BAYESIAN TREATY MONITORING: PRELIMINARY REPORT

Stuart J. Russell,[1] Stephen C. Myers,[2] Nimar S. Arora,[1] David A. Moore,[1] and Erik Sudderth[3]

University of California, Berkeley,[1] Lawrence Livermore National Laboratory,[2] and Brown University[3]

Sponsored by the Defense Threat Reduction Agency

Award Nos. BRBAA08-Per5-I-2-0042
Proposal No. HDTRA1-08-10-BRCWMD-BAA

ABSTRACT

Our project has initiated and will develop and evaluate a new Bayesian approach for nuclear test monitoring. We anticipate that the new approach will yield substantially lower detection thresholds, possibly approaching a theoretical lower bound that we hope to establish. We will also develop new techniques to implement such monitoring capabilities within a general-purpose Bayesian modeling and inference system that may eventually support a wide range of information-system needs for arms treaties.

In ongoing work that is moving towards possible deployment, we have completed a prototype seismic monitoring system based on a generative, vertically integrated statistical model linking hypothesized events to “detections” extracted from raw signal data by classical algorithms. On test data sets of naturally occurring events curated by human experts, our system exhibits roughly 60% fewer detection failures than the currently deployed automated system, SEL3, that forms part of the International Monitoring System.

The current phase of the project moves away from hard-threshold detections altogether. Instead, the generative model spans the full range from events to measured signal properties. Given the observed signal traces, the statistical inference algorithm attempts to maximize a whole-network statistical measure of the likelihood that an event – or collection of events – has occurred. Specialized techniques such as *waveform matching* and *double differencing* are realized within our framework as special cases of probabilistic inference; our initial experiments using 2D simulated data indicate that a full Bayesian analysis can provide more accurate absolute and relative locations than double differencing, while simultaneously estimating the velocity structure of the observed region.

As we move toward a full-scale implementation, the primary tasks will involve the development of accurate predictive models of waveform properties. These models will combine both parametric forms (for example, triangular envelopes in multiple frequency bands) and nonparametric forms based on previously observed waveforms from nearby events. Hybrid models will smoothly interpolate between these two forms depending on the distance of the hypothesized event from previously observed events.

OBJECTIVES

Our objective is to realize dramatic improvements in sensitivity, accuracy, and robustness of global monitoring systems for nuclear tests through a novel Bayesian approach to whole-network data analysis.

The project will develop a new mathematical and computational approach to the analysis of the sensor data collected from global monitoring systems. The approach involves the application of rigorous Bayesian statistical analysis to the entire monitoring problem and requires the development of a complete, vertically integrated, empirically validated, generative statistical model of event occurrence, signal propagation, and sensing, as well as efficient and provably correct inference algorithms for extracting the most likely event history and/or a posterior distribution over event histories from the measured sensor data. At a fundamental level, we hope to gain a much deeper and more accurate understanding of the limits of detectability than is given by a classical per-station SNR analysis. We expect to realize the following benefits: (1) Substantially more accurate and sensitive detection and localization of events, particularly at lower magnitudes. (2) A monitoring software architecture that allows straightforward incorporation of multiple sensor modalities (including new modalities as they arise) and new and improved physics-based models such as source models and phase velocity and attenuation models. (3) Extensibility to other monitoring problems arising in treaty verification.

RESEARCH ACCOMPLISHED

The period of performance for the DTRA-funded project on the title page has not yet begun at the time of writing; therefore, this report covers some background material, our relevant past work, some initial steps taken for the new project, and a brief summary of our planned activities.

Background: Bayesian monitoring

The proposed research is motivated by a fundamental problem with current approaches to seismic monitoring, namely, the problem of robust detection of signals with low signal-to-noise ratio (SNR). Standard detection practice is to declare a detection when the average amplitude in a short time window increases above a predefined multiple of average amplitude in a long time window (a.k.a. short-term average/long-term average, STA/LTA). Independent algorithms then attempt to associate detections to seismic events. To avoid creating a flood of spurious events, a very high SNR threshold is typically used for detection. This approach is flawed in at least three important ways. First, it ignores the *absence* of detections in evaluating a proposed event. Second, it fails to take advantage of spatiotemporally *correlated* signals at multiple stations to detect that an event has occurred. Third, it discards details of the signal that support more accurate estimation of event properties via techniques such as waveform matching.

Our proposed approach is based on Bayesian inference, a method well-established in many areas of science that solve inverse problems including seismology itself (Duijndam, 1988ab). They have been very effective in tomographic applications (Taylor *et al*., 2001; Simmons *et al*., 2010b) and event localization (Myers *et al*., 2007, 2009) and have been applied to the general beamforming problem (Bell *et al*., 2000). Our work (Russell *et al*., 2009, 2010; Arora *et al*., 2009, 2010) is to our knowledge the first complete Bayesian monitoring system, solving both association and localization problems within a unified probability model.

