Global seismic monitoring: A Bayesian approach

Nimar S. Arora and **Stuart Russell**
University of California, Berkeley
Berkeley, CA 94720
nimar,russell@cs.berkeley.edu

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

Erik Sudderth
Brown University
Providence, RI 02912
sudderth@cs.brown.edu

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. 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 to the currently deployed system. It also finds events that are missed even by the human analysts who post-process the IMS output.

1 Introduction to the Problem

Seismic *events* are large disturbances in the earth's crust, caused primarily by earthquakes and explosions (nuclear or conventional). They generate seismic *waves* that travel through the earth and generate measurable *signals* at detector *stations*. The basic problem is this: given continuous signal traces from multiple stations, determine what events have occurred and their onset times, latitudes, longitudes, depths, magnitudes, and types (natural or man-made).

What might seem at first sight to be a simple triangulation problem is actually extraordinarily complex and far from being solved. There are several sources of difficulty:

- Seismic waves occur in several types and follow a variety of qualitatively distinct paths through the earth; seismologists recognize over 100 different type/path combinations, called *phases* (Storchak, Schweitzer, and Bormann 2003). Velocities vary with wave type, depth, and geological properties of the medium, and the *travel time* between any two points on the earth and the *attenuation* of various frequencies and wave types are not known accurately.
- Each detector is subject to local *noise* that may mask true signals and cause false detections; in wide-area monitoring systems, up to 90% of all detections are false.
- Phases from a given event may arrive at a distant station between fifteen minutes and several hours after the event, while dozens or hundreds of events may occur each day, generating thousands of overlapping sets of detections. The combinatorial problem of proposing and comparing possible events (subsets of detections) is daunting.

The mathematics of seismic event detection and localization has been studied for almost 100 years (Geiger 1912). Most systems operate in two-stage pipeline: *station processing* is responsible for *detection* of arriving signals that exceed ambient noise levels by a given threshold, while *network processing* groups detections together to form event hypotheses. (Thus, it resembles the *data association* problem in multitarget tracking (Bar-Shalom and Fortmann 1988).) Station processing attaches attributes to each detection: *onset time*, *azimuth* (direction from the station to the wave source), *slowness* (related to the angle of declination of the signal path), *amplitude*, *phase label*, etc., all of which may be erroneous. The time and location of each event are typically found by heuristic methods such as grid search (Shearer 1997), soft-constraint solving (Waldhauser and Ellsworth 2000), and wavefront intersection (Pujol 2004), combined with *ad hoc* scoring functions to resolve ambiguity.

Our work is aimed in particular at seismic monitoring for nuclear explosions using the UN's International Monitoring System (IMS) for the Comprehensive Nuclear-Test-Ban Treaty (CTBT). The IMS is the world's primary global-scale, continuous, real-time system for seismic event monitoring. Data from 120 IMS stations are transmitted via satellite in real time to the International Data Center (IDC) in Vienna, where event-bulletins are issued at predefined latency. Perfect performance remains well beyond the reach of current technology: the final (SEL3) bulletin from 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.

2 Bayesian Framework and Model

The pervasive presence of model uncertainty, noise, and combinatorial complexity suggest that a Bayesian framework might be effective in extracting maximally accurate hypotheses from the data. Whereas pipelined approaches using local decisions and hard thresholds are ubiquitous in

large sensing systems of all kinds, we believe that a *vertically integrated* probability model connecting raw data to high-level hypotheses can be far more effective, since it enables all available information to be brought to bear on the interpretation of each locally ambiguous datum. Our work to date (Arora et al. 2010; Russell, Vaidya, and Le Bras 2010) demonstrates the effectiveness of even a partial realization of this approach. Our initial system, NET-VISA, handles the network processing stage, [1] relying on the IDC's pre-existing signal detection algorithms. The second phase, SIG-VISA, will incorporate a signal waveform model and thereby subsume the detection function.

