PROGRESS IN SIGNAL-BASED BAYESIAN MONITORING

David A. Moore[1], Kevin M. Mayeda[1], Steve M. Myers[2], Min Joon Seo[1], and Stuart J. Russell[1]

University of California at Berkeley[1] and Lawrence Livermore National Laboratory[2]

Sponsored by the Defense Threat Reduction Agency and the Comprehensive Nuclear-Test-Ban Treaty Organization

Award Nos. HDTRA1-11-1-0026 and 2011-2000

ABSTRACT

Localization tools such as the CTBTO's GA (Global Association) system work by analyzing detections of arriving phases; these detections constitute a discretized, thresholded summary of information from the underlying seismic waveforms. By contrast, our SIG-VISA (SIGnal-based Vertically Integrated Seismic Analysis) system operates directly at the level of the raw waveforms or envelopes, applying Bayesian inference to a generative probabilistic model of seismic traces to search for the event bulletin having the highest posterior probability given the observed signals. The signal-based approach has the potential for improved sensitivity and localization performance compared to purely detection-based methods.

The SIG-VISA generative envelope model views the observed signal envelope as the composition of a background noise process plus a set of arriving phase envelopes; in its simplest form, each phase envelope consists of a parameterized template perturbed by an autoregressive process. We build station-specific models for the template shape parameters, including amplitude and coda decay rates across multiple frequency bands, as well as the parameters of the noise and signal perturbation processes. The signal-based model leads to increased sensitivity compared to a purely detection-based model, since by comparing the observation likelihoods under the signal and noise models we can extract statistical evidence from sub-threshold arrivals (or their absence) which would otherwise be ignored. This capability is especially valuable for faintly detected (low-magnitude and/or teleseismic) events.

A further advance of SIG-VISA is the incorporation of *nonparametric* modeling using Gaussian processes (GPs), a class of statistical models closely related to the kriging method in geophysics. GPs provide a probabilistic framework for predicting the attributes of future events based on past events in similar or nearby locations. SIG-VISA makes use of GP models for the template shape parameters, improving localization performance relative to simpler parametric models of the types used in previous systems (e.g., NET-VISA). Moreover, we provide preliminary results from a nonparametric model for the template perturbations that captures the phenomenon of correlated waveforms from nearby events; this allows a waveform matching effect to fall out naturally from the probabilistic inference.

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 is developing 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

This report presents a fully specified generative model for seismic waveform envelopes observed across a network, implemented in a prototype version of the SIG-VISA system. We describe the components of the model, demonstrate the process of learning model parameters, and demonstrate localization of events using data from the CTBTO's International Data Centre (IDC).

Background: Signal-based Bayesian monitoring

Seismic monitoring can be described in the language of statistics as an *inference* problem: given as input the waveform data recorded by a network of monitoring stations over a period of time, the goal is to infer an event bulletin, i.e., a list of seismic events hypothesized to have occurred during that period, identified by their locations and estimated magnitudes. This is sometimes referred to as an *inverse* problem, since it requires reasoning backwards from the observations to the events that produced them. Bayesian statistics provides a rigorous mathematical framework for the solution of inverse problems, and has been well-established throughout the sciences including seismology (Duijndam, 1988a,b), having been applied successfully in areas including tomography (Taylor et al., 2001; Simmons et al., 2010b) and event localization (Myers et al., 2007, 2009). Recently, the NET-VISA system (Russell et al., 2009, 2010; Arora et al., 2011a,b) has demonstrated what is to our knowledge the first complete Bayesian monitoring system, yielding more accurate locations and fewer missed events when compared to traditional systems such as IDC's Global Association (GA) system.

A Bayesian approach to inference involves two components: a *prior probability distribution* $\Pr(X = x)$ over a set of hypotheses X, and a *likelihood* or *conditional probability model* $\Pr(Y = y \mid X = x)$ which gives the probability of observing the evidence y under the given hypothesis x. The prior and the likelihood together comprise a *generative model* specifying a complete joint distribution over hypotheses and observations, as well as any additional latent variables introduced by the model. In seismic monitoring, the hypotheses are the event bulletins, the evidence consists of the raw waveforms recorded by the network, and the conditional probability model (sometimes called the *forward* 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 *posterior probability distribution* allows us to reason backwards from the evidence (observed waveforms) to hypotheses (event bulletins), and is given by *Bayes' rule*:

$$Pr(X \mid Y = y) = \alpha \, Pr(Y = y \mid X) \Pr(x)$$

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. Computing the posterior distribution (or a point estimate of it, e.g., the single most probable hypothesis) is generally nontrivial and often requires sophisticated search or sampling algorithms.