In general, Bayesian inference yields a *posterior probability distribution* over a set of *hypotheses* X given some *evidence* $Y = y$. In seismic monitoring, a hypothesis is a collection of events occurring over space and time; the evidence consists of the raw seismic (and other) waveform signals from all sensors over the time period of interest. The inference process is based on a *probability distribution* or *model* with two components:

- The *prior probability distribution* $P_\theta(X)$ over hypotheses; for the monitoring problem, this would include the natural seismicity distribution on Earth. The subscript θ represents the parameters of this distribution, which can be learned from empirical data.
- The *conditional probability distribution* $P_\phi(Y=y \mid X)$ for the evidence given each possible hypothesis; in our case, this part of the model describes how signals propagate through the earth and how they are detected by sensors, as well as the ways in which noise signals arise. The subscript ϕ represents the parameters of this distribution; again, these can be learned from empirical data, but there is also a great deal of physics-based knowledge that constrains the possible values. In seismology, the model of propagation is often called the *forward model*, although it is important to note that in the Bayesian context it includes a quantitative characterization of uncertainty.

Bayes' rule simply multiplies these two components together to give the posterior probability distribution over the set of hypotheses, given the available evidence:

$$P(X \mid Y=y) = \alpha\, P_{\phi}(Y=y \mid X)\, P_{\theta}(X)$$

where α is a normalizing constant. To the extent that the prior and conditional distributions correctly describe knowledge of seismicity, propagation, and so on, the posterior distribution represents an optimal inference from the available data. An inference algorithm may compute a single *most likely explanation* or MLE hypothesis, instead of the full posterior distribution over hypotheses. Because there are infinitely many possible hypotheses (each a set of seismic events), the calculations involved are nontrivial and require efficient inversion of the forward model.

As a side effect of the inference process, the Bayesian approach generates information that can be used to continuously adapt the model parameters θ and ϕ to better explain the data. This adaptation requires no "ground truth" (unlike supervised learning methods) and hence allows for continuous self-calibration and sensor diagnostics.

Background: Detection-based Bayesian monitoring

When the raw data *y* are complex – as in the case of complete seismic waveforms – it is common for Bayesian methods to deal with a simplified representation $f(y)$ of the raw data. Here f can be any deterministic function. In *detection-based Bayesian monitoring*, f represents the signal processing algorithms that are used to pull out detections from the signal traces at each station. That is, $f(y)$ is the set of all detections (with associated characteristics) from all stations during the period of interest. The performance of a detection-based Bayesian monitoring system is limited by the underlying detection algorithm f, which typically abstracts away important details of the arriving waveform and which may fail to detect some arriving signals and may generate many spurious detections. Nonetheless, as we show below, the performance may be significantly better than state-of-the-art systems based on the same detection algorithms.

NET-VISA (Arora *et al*., 2010, Arora *et al*., 2011) consists of a collection of probability distributions and an inference algorithm that computes an MLE hypothesis. The probability distributions include:

- A prior distribution $P_T(e)$ over sets of events e for a given interval T. Each event is defined by time, location, depth, and magnitude. Event times are distributed according to a time-homogeneous Poisson process. (This is easily modified to allow for aftershocks.) Magnitudes are distributed according to the Gutenberg–Richter distribution. Locations for naturally occurring events follow a distribution estimated from historical data, whereas man-made events have a spatially uniform distribution with zero depth.
- For each station and each true phase, a detection probability distribution given the event magnitude, depth, and distance. The probability model (logistic regression with local input features) is calibrated from historical data that record which events were detected by which stations. (**Figure 1**(a).)
- For each detection, the following predictive distributions over attributes of the detection:
    - A distribution for the predicted arrival time, given by the event time plus a travel-time distribution for the true phase from the event location to the station. NET-VISA currently uses the IASPEI91 1-D model with station-specific corrections, i.e., the same model as the IMS. Uncertainty is modeled as a Laplacian residual, as illustrated in **Figure 1**(b). There is also an additional phase pick error distribution, estimated from historical ground truth data for each station.
    - A distribution for the measured amplitude given the event magnitude and distance – i.e., an attenuation model. The log amplitude is assumed to decay linearly with distance, with Gaussian error. (We expect to improve this with a more detailed attenuation model.)
    - A multinomial distribution for the measured phase label given the true phase label, again estimated from historical data.
    - Distributions for the measured azimuth and slowness given the true azimuth and slowness for the event also follow a Laplacian distribution.
- Finally, for each station a time-homogeneous Poisson for spurious detections, each with its amplitude drawn from an empirically estimated distribution (a mixture of Gaussians), its phase label from an empirically estimated multinomial distribution, and its azimuth and slowness from uniform distributions.