In simple terms, let X be a random variable ranging over all possible sets of events, and let Y range over all possible sets of detections at all detection stations. Then $P_\theta(X)$ describes a parameterized generative prior over events and their properties, and $P_\phi(Y \mid X)$ describes how the signal is propagated and measured (including travel time, attenuation, noise, artifacts, sensor error, etc.). Given observed recordings $Y = y$, we are interested in the posterior $P(X \mid Y = y)$ or the MAP $\arg\max_x P(X = x \mid Y = y)$. We also learn the model parameters θ and ϕ from historical data. Below, we describe the model components in enough detail to illustrate their close relationship to the way in which a seismologist understands the physical processes involved. For readers familiar with Bayesian networks, we should note that although the model is a composition of prior and conditional distributions, it is not a Bayesian network as classically defined because the number of events, and hence the number of random variables in the model, is unknown; it is, however, expressible as a BLOG model (Milch et al. 2005).

2.1 Events

Geophysically speaking, each point on the earth is considered to generate seismic events according to a time-homogeneous Poisson process whose rate parameter varies with location and depth. Mathematically, this is equivalent to a single Poisson process with rate λ_e, generating events whose locations and depths are sampled from a spatial density proportional to the seismic event rate at that point. The magnitude of each event is drawn from an exponential distribution with decay parameter λ_m (known in seismology as the *Gutenberg–Richter* distribution). Maximum likelihood estimates of λ_e and λ_m may be determined from historical event frequencies and magnitudes. To approximate the spatial density of natural events, we use a kernel density model with an exponentially decaying kernel; the decay parameter was estimated by cross-validation (see Figure 1). To this density model we added a uniform distribution, with prior probability 0.001, to allow for explosions at an arbitrary location.

[1] NET-VISA computes a single *max a-posteriori* (MAP) bulletin — a set of hypothesized events with their associated detections; this input-output specification, while not fully Bayesian in spirit, is compatible with and enables direct comparison to the current automated system bulletin, SEL3.

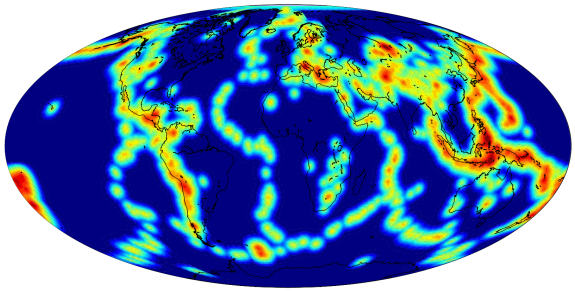

Figure 1: Heat map (large values in red, small in blue) of the prior event location density.

2.2 Correct Detections

An event can generate up to J distinct phases (we consider just the 14 most commonly detected phases). For each phase and each station, we model the probability of detection as a function of the event's magnitude, depth, and the travel time to the station. Specifically, we use a logistic regression with some predefined features and a hierarchical combination of station-specific and station-independent models to improve estimation at stations with sparse data.

If an event phase is detected at a station, the model specifies probability distributions for the observed attributes of that detection, conditioned on the event properties:

- The arrival time is the event time plus a travel time whose distribution is a Laplacian; the mean of the distribution is the sum of the standard spherically symmetric "IASPEI" travel time prediction for that phase, which depends only on the event depth and the distance between the event and station, and a learned station-specific correction that accounts for inhomogeneities in the earth's crust and corrects for any systematic biases in picking onset times from waveforms. The variances are estimated per station from historical data. (A more sophisticated model would include distance-dependent variances.)
- The azimuth and slowness also follow Laplacian distributions, with peaks pointing to the event location and variances estimated per station from historical data—some stations are *much* more accurate than others.
- The arrival amplitude 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.
- Finally, we model the phase label assigned by the station processing software as being randomly generated from a multinomial distribution whose parameters depends on the true phase; again, the multinomial distribution is learned from historical data.

Figure 2 shows two of the empirical and modeled distributions for one phase-site.

