_a^{ijk} is similar to the detection probability in that it depends only on the event magnitude, depth, and distance to the station. We model the log of the amplitude as a linear function with Gaussian error.

$$P_\phi(a_a^{ijk}|d^{ijk}=1,e^i) = \frac{1}{\sqrt{2\pi}\sigma_a^{jk}} exp\left(-\frac{(log(a_a^{ijk}) - \mu_a^{jk}(e_m^i, e_d^i, \Delta_{ik}))^2}{2{\sigma_a^{jk}}^2}\right) \tag{9}$$

Finally, the phase, a_h^{ijk}, assigned to the detection follows a simple multinomial distribution which depends on the true phase, j.

$$P_\phi(a_h^{ijk}|d^{ijk}=1,e^i) = p_h^{jk}(a_h^{ijk}) \tag{10}$$

The phase- and station-specific detection probabilities, p_d^{jk}, were obtained using logistic regression models estimated via an empirical Bayes procedure. Each phase defines among other things the general physical path taken from an event to a station, therefore a distinct set of features were learned from the event characteristics for each phase via maximum likelihood. An empirical distribution for detection probabilities was then constructed for each phase by pooling data among all stations. Finally, pseudo data was generated from the empirical distribution conditional with respect to the features to form a prior [10] which was combined with the individual station data to estimate the respective models. The estimation of the other distributions was done similarly by maximum likelihood estimates on phase-site specific data combined with empirical data for the phase from all sites. Figure 3.2 shows two of the empirical and modeled distribution for one phase-site.

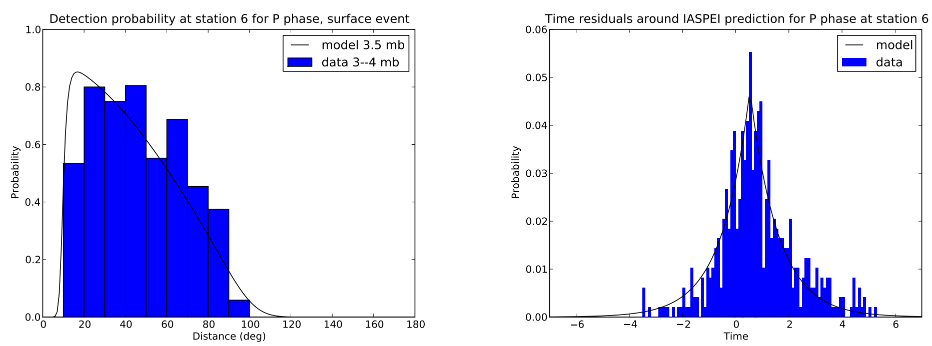

Figure 2: Detection probability and Arrival Time distribution (around IASPEI prediction) for the P phase at Station 6

3.3 False Detections

Each station, k, also generates a set of false detections f^k through a time-homogeneous Poisson process with rate, λ_f^k.

$$P_\phi(|f^k|) = \frac{(\lambda_f^k \cdot T)^{|f^k|} exp\left(-\lambda_f^k \cdot T\right)}{|f^k|!} \tag{11}$$

The time, azimuth, and, slowness of these false detections—f_t^{kl}, f_z^{kl}, and f_s^{kl} respectively—are generated from the uniform distribution. The amplitude of the false detection, f_a^{kl} is generated from a mixture of two Gaussians, p_a^k. Finally, the assigned phase of the false detection, f_h^{kl}, follows a multinomial distribution, p_h^k. If M_z is the range of values for azimuth (360 degrees) and M_s is the range of values for slowness, then the probability of the lth false detection is given by.

$$P_\phi(f^{kl}) = \frac{1}{T}\frac{1}{M_z}\frac{1}{M_s} p_a^k(f_a^{kl}) p_h^k(f_h^{kl}) \tag{12}$$

Since the false detections at a station are exchangeable, we have

$$P_\phi(f^k) = P_\phi(|f^k|) \cdot |f^k|! \cdot \prod_{l=1}^{l=|f^k|} P_\phi(f^{kl}) \tag{13}$$

$$= exp\left(-\lambda_f^k \cdot T\right) \prod_{l=1}^{l=|f^k|} \frac{\lambda_f^k}{M_z M_s} p_a^k(f_a^{kl}) p_h^k(f_h^{kl}) \tag{14}$$

3.4 Overall

The overall probability of any hypothesized sequence of events e, detected event phases d, arrival attributes a for detected event phases, and arrival attributes f for falsely detected events equals:

$$P(e, d, a, f) = P_\theta(e) P_\phi(d|e) P_\phi(a|d, e) P_\phi(f) \tag{15}$$

4 Inference

We will attempt to find the most likely explanation consistent with the observations. In other words, we will find e, d, a, and f which maximize $P(e, d, a, f)$, such that the set of detections implied by d, a, and f correspond exactly with the observed detections. Since the detections in the real world are observed incrementally and roughly in time-ascending order our inference algorithm also produces an incremental hypothesis which advances with time. Our algorithm is mainly a form of greedy search where we try to improve the current hypothesis with a set of local moves.