As a simplifying assumption, current seismic monitoring systems generally include a concept of *detections*; standard 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). The

set of detections thus functions as an intermediate representation of the raw waveforms recorded throughout the network, yielding a simpler inference problem at the cost of throwing away potentially valuable information. By contrast, the signal-based approach developed in SIG-VISA models the waveform traces directly, reducing information loss and yielding a number of potential advantages:

- Detection-based systems face a tradeoff in setting the detection threshold: too low and the system is swamped with false detections, too high and it could miss genuine events. Modeling the signal directly eliminates the need for such a threshold and allows SIG-VISA to incorporate arrivals of all strengths as statistical evidence; thus, a large number of very weak (sub-threshold) arrivals might have the same inferential power as a smaller number of stronger arrivals. Roughly speaking, signal-based Bayesian monitoring improves over detection-based methods in much the same way that array stations improve over single seismometers, by aggregating signals recorded at multiple sensors to increase SNR.
- When signals from two events overlap at a station, a detection-based system might record only a single large detection incompatible with the locations and magnitudes of both events. By contrast, the signal-based approach simply models the observed waveform signal and has the ability to assign a fraction of the received energy at each moment to each event, in proportions determined by the inference process.
- Finally, signal-based approaches have the ability to take into account more detailed information regarding the shape and characteristics of the observed waveforms. This can involve modeling high-level characteristics such as the decay rate of the seismic coda, which can in some cases be predicted quite accurately from an event's location (Mayeda et al., 2003), as well as lower-level details including the fluctuations of the waveform itself, which are known to be correlated for events in nearby locations; folding this information into the probabilistic model allows the inference process to achieve some of the benefits of waveform matching algorithms (Thorbjanrardottir et al., 1987; Harris, 1991; Withers et al., 1999, Waldhauser and Ellsworth, 2000; Schaff and Richards, 2004).

In this paper we focus primarily on modeling the shape and characteristics of observed signal envelopes, describing a generative probabilistic model for signal envelopes that includes a location-dependent template model learned from historical data, as well an extension to this model that implements an additional location-specific component, modeling shorter-term fluctuations determined by path effects.