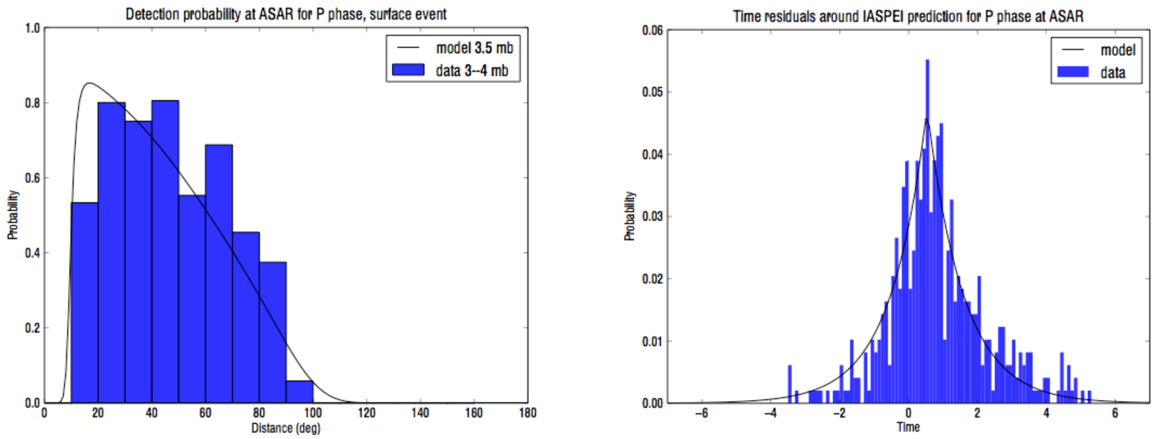

Figure 1: (a) P-phase detection probability at ASAR for magnitude 3.5 events, as a function of distance. (b) Laplacian fit to ASAR travel-time residual w.r.t. IASPEI91 predictions with station-specific corrections. Note the bias of 0.6 seconds even after station-specific correction; our model automatically compensates for this bias.

NET-VISA is currently being evaluated with a view to eventual deployment by the CTBTO. The evaluation compares the output of NET-VISA to the SEL3 bulletin (the final automated bulletin) produced by the CTBTO. For the purposes of this comparison, "ground truth" was defined by the LEB (Late Event Bulletin), which is generated by a team of expert analysts post-processing the SEL3 output. (The LEB itself is imperfect, as we see below.) The distribution parameters for NET-VISA were trained on 2.5 months of IMS data and tests were run on one week of held-out data. Details of the evaluation appear in the paper by Le Bras *et al*. (2011) in these proceedings. In summary, the low-magnitude event-detection failure rate for NET-VISA is 2.5–3 times lower than that for SEL3, as shown in **Figure 2**(a). This improvement is achieved despite the fact that the "ground truth" LEB is derived from SEL3 and is therefore perhaps biased towards SEL3's interpretation of the data. As shown in **Figure 2**(b), NET-VISA also finds events recorded by local networks (e.g., NEIC, JMA, PRU, NNC) that do not appear in LEB. Finally, NET-VISA was tested on one week of data that included the DPRK test of May 25, 2009; NET-VISA formed the event correctly with an accurate location based on associating 53 detections from the IMS, of which 50 were also associated in the LEB. (SEL3 included only 39 detections for the event.) This result suggests that NET-VISA's capability to detect manmade events is not compromised by its use of a natural seismicity prior.