2.3 False Detections

Each station generates false detections, which may be the result of thermal and electrical noise or real but local events (ocean waves breaking, trees falling, ice cracking, and so

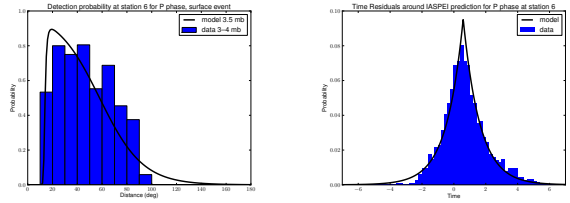

Figure 2: Conditional detection probabilities and arrival time distributions (relative to the IASPEI prediction) for the P phase at Station 6.

on). We assume a time-homogeneous Poisson process with a constant rate specific to each station. (A more sophisticated model would allow for time-varying and seasonal noise levels.) The station processing software, of course, does not know that the detections are false, and assigns them all the attributes of normal detections. We model the time, azimuth, and slowness of these false detections as drawn from a uniform distribution over their respective ranges. The amplitude of the false detection is modeled by an empirically estimated mixture-of-Gaussians model. Finally, the phase label follows an empirically estimated multinomial distribution,

2.4 Inference

The model components described above, when combined, determine the joint probability distribution over events and detections. As noted earlier, for compatibility reasons NET-VISA computes an approximate MAP bulletin—the most likely set of hypothesized events given the observed detections. Because detections from real seismic sensors are observed incrementally and roughly in time-ascending order, our inference algorithm also produces an incremental hypothesis which advances with time. Although the sequence of observations is of unbounded length, there is a physical limit on the travel time of any phase, so we can use a moving observation window of fixed length.

The most natural algorithm for finding a MAP bulletin in the temporal context would be some form of Viterbi algorithm, adapted to the context of a hypothesis space of unbounded size using, say, an online version of simulated annealing. In practice, we have found that an incremental, deterministic hill-climbing search suffices to find high-quality hypotheses by a sequence of local moves that improve the posterior likelihood. (Calculating the likelihood change due to each potential move is rendered very efficient because of decomposability across events and across detections assigned to an event.) Beginning with the hypothesis that all new detections in our window are false detections and there are no events, we repeatedly apply the following moves:

Birth Move We randomly pick a detection, invert it into an event location (using the detection's time, azimuth, and slowness), and sample an event in a 10 degree by 100 second ball around this inverted location. The depth of the event is fixed at 0, and the magnitude is uniformly sampled.

Improve Detections Move For each detection in the detection window, we consider all possible phases for all events that could have caused the detection. We then associate the best event-phase for this detection that is not already assigned to a higher-likelihood detection at the same station. If this best event-phase does not improve the overall likelihood, the detection is changed to a false detection.

Improve Events Move For each event, we consider 10 points chosen uniformly at random in a small ball around the event (2 degrees in longitude and latitude, 100 km in depth, 5 seconds in time, and 2 units of magnitude), and choose whichever point has highest likelihood (if better).

Death Move Any event making a negative contribution to the overall likelihood is deleted, and all of its currently associated detections are marked as false alarms.

Theoretically, this algorithm can reach a local maximum. We have experimented with various stochastic optimization algorithms that, in the limit, will find a MAP bulletin; occasionally, at great computational expense, these algorithms find a better maximum than our deterministic search, but the likelihood differences are usually negligible compared to those resulting from model improvements.

3 Experimental Results

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. A distance weighted edge is added between a predicted and a true event that are at most 5 degrees in great-circle distance and 50 seconds in time apart. Finally, a min-weight max-cardinality matching is computed on the graph. We report 3 quantities from this matching—*precision*, *recall*, and *average error* (average distance in kilometers between matched events).

Using the final expert-generated bulletin, LEB, as ground truth, we compared NET-VISA and SEL3 on 7 days of held-out data. Using the probabilities for hypothesized events, we have generated a precision-recall curve for NET-VISA, and marked SEL3 on it as a point (see Figure 3). Also in this figure, we show a precision-recall curve for SEL3 using scores from an SVM trained to classify true and false SEL3 events (Mackey, Kleiner, and Jordan 2009) (SEL3 extrapolation). NET-VISA has 18.4% more recall at the same precision as SEL3, and 32.6% 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.