Initially, we start with a event-window of size W from $t_0 = 0$ to $t_1 = W$ and a detection-window of size $W + M_T$ (where M_T is the maximum travel time for any phase) from $t_0 = 0$ to $t_1 = W + M_T$. We start with the hypothesis that all the detections in our detection-window are false detections and there are no events. Then we repeatedly apply the birth, death, improve-event, and improve-detection moves (described below) for a fixed number, N, of iterations and then move our windows forward by a step size S. The new events added to the detection window are again assumed to be false detections. As the windows move forward the events older than $t_0 - M_T$ become stable in the sense that none of the moves modify either the event or detections associated with them. These events are then output. In theory, this algorithm never needs to terminate. However, in practice we terminate it when the dataset is fully consumed.

In order to simplify the probability computations needed to compare alternate hypotheses we will decompose the probability into the contribution from each event. We define the score of an event as the probability ratio of two hypotheses—one in which the event exists and the other in which the event doesn't exist and all of its associated detections are noise. If an event has score less than 1 then the alternate hypothesis in which the event doesn't exist, clearly has a higher probability. The most important property of this score is that it is unaffected by other events in the current hypothesis. From Equations 4, 14.

$$\text{score-ev}(e^i) = \frac{p_l(e_l^i)\lambda_e\lambda_m}{D} exp\left(-\lambda_m e_m^i\right) \prod_{j,k} P_\phi(d^{ijk}|e^i) \left(1_{(d^{ijk}=0)} + 1_{(d^{ijk}=1)} \cdot \frac{P_\phi(a^{ijk}|d^{ijk}, e^i)}{\frac{\lambda_f^k}{M_z M_s} p_h^k(f_h^{kl})}\right)$$

Note that the quantity in the last fraction is comparing the likelihood of the same detection as either the detection of event i's jth phase at station k or the lth false detection at station k. We can further decompose the score into a score for each detection. The score of d^{ijk} (defined if $d^{ijk} = 1$) is the ratio of the probabilities of the hypothesis where the detection is associated with the event-phase $i - j$ at station k and one in which event-phase $i - j$ is not detected at station k and this detection is a false detection

$$\text{score-det}(d^{ijk}) = \frac{p_d^{jk}(e_m^i, e_d^i, \Delta_{ik})}{1 - p_d^{jk}(e_m^i, e_d^i, \Delta_{ik})} \frac{P_\phi(a^{ijk}|d^{ijk}, e^i)}{\frac{\lambda_f^k}{M_z M_s} p_a^k(f_a^{kl}) p_h^k(f_h^{kl})} \tag{16}$$

Again, it is worth nothing that a score less than 1 suggests that changing a detection to a false detection would improve the probability. Also, the score of a detection doesn't depend on other detections or events in the hypothesis.

1. *Birth Move*: The birth move is executed only once in each event window. To propose the birth move we search for events on a fixed grid of points. We divide longitude and latitude in 1 degree buckets, depth in 100 km buckets, time in 5 second buckets within the event window and magnitude in steps of .5. At each of these grid points we hypothesize an event and attempt to add the best detections to it. At each station we add the best detection with a score > 1. Finally, the score of all the possible events is computed and we pick the event with the greatest score > 1. If an event is found then it is added to a list and all of its associated detections are marked unavailable. The algorithm is repeated over the same grid of events but with fewer detections. Finally, when no more events are found the list of events is added to the hypothesis.

   The events are added to the hypothesis without any associated detections. Therefore this move is technically a downhill move. However, it is followed by other moves which will either add more detections and make these events viable or kill them.

2. *Improve Detections Move*: For each detection in the detection window we consider all possible events from $t_0 - M_T$ up to the time of the detection and consider all possible phases for these events to find the best event-phase for it. If the best event-phase has a score < 1 or the best event-phase has already another detection with a higher score then this detection is changed to a false detection.

3. *Death Move*: Any event with a score < 1 is killed and all its detections are marked as noise.

4. *Improve Events Move*: For each event we look for 1000 points chosen uniformly at random in a small ball around the event (5 degrees in longitude and latitude, 200 km in depth, 50 seconds in time, and 1 units of magnitude) for an event with a higher score. If such an event is found we change the event parameters.