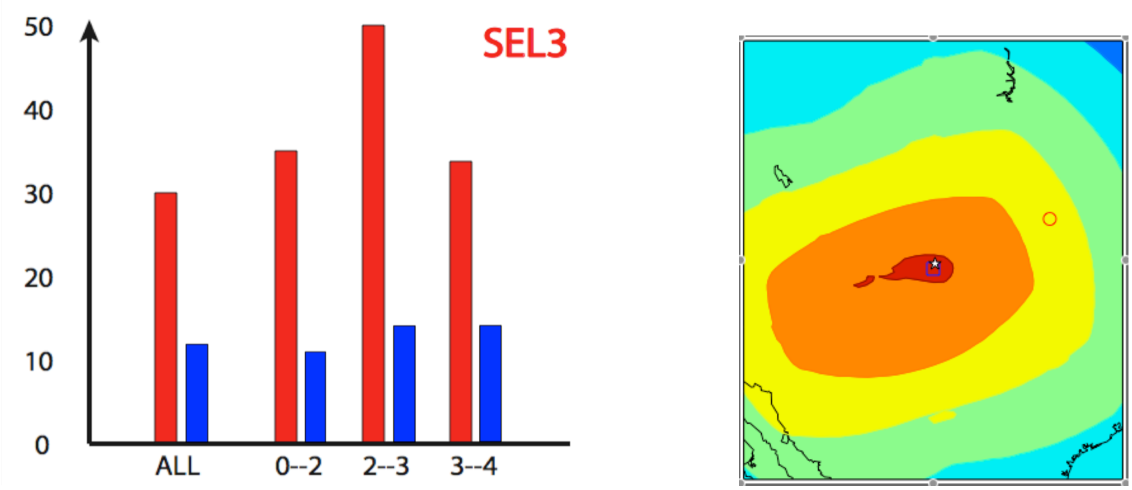

Figure 2: (a) Comparative detection failure rate for SEL3 and NET-VISA, for a test set of 852 events (1 week). Rates are shown for all events and for specific magnitude ranges. (b) Posterior location density calculated by NET-VISA for NEIC Event 12054557 (Colorado) using detections from three IMS stations (TXAR, PDAR, and ANMO), with the most likely location shown by the blue square. The white star shows the NEIC location using 19 stations; the red circle shows the nearest SEL3 location (approx. 10 degrees away); the event was not recorded in LEB.

Signal-based Bayesian monitoring

Detection-based systems such as NET-VISA are fundamentally limited by the ability of the underlying signal detection algorithm. Our current project will develop the required statistical models and algorithms for a *signal-based Bayesian monitoring system*, which takes as its observed data the network-wide raw signal (waveform) data itself, thereby avoiding the need for hard-threshold local detection algorithms. Whereas NET-VISA has established the feasibility and credibility of Bayesian monitoring, SIG-VISA will be a qualitatively different and technically more sophisticated approach requiring fundamental advances in understanding. The basic relationship between NET-VISA and SIG-VISA is shown in Figure 3.

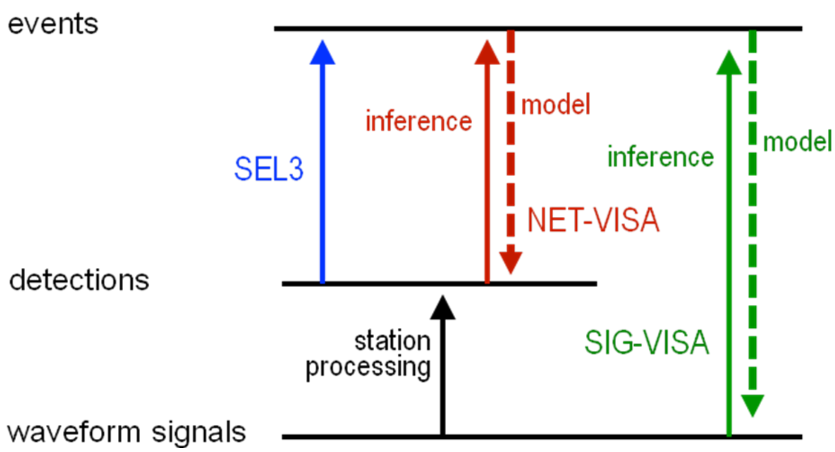

Figure 3: SEL3, NET-VISA, and SIG-VISA.

Signal-based Bayesian monitoring has several advantages over detection-based approaches:

- Signal-based Bayesian monitoring is potentially more sensitive to low-SNR signals because it uses Bayesian inference to determine whether or not signal characteristics at *all* stations are consistent with the occurrence of an event at a given location and magnitude. Roughly speaking, signal-based Bayesian monitoring improves over detection-based methods in much the same way that array stations improve over single seismometers, by using multiple sensors to overcome the difficulty of detection in any one sensor.
- Signal-based Bayesian monitoring can also take advantage of correlations in waveform properties across multiple signals, as used in waveform matching algorithms such as double-differencing (Waldhauser and Ellsworth, 2000; Schaff and Richards, 2004).
- If signals from two events overlap at a given station, a detection-based system may record only a single, large detection, which may be incompatible with the locations and magnitudes of both events. In the signal-based approach, the overlapping signal is expected and is effectively separated into its additive components by the inference process (given time-separated signals from at least one other station). Stone *et al*. (1999) cite a similar advantage for their "unified tracking" approach.

We report here on two aspects of the problem: formulation of spatially correlated waveform models and connections between Bayesian methods and other "joint multi-station methods" such as double-differencing and joint hypocenter determination.

Waveform models: Preliminary proposal

The first research challenge lies in formulating a complete and consistent probability model that supports the computation of the likelihood of observing the measured signal traces, given any hypothesized collection of events. The simplest approach involves a low-dimensional parametric envelope descriptor (cf. Huseby et al., 1998) (e.g., triangular or paired-exponential) whose arrival time, amplitude, and spread depend on the distance and magnitude of the event; an example of a paired-exponential envelope template is shown in Figure 4.