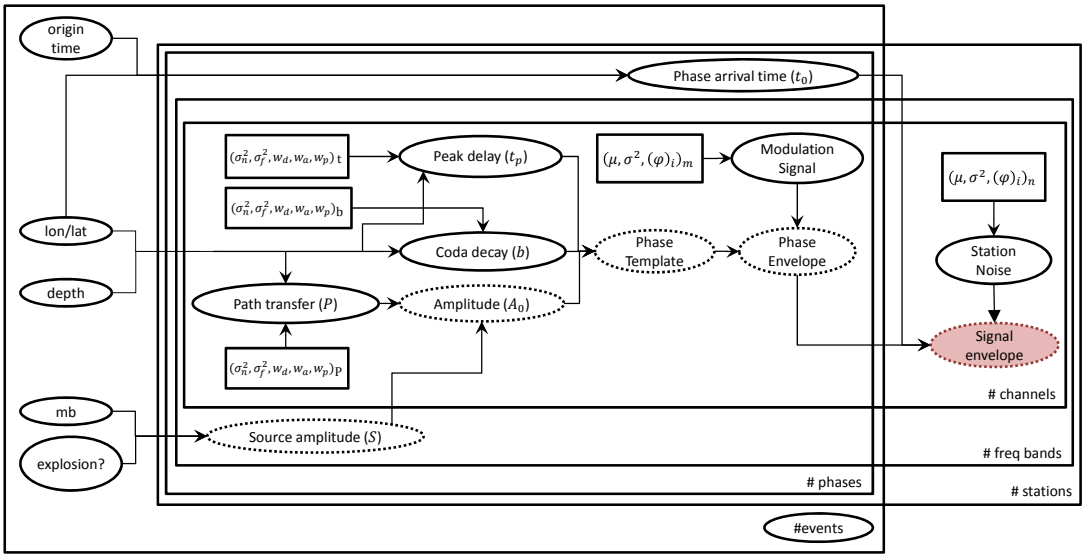

Figure 1. The SIG-VISA generative envelope model. The graph shows dependence relationships among variables in the model, beginning with the event attributes on the left and ending on the right with the observed signal envelopes, in all channels and frequency bands at all stations in the network. The model specifies a conditional probability distribution for each variable given its parents; variables shown with dotted borders are deterministic functions of their parents. Parameters for the Gaussian and autoregressive process components are indicated in square boxes. The "plate" notation is standard for graphical models and should be interpreted as creating a number of independent copies of each variable, e.g., each event has an origin time sampled independently of the other events.

The SIG-VISA Model

The prototype SIG-VISA implementation takes as input the waveforms observed by a network of three-component stations (an extension to array stations is planned as future work). Rather than modeling the raw waveforms directly, we instead model their envelopes within a range of narrow frequency bands, from 0.1-0.2Hz up to 6-8Hz. The envelope representation removes phase information while retaining the level of energy arriving over time; using multiple frequency bands allows us to more accurately model frequency-dependent characteristics of the signal such as the coda decay rate (Mayeda et al, 2003) and spectral content.

Figure 1 shows a graphical view of the SIG-VISA generative model. The model can be understood intuitively as telling a generative story, i.e., a series of steps to generate a sample observation:

1. First, the number of events in the world is sampled from a prior distribution.
2. Characteristics of each event are then sampled independently from prior distributions, including origin time, location, depth, body wave magnitude, and whether the event is a natural event or an explosion. Note that the "explosion" attribute is not intended to be a determinative feature of the system output, but is used internally to allow the model to fit events whose source spectra are not consistent with the predicted spectrum of a natural event.
3. Each event generates a P and S arrival at each site. (Other phase types are supported by the modeling framework but are not currently implemented.) The arrival time of each phase depends on the origin time of the event as well as the event's location relative to the station.
4. Each arriving phase generates a phase envelope, which is taken to be the product of a parameterized template with a stochastic modulation signal. The template captures the general pattern of arriving energy and subsequent decay, while the modulation signal corresponds to shorter-term, less predictable fluctuations (though these fluctuations can sometimes be predicted due to waveform correlations between nearby events, as described below).
5. Finally, the observed envelope for each channel in each frequency band at each station is modeled as the sum of all the arriving phase envelopes, plus a background noise process. Figure 2 shows an example.

The first three steps are essentially the same in both the SIG-VISA and NET-VISA models; Arora et al. (2011a,b) provide details on the event prior distributions and phase arrival time model. Here we focus on the later steps of the process, including the template model and its parameterization as well as models for the modulation signal and station noise processes.

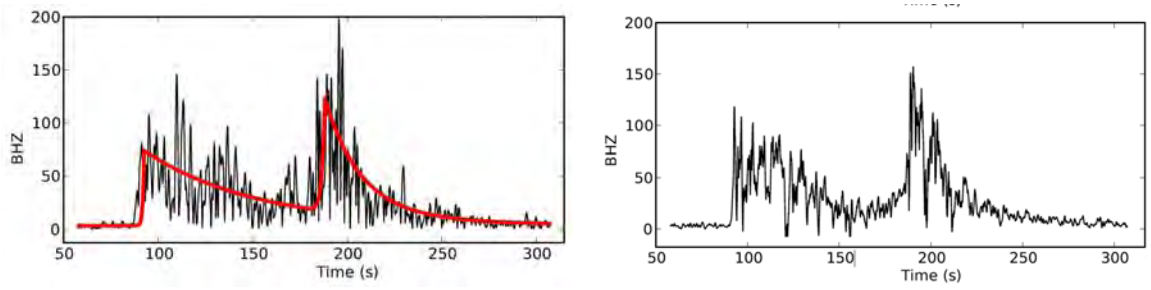

Figure 2. (a) An observed envelope at a station (IDC evid 5305385 at AFI, 2-3Hz), showing the P and S arrivals and subsequent coda decays. The red line indicates the template fit. (b) A synthetic envelope generated from the template and an autoregressive modulation process.