5. *Final Pruning*: Before outputing events we perform a final round of pruning to remove some duplicate events. We prune out any event which has another event with a higher score and within 5 degrees distance and 50 seconds time. The reason for these spurious events is that true events generate a number of superfluous detections at each site. These detections don't correspond to any event-phase of the original event but taken together they do suggest a new event at about the same location and time as the original event. Our model does allow for false detections but not false detections which are associated with a true event. Thus we need to explicitly prune these events for now.

5 Experimental Results

As we described in Section 2, we normally measure the precision and recall of our prediction versus an assumed ground truth. First, we consider the LEB as ground truth and compare the performance of NET-VISA and SEL3. Using our event scores we have generated an ROC curve for NET-VISA and marked SEL3 on it as a point, see Figure 3. As shown in the figure, NET-VISA has 13% more recall at a slightly higher precision than SEL3 and 24% more precision at the same recall as SEL3. On the same figure we plot the results of two other experiments. NET-VISA(-amp) was run without observing the detection amplitudes and in NET-VISA(-phase) the detection phase was not observed. Both of these results are inferior by about $2 - 3\%$. All the predicted NET-VISA, SEL3 and LEB events are plotted in Figure 4.

The true precision of NET-VISA is perhaps higher than these figures show. We have evaluated the recall of LEB and NET-VISA with the NEIC dataset as ground truth. Since the NEIC has many more sensors in the United States than the IMS, it is considered a better ground truth for this region. Out of 33 events in the continental United States, LEB found 4 and NET-VISA found 6 including the 4 found by LEB. Interestingly, both of the additional events found by NET-VISA had arrival times which deviated significantly from the expected arrival time. However, NET-VISA was able to discount these discrepancies as within the reasonable bounds of uncertainty for the relevant stations. In Figure 5 we show a heatmap of the posterior density around one such event in Colorado plus the predictions made by NET-VISA, NEIC, and SEL3[2].

[2]The SEL3 prediction was off by more than 5 degrees and hence counts as a miss per our evaluation criteria. LEB did not report this event.

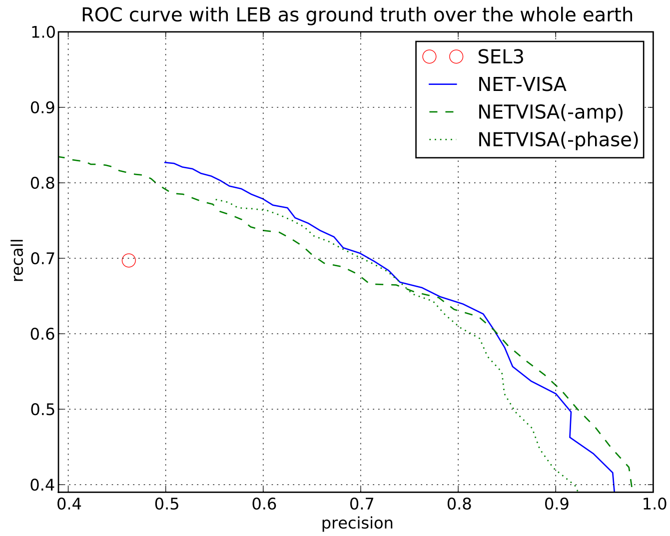

Figure 3: ROC curve for NET-VISA with LEB as ground truth. NET-VISA(-amp) doesn't observe the detection amplitude. NET-VISA(-phase) doesn't observe the detection phase.