Figure 4 shows the recall and error divided among different types of LEB events with NET-VISA precision fixed to that of SEL3. The table on the top summarizes by LEB event magnitude. For magnitudes up to 4, NET-VISA has nearly 20% higher recall with similar error. The table on the bottom shows a break-down by *azimuth gap*, defined as the largest difference in consecutive event-to-station azimuths for sta-

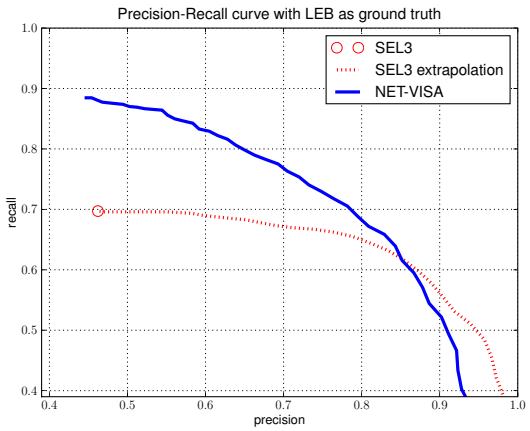

Figure 3: Precision-recall performance of the proposed NET-VISA and deployed SEL3 algorithms, treating the analyst-generated LEB as ground truth.

m_b	Count	SEL3		NET-VISA	
		Recall	Err	Recall	Err
0 – 2	74	64.9	101	89.2	103
2 – 3	36	50.0	186	86.1	140
3 – 4	558	66.5	104	86.2	121
> 4	164	86.6	70	93.9	77
Azimuth Gap	Count	SEL3		NET-VISA	
		Recall	Err	Recall	Err
0 – 90	72	100.0	28	100.0	39
90 – 180	315	88.9	76	93.7	75
180 – 270	302	51.0	134	84.4	137
270 – 360	143	51.0	176	76.9	198
all	832	69.7	99	88.1	112

Figure 4: Recall and error (km) broken down by LEB event magnitude and azimuth gap (degrees).

tions which detect an event. Large gaps indicate that the event location is under-constrained. For example, if all stations are to the southwest of an event, the gap is greater than 270 degrees and the event will be poorly localized along a line running from southwest to northeast. By using evidence about missed detections ignored by SEL3, NET-VISA reduces this uncertainty and performs much better.

Results in this section were produced using the validation dataset. There were a total of 832 LEB events during this 7 day period, and roughly $120,000$ detections at 117 stations. The inference took about 5.5 days on a single core running at 2.5 GHz. Estimating model parameters from 2.5 months of training data took about 1 hour.

4 Conclusions and Further Work

NET-VISA is an instance of a vertically integrated, Bayesian approach to sensor-based monitoring in an important application area where previous penetration of Bayesian methods had been only fragmentary (Myers, Johannesson, and Hanley 2009). NET-VISA misses half as many events as the currently deployed automated system and finds events missed by human experts; thus, it lowers 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.

In addition to a variety of physics-based model improvements and the integration of other sensor modalities (hydroacoustic and infrasound), some more generic advances are needed. The first is to have NET-VISA output (marginals of) the posterior distribution (easy) and integrate such output into the overall operations of the IDC (hard). The second is to reimplement NET-VISA within a declarative language such as BLOG, given suitable improvements to the inference engine; this would greatly facilitate further improvements to the model by domain experts. The third is to extend the generative model all the way to the raw signal level, so that detection itself becomes part of a globally integrated inference process rather than being a purely local, bottom-up, hard-threshold decision.

References

Arora, N.; Russell, S.; Kidwell, P.; and Sudderth, E. 2010. Global seismic monitoring as probabilistic inference. In *Advances in Neural Information Processing Systems (NIPS) 23*, 73–81. MIT Press.

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

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

Mackey, L.; Kleiner, A.; and Jordan, M. I. 2009. Improved automated seismic event extraction using machine learning. *Eos Trans. AGU* 90(52). Fall Meet. Suppl., Abstract S31B-1714.

Milch, B.; Marthi, B.; Russell, S. J.; Sontag, D.; Ong, D. L.; and Kolobov, A. 2005. BLOG: Probabilistic models with unknown objects. In *IJCAI*, 1352–1359.

Myers, S. C.; Johannesson, G.; and Hanley, W. 2009. A Bayesian hierarchical method for multiple-event seismic location. *Geophysical Journal International* 171:1049–1063.

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

Russell, S.; Vaidya, S.; and Le Bras, R. 2010. Machine learning for comprehensive nuclear-test-ban treaty monitoring. *CTBTO Spectrum* 14:32–35.

Shearer, P. M. 1997. 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.

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

Waldhauser, F., and Ellsworth, W. L. 2000. 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.