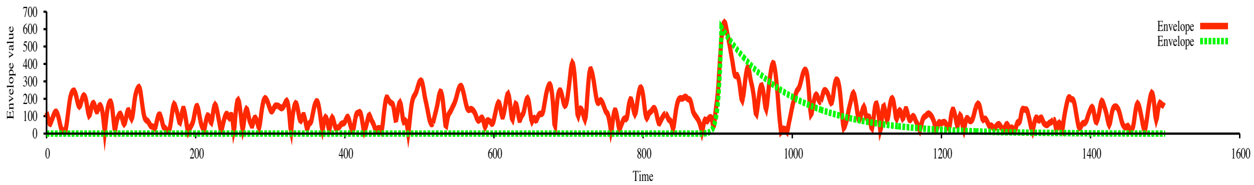

Figure 4: A paired-exponential envelope template (green) superimposed on an actual waveform envelope (red).

Actual envelopes are not well-modeled by "template plus i.i.d. noise"; typically, they exhibit more temporally correlated "macro-variation" resulting from path effects. Furthermore, envelope (or waveform) shape is highly repeatable across events with the same location and type – this is the basis for waveform cross-correlation. To achieve these properties we adopt a hybrid parametric/nonparametric model in which the realized envelope for a given event is the product of a parametric mean envelope (magnitude and distance dependent) and a stochastic modulation signal. A simple modulation model might be based on a random linear combination of Fourier basis functions (see Figure 5). Ultimately we will learn a model from historical data.

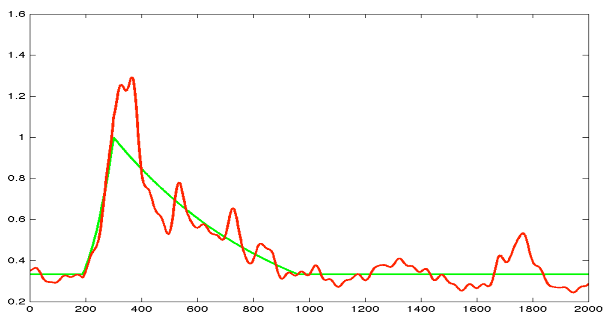

Figure 5: A sampled envelope (red) formed from a parametric template (green) multiplied by a random linear combination of Fourier basis functions.

A basic model of envelope variation would assign each event an independently sampled log-modulation signal, using a Gaussian prior on the basis coefficients. A nonparametric extension adopts a *Gaussian process* model for the basis function parameters. This captures correlations among event envelopes that decay with distance (analogous to *correlation matching*); an illustrative example is shown in Figure 6.

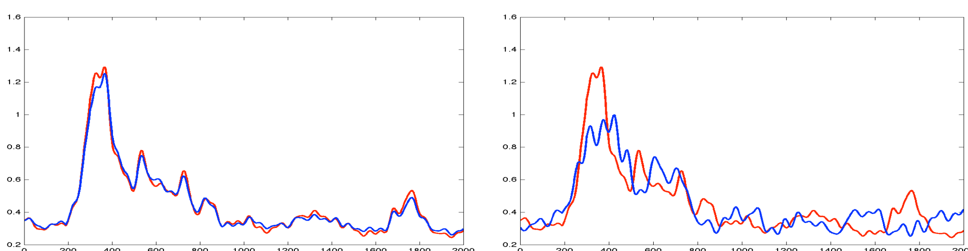

Figure 6: (a) A reference envelope (red) and a sampled envelope (blue) from the Gaussian-process generative model for a nearby event location. (b) The same reference envelope (red) and a sampled envelope (blue) from a distant event location.

Bayesian localization, joint hypocenter determination, and double-differencing

Accurately inverting event hypocenters is made difficult by velocity heterogeneity, which introduces correlations to the travel-time residuals. This issue has been addressed by the simultaneous inversion of hypocenters with station-specific corrections (e.g., Douglas, 1967) or with velocity structure (e.g., Thurber, 1983; Zhang & Thurber, 2003), and by difference-based methods such as double-differencing (Waldhauser and Ellsworth, 2000). The latter approach has the advantage of naturally integrating differential arrival time data obtained through waveform matching techniques such as cross-correlation; differences obtained in this way are more precise than those obtained by subtracting picked arrival times and so provide higher-quality location estimates (Schaff and Richards, 2004). We have developed a Bayesian method for the joint inversion of event hypocenters that accounts for correlated travel-time residuals as a fully integrated part of the probabilistic inference. We refer to this method as Bayesian double-differencing (BDD), because it resembles the method of double-differencing both in the sense of correcting for velocity heterogeneity and in its ability to exploit the difference information provided by waveform matching to achieve higher-quality localizations. Unlike double-differencing and related techniques, we solve for absolute as well as relative locations.