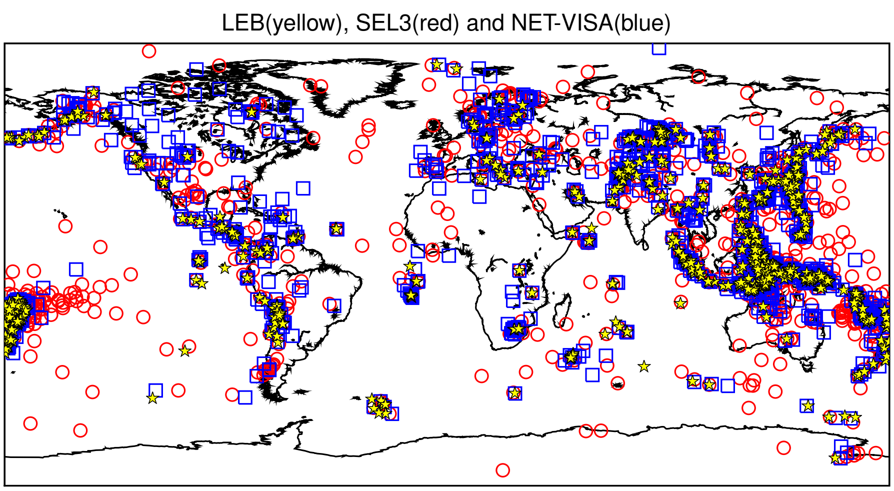

Figure 4: All the predicted events

All the results in this section were produced with 7 days worth of data from the validation set. The inference used a window size, W, of 30 minutes, a step size, S, of 5 minutes, and $N = 10$ iterations. There were a total of 832 LEB events during this period. All the computation was performed on a single core of a $2.5GHz$ Intel Core 2 Quad processor running Linux 2.6. It took 1.6 days to create and cache the results of the birth moves and 4 hours to complete the rest of the inference on the

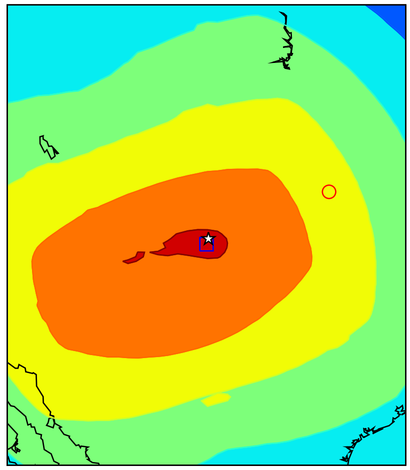

Figure 5: Posterior density in a 10 degree ball around an NEIC event. The NEIC event is marked in white, the NET-VISA prediction in blue, and SEL3 in red.

7 days of validation data. The parameter estimation on the 2.5 months of training data took 30 minutes.

6 Conclusions and Further Work

Our results demonstrate that a Bayesian approach to seismic monitoring can improve significantly on the performance of classical systems. The NET-VISA system can not only reduce the human analyst effort required to achieve a given level of accuracy, but can also lower the magnitude threshold for reliable detection.

Putting monitoring onto a sound probabilistic footing also facilitates further improvements such as continuous estimation of local noise conditions and travel time and attenuation models without the need for ground-truth calibration data such as controlled explosions or earthquakes that are localized within a few km by a dense local network of sensors. We also expect to lower the detection threshold significantly by extending the generative model to include waveform characteristics, so that 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.

Acknowledgments

We are very grateful to the help of various seismologists who very patiently explained to us the intricacies of seismology and provided us the data used in this paper. We would like to thank Ronan LeBras, Robert Engdahl, David Bowers, Bob Pearce, Dmitry Storchak, Istvan Bondar, and Barbara Romanowicz. We acknowledge the help provided by various undergraduate students in Berkeley—Matthew Cann, Hong Hu, Christopher Lin, and Andrew Lee. The third author's work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The other authors were partially supported by the Preparatory Commission for the CTBT. Finally, the first author wishes to thank his family for their infinite patience and support without which this work would not have been possible.

References

[1] Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*. Academic Press, 1988.

[2] P. M. Shearer. Improving local earthquake location using the L1 norm and waveform cross correlation: Application to the Whittier Narrows, California, aftershock sequence. *J. Geophys. Res.*, 102:8269 – 8283, 1997.

[3] F. Waldhauser and W. L. Ellsworth. A double-difference earthquake location algorithm: method and application to the Northern Hayward Fault, California. *Bulletin of the Seismological Society of America*, 90:1353 – 1368, 2000.

[4] J. Pujol. Earthquake location tutorial: graphical approach and approximate epicentral location techniques. *Seis. Res. Letter*, 75:63 – 74, 2004.

[5] Stephen C. Myers, Gardar Johannesson, and William Hanley. A Bayesian hierarchical method for multiple-event seismic location. *Geophysical Journal International*, 171:1049–1063, 2009.

[6] L. Geiger. Probability method for the determination of earthquake epicenters from the arrival time only. *Bull. St. Louis Univ.*, 8:60 –71, 1912.

[7] D. A. Storchak, J. Schweitzer, and P. Bormann. The IASPEI standard seismic phase list. *Seismol. Res. Lett.*, 74(6):761 – 772, 2003.

[8] Brian Milch, Bhaskara Marthi, Stuart J. Russell, David Sontag, Daniel L. Ong, and Andrey Kolobov. BLOG: Probabilistic models with unknown objects. In *IJCAI*, pages 1352–1359, 2005.

[9] Stanley F. Chen and Joshua Goodman. An empirical study of smoothing techniques for language modeling. In *ACL*, pages 310–318, 1996.

[10] B.P. Carlin and T.A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, New York, second edition edition, 2000.