Template Model

The Bayesian framework allows wide flexibility in the choice and parameterization of a template model. Currently we model each phase envelope using a simple paired-exponential template (cf. Huseby et al, 1998) with three parameters: the peak delay t_p(in seconds, the time from the first particle motiont_0 until the envelope reaches its peak), the amplitude/peak height A_0, and the decay rate b. The decaying portion of the paired-exponential template models the seismic coda, and corresponds to a special case (with γ=0 fixed) of the coda decay model introduced by Aki (1969) and used in subsequent coda-modeling work, by, e.g., Mayeda et al. (2003):

$$A(f,d,a,t) = A_0\,(f,d,a) \cdot (t - (t_0 + t_p))^{\gamma} \cdot \exp\,(b(f,d,a) \cdot (t - (t_0 + t_p)))$$

This expression indicates that the template parameters A_0 and b are dependent on the frequency band *f* as well as the station-event distance *d* and azimuth *a* (our model also allows t_p to depend on $f, d, and\ a$, but in practice we have not found any significant empirical relationship between these quantities) . The time in seconds elapsed from the template peak is given by $t - (t_0 + t_p)$. The template amplitude $A_0\ (f, d, a) = S(f) \cdot P(f, d, a)$ is a function of the source amplitude *S*, as well as a path-dependent amplitude correction term *P* that encapsulates the site response, transfer function, and geometrical spreading/attenuation terms used by Mayeda et al. (2003). Models for the source amplitude and path correction terms are described in the sections that follow.

Source Spectra

The source amplitude *S* is computed deterministically as a function of frequency using a source spectrum model. Since natural events and explosions have different spectral signatures, we use different models for the two event types; the inference algorithm considers both hypotheses for each event. For explosion spectra we plan to implement the Mueller–Murphy model (Mueller and Murphy, 1971), whereas for events hypothesized to be natural we currently use a version of the Brune source model (Brune, 1970):

$$S(f) = \frac{FM_0}{\left(1+\frac{f}{f_0}\right)^2}$$

$$f_0 = K\beta^3 \sqrt{\frac{16\Delta\sigma}{7M_0}}$$

where f_0 is a corner frequency calculated following Myers et al (1998), assuming a stress drop ($\Delta\sigma$) of 10 bars for shallow events, S wave velocity of 4.5m/s, and with the constant K set to 1.5 for P waves and 0.33 for S waves. The constant *F* is not set a priori, but is instead learned empirically as part of the path-specific correction *P(f,d,a)* described below. Finally, the seismic moment M_0 is estimated from the IDC body wave magnitude (m_b) using the following relationships:

$$M\_w = 1.46 m\_b\ 0 - 2.42 \qquad \text{(Sipkin, 2003)}$$

$$M\_0 = \exp\left(1.5 Mw + 9.1\right) \qquad \text{(Kanamori, 1977; Hanks and Kanamori, 1979)}$$

Nonparametric Models of Template Parameters

Since SIG-VISA's templates are defined by three parameters, training the template model involves learning models for each of those parameters conditioned on the event properties (note that in the case of the amplitude A_0 this reduces to learning the path-specific corrections $P(f, d, a)$). In particular, we learn these models using Gaussian process regression (Rasmussen and Williams, 2006), also known as kriging. Since GP regression is a fully probabilistic technique — i.e., it computes a Bayesian posterior distribution on the regression function — it delivers a fully quantified estimate of uncertainty along with each prediction. Thus a GP regression model can produce highly confident predictions in areas where historical data are abundant, while still providing its best guess (and admitting its uncertainty) in areas where no data are available.

The extent to which a training example in a particular location influences the predictions in a different location is determined by the *covariance kernel* of the Gaussian process. The kernel currently used by SIG-VISA represents event locations using cylindrical coordinates centered at the station location; that is, $x = (x\_d, x\_a, x\_p)$ where x_d is the station-to-event distance, x_a is the station-to-event azimuth, and x_p is the event depth. The functional form of the kernel is given by:

$$k\left(x^{(1)}, x^{(2)}\right) = \sigma_f^2 \cdot \exp\left(-\left(\frac{\sqrt[3]{x_d^{(1)}} - \sqrt[3]{x_d^{(2)}}}{w_d} \right)^2 - \left(\frac{d_a\left(x_a^{(1)}, x_a^{(2)}\right)}{w_a} \right)^2 - \left(\frac{\sqrt[3]{x_p^{(1)}} - \sqrt[3]{x_p^{(2)}}}{w_p} \right)^2 \right) + \sigma_n^2 \cdot \delta\left(x^{(1)}, x^{(2)}\right)$$

where $\sigma\_f, \sigma\_n, w\_d, w\_a, w\_p$ are hyperparameters specifying the regression function variance, iid noise variance, and correlation length-scales for distance, azimuth, and depth respectively. Note that the distances and depths are transformed by a cube root; this serves to emphasize the effect of small changes in those coordinates at regional distances and shallow depths, and correspondingly to de-emphasize them for teleseismic or deep events. The azimuth-comparison function d_a computes the difference in degrees between $x_a^{(1)}$and $x_a^{(2)}$, and then scales the result by (the cube root of) the mean distance of the two events, $\sqrt[3]{\frac{x_d^{(1)} + x_d^{(2)}}{2}}$. This gives greater weight to differences in azimuth between teleseismic events, corresponding to the greater physical distances between such events.

We construct the training sets for the GP models using the iterative process described in the section on "Learning" below. Given a training set, the GP hyperparameters are learned via maximum likelihood on the training set and then used to train the full GP model (Rasmussen and Williams, 2006). Figure 3 evaluates the effectiveness of this process, showing that the nonparametric, location-sensitive models output by the GP often perform significantly better than a simple linear trend based on distance from the station, both when predicting the coda decay rate as well as when predicting the observed amplitude of an arrival. Figure 4 shows the relationship between coda decay rates and station-to-event distance and azimuth at a single station; note that despite the lack of clear linear trends in the data with respect to distance or azimuth individually, GP regression was able to perform well at this station. This suggests that the regression is successfully learning the properties of events in specific regions. By way of illustration, Figure 6 shows the predicted coda decay rates, and the uncertainty in those predictions, for locations around the world. Note that the GP regression predictions are most confident in areas where the most training data are available, as one would expect.

Modulation and Noise Models

The simplest form of the SIG-VISA model takes both the template modulation and station noise processes to be autoregressive (AR) processes (Box et al., 1994), parameterized by a mean μ, white noise variance σ^2, and a set of n weights (φ_i), where n is the order of the AR process (a more sophisticated modulation model is described below). The use of an AR model allows these signals to exhibit relatively smooth and coherent fluctuations over time, as opposed to, e.g., an iid Gaussian noise model which would be much more "jumpy" and would introduce a strong dependence on the signal sampling rate. The parameters of the modulation process are learned from historical data at each station for each channel and frequency band, while the AR station noise models are re-learned for every hour of incoming data, to account for potential changes in background noise levels over time. When a signal can be fully observed, as is the case with the station noise process in the absence of any arriving phases, then learning the AR parameters for a process of a given order is straightforward using the Yule–Walker equations. We search for the process order that maximizes the log-likelihood of the data under five-fold cross-validation; in our experiments the optimal order tends to be relatively constant around 17 (with signals sampled at 40Hz) across a range of stations and time periods.

Learning the Model

Although we have discussed straightforward methods for learning both the template parameter models and the modulation signal model (GP regression and the Yule–Walker equations respectively), we are generally not so lucky as to be able observe or train models for these quantities directly. Instead, the envelope that we observe is merely the product of an unknown modulation process with an unknown template, leaving us with some uncertainty about the contributions of each component. To learn models for these unobserved variables, we apply "hard" expectation-maximization (Dempster et al., 1977): beginning with an initial guess for a set of AR parameters describing the modulation process, we find the most likely template for every phase arrival in our training set (that is, we run an optimization algorithm to find the set of template parameters that maximizes the probability of the observed signal under the assumed modulation model parameters). We then use these templates to re-estimate the modulation process parameters, by dividing the observed envelope by the optimized template to extract the modulation signal (which we cut off at some threshold above the noise floor to reduce the potential impact of station noise on the extracted signal), then using the Yule–Walker equations to directly learn a new set of modulation parameters. We also use these templates as a training set for learning the template parameter models. By iterating these steps,
SIG-VISA converges to increasingly accurate estimates of modulation signal parameters as well as the template parameters.