Model: Our approach models the slowness (inverse velocity) field as a Gaussian process (GP). In a GP, values at any pair of points are jointly Gaussian, with a covariance kernel that typically depends on the distance between the points. In standard GP regression (familiar to geophysicists as *kriging*), the posterior random field is inferred based on observations at a finite set of points; in our case, however, we do not observe the slowness values directly, and our goal is not to estimate the slowness field itself (although a posterior estimate does emerge naturally from the probabilistic inference process). Instead, we make use of the fact that a GP slowness field induces a GP residual field at each station, with the covariance kernel of each residual field given by a double integral along event-station paths of the slowness field covariance kernel. Because these station-specific residual fields are all induced by the same underlying GP slowness field, the travel-time residuals themselves are jointly Gaussian across stations. (Note that unlike difference methods which consider the relationships between event residuals independently at each station, our model shares information naturally between nearby stations.)

The joint distribution on the residuals given the event locations (which is equivalently a distribution on the arrival times themselves) constitutes a forward model, which we combine with a prior distribution on event locations and invert using Bayes' rule to yield a posterior distribution on event locations conditioned on arrival times. Of course, in reality we do not observe the true arrival times; we must extend the model to reflect sensor noise in the form of "pick errors." This is accomplished by an additional Gaussian noise term that is added independently for each observation as part of the forward model.

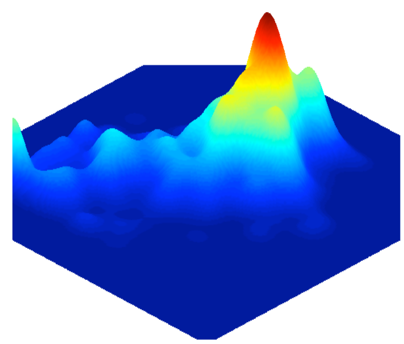

Figure 7: Example posterior marginal distribution for a single event location.

It is similarly possible to incorporate arrival-time differences obtained by waveform matching: these differences are represented in the forward model as the differences in the true arrival times, plus a Gaussian noise term (again independent for each difference observation) with lower variance than that of the pick error to reflect the higher precision attained by waveform matching. The Bayesian inversion seamlessly makes use of both the absolute time constraints provided by picked arrival times along with the relative time constraints of the difference observations when available; the lower variance ascribed to difference observations causes them to be treated as stronger constraints, analogously to the higher weighting commonly given to these constraints in double-difference methods.

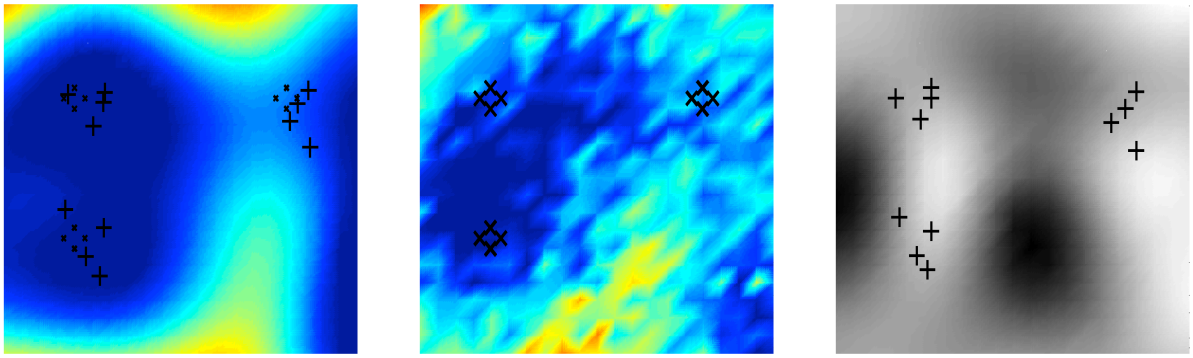

Figure 8: (a) Simulated event clusters located using BDD (+) along with their true locations (x). Stations are positioned at the four corners and each event is observed by three of the four stations. The background shows the MLE slowness field given the inferred locations. (b) The true slowness field, with true event locations marked. (c) Variance of the slowness field posterior (darker is higher).

Inference: Unfortunately, although the forward distribution over the travel-time residuals is Gaussian for any fixed set of event locations, the posterior distribution over the event locations themselves has no simple analytic form (Figure 7 shows an example), so the inversion is nontrivial to perform. We instead apply Markov Chain Monte Carlo (MCMC), a standard technique for Bayesian inference, to sample from the posterior and estimate the expected locations of the joint set of events. This is time-consuming but tractable for the small-scale simulations we have considered thus far.

Although the goal of the inference is to produce accurate locations, the model also defines a posterior distribution over the slowness field, which is simple to calculate given a point estimate of the event locations. Figures 8(a) and 8(b) compare a slowness field inferred from 12 events to the true slowness field, showing that our method effectively recovers the tomography while inverting for event locations. In this sense it resembles other tomographic approaches (e.g. Zhang & Thurber, 2003), but maintains a full posterior distribution over the tomography instead of a point estimate. As illustration, Figure 8(c) shows the variance of the slowness field posterior; note that we are most confident of the tomography along ray paths (straight lines in this simulation) between events and nearby stations (located at the four corners).

Experiments: We have conducted initial experiments on simulated data, with events clustered along a fixed "fault line" and with each event detected by three of four stations with arrival times sampled from the Gaussian process prior. These experiments did not make use of waveform matching. Several distance-weighting strategies for double-differencing were evaluated, including the one described by Waldhauser & Ellsworth (2000); the results presented in Figure 9 are from the best-performing strategy, which sets the weight of each difference residual constraint equal to the prior covariance of the two corresponding residuals in the Bayesian model. As Figure 9 shows, the Bayesian method significantly outperformed both double-differencing as well as a simple independent inversion of the event locations, both in the absolute location error and in the relative locations between events.

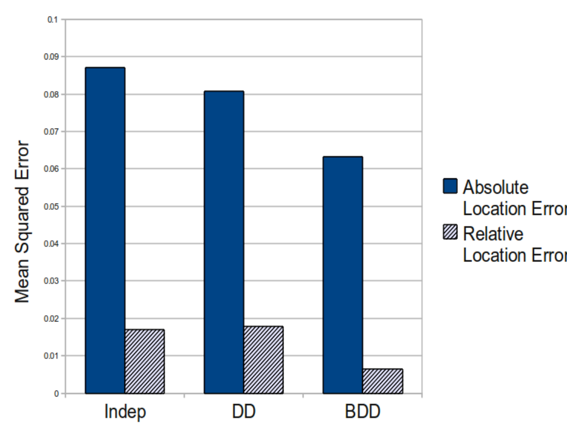

Figure 9: Mean squared error in absolute and relative event locations, over 22 simulation runs with 5 events in each run. Relative error is the difference between true and inferred distance for any pair of events.

CONCLUSIONS AND RECOMMENDATIONS

Prior results suggest that Bayesian monitoring is a promising technique for analyzing streams of parametric detections from multiple stations to form a global event bulletin and may be preferable to existing deployed methods for global association. Our current work, in its very early stages, is aimed at extending Bayesian monitoring with generative models of waveform envelopes to improve detection and association and performing joint inference of event locations and slowness fields to improve localization accuracy.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of DTRA for this work; also the support of CTBTO for the work on NET-VISA and the technical cooperation of CTBTO personnel including Ronan Le Bras and Jeff Given. We thank Sheila Vaidya of LLNL and Ola Dahlman, former Chair of CTBTO Working Group B, for their moral support and encouragement in this project.

REFERENCES

Antolik, M., G. Ekstrom, and A. Dziewonski (2001). Global event location with full and sparse datasets using three-dimensional models of mantle *P*-wave velocity, *Pure Appl. Geophys.* **158,** 291–317.

Arora, N. S., S. J. Russell, P. Kidwell, and E. Sudderth (2011b). Global seismic monitoring: A Bayesian approach. In *Proc. AAAI-11*, San Francisco.

Arora, N. S., S. J. Russell, P. Kidwell, and E. Sudderth (2011a). Global seismic monitoring as probabilistic inference. In *Advances in Neural Information Processing Systems 23*, MIT Press.

Arora, N., Russell, S., and Sudderth, E. (2009). Vertically Integrated Seismic Analysis II: Inference. *Eos Transactions of the American Geophysical Union*, 90(52), Fall Meeting Supplement, Abstract S31B-1713.

Arora, N., Russell, S., de Salvo Braz, R., and Sudderth, E. (2010b). Gibbs sampling in open-universe stochastic languages. In *Proc. 26th Conference on Uncertainty in Artificial Intelligence (UAI 2010)*, Catalina Island, California.

Bell, K. L., Ephraim, Y., and Van Trees, H. L. (2000). A Bayesian Approach to Robust Adaptive Beamforming. *IEEE Transactions on Signal Processing*, 48(2),

Candy, J. V. (2009). *Bayesian Signal Processing: Classical, Modern, and Particle Filtering Methods*. Wiley Interscience.

Dodge, D.A., and K.F. Sprenke (1992). Improvements in mining induced microseismic source locations at the Lucky Friday mining using an automated whole-wave-form analysis system, *Pure and Applied Geophysics*, 139, 609-626.

Douglas, A. (1967). Joint epicentre determination. *Nature*, *215*, 47–48.

Duijndam, A. J. W. (1988a). Bayesian estimation in seismic inversion—Part I: Principles. *Geophys. Prosp*., 36, 878–898.

Duijndam, A. J. W. (1988b). Bayesian estimation in seismic inversion—Part II: Uncertainty analysis. *Geophys. Prosp*., 36, 899–918.

Flanagan, M.F., S.C. Myers, K.D. Koper (2007). Regional travel-time uncertainty and seismic location improvement using a 3Dimensional *a priori* velocity model, *Bull. Seismol. Soc. Am*., **97**, 804-825.

Gibbons, S. J. and Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophys. J. Int.*, 165, 149 – 166.

Harris, D.B (1991). A wave-form correlation method for identifying quarry explosions, *Bull. Seismol. Soc. Am*., **81**, 2395-2418.

McNamara, D. E., Hutt, C. R., Gee, L. S., Benz, H. M., and Buland, R. P. (2009). A Method to Establish Seismic Noise Baselines for Automated Station Assessment. *Seismological Research Letters*, 80(4), 628-637.

Milch, B., Marthi, M., Sontag, D. Russell, S., Ong, D. L., and Kolobov., A. (2005). BLOG: Probabilistic Models with Unknown Objects. In *Proc. IJCAI-05*, Edinburgh, Scotland.

Myers, S.C., G. Johannesson, and W. Hanley (2007). A Bayesian hierarchical method for multiple-event seismic location. *Geophys. J. Int.*, 171, 1049-1063.

Myers, S.C., G. Johannesson, and W. Hanley (2009). Incorporation of probabilistic seismic phase labels into a Bayesian multiple-event seismic locator. *Geophys. J. Int*., 177, 193-204.

Oh, S., Russell, S., and Sastry, S. S. (2009). Markov Chain Monte Carlo Data Association for Multi-Target Tracking. *IEEE Transactions on Automatic Control*, 54(3), 481-497.

Ó Ruanaidh, J. J. K. and Fitzgerald, W. J. (1995). *Numerical Bayesian Methods Applied to Signal Processing.* Springer.

Peterson, J. (1993). *Observation and Modeling of Seismic Background Noise*. USGS Open-File Report 93-322.

Pinsky, V. (2006). Using beamforming for the global network location. *Physics of The Earth and Planetary Interiors*, 158(1), 75-83.

Russell, S., Vaidya, S., and Le Bras, R. (2010). Machine Learning for Comprehensive Nuclear-Test-Ban Treaty Monitoring. *CTBTO Spectrum*, 14, 32-35.

Russell, S., Arora, N., Jordan, M., and Sudderth, E. (2009). Vertically Integrated Seismological Analysis I: Modeling. *Eos Transactions of the American Geophysical Union*, 90(52), Fall Meeting Supplement, Abstract S33D-08.

Schaff D.P., and P.G. Richards (2004). Lg-wave cross correlation and double-difference location: Application to the 1999 Xiuyan, China, sequence, *Bull. Seismol. Soc. Am*., **94**, 867-879.

Schultz , C.A., S.C. Myers, J. Hipp, and C. Young (1998). Nonstationary Bayesian kriging: a predictive technique to generate corrections for detection, location, and discrimination, *Bull. Seismol. Soc. Am.,* **88**, 1275-1288.

Simmons, N.A., A.M. Forte, L. Boschi and S.P. Grand (2010a), GyPSuM: A joint tomographic model of mantle density and seismic wave speeds, J. Geophys. Res. (under review).

Simmons, N.A., S.C. Myers, and G. Johannesson (2010b). Global-scale P-wave tomography designed for accurate prediction of regional and teleseismic travel times for Middle East events, *Geophys. J. Int.*, In preparation.

Stone, L. D., Corwin, T. L., and Barlow, C. A. (1999). *Bayesian Multiple Target Tracking*. Artech House.

Taylor, S. R., Velasco, A., Yang, X., Maceria, M., and Phillips, W. S. (2001). Bayesian tomography applied to seismic event identification problems. In *23rd Seismic Research Review*: *Worldwide Monitoring of Nuclear Explosions*.

Thurber, C. H. (1983). Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California. *Journal of Geophysical Research*, *88*(B10), 8226–8236. American Geophysical Union.

Waldhauser, F. and Ellsworth, W. L. (2000). A Double-Difference Earthquake Location Algorithm: Method and Application to the Northern Hayward Fault, California. *Bull. Seism. Soc. Amer.*, 90(6), 1353-1368.

Yang, X., Bondar, I., Bhattacharyya, J., Ritzwoller, M., Shapiro, N., Antolik, M., Ekström, G., Israelsson, H., and McLaughlin, K. (2004). Validation of Regional and Teleseismic Travel-Time Models by Relocating Ground-Truth Events. *Bull. Seism. Soc. Amer*., **94**, 897–919.

Zhang, H., & Thurber, C. H. (2003). Double-difference tomography: The method and its application to the Hayward fault, California. *Bulletin of the Seismological Society of America*, *93*(5), 1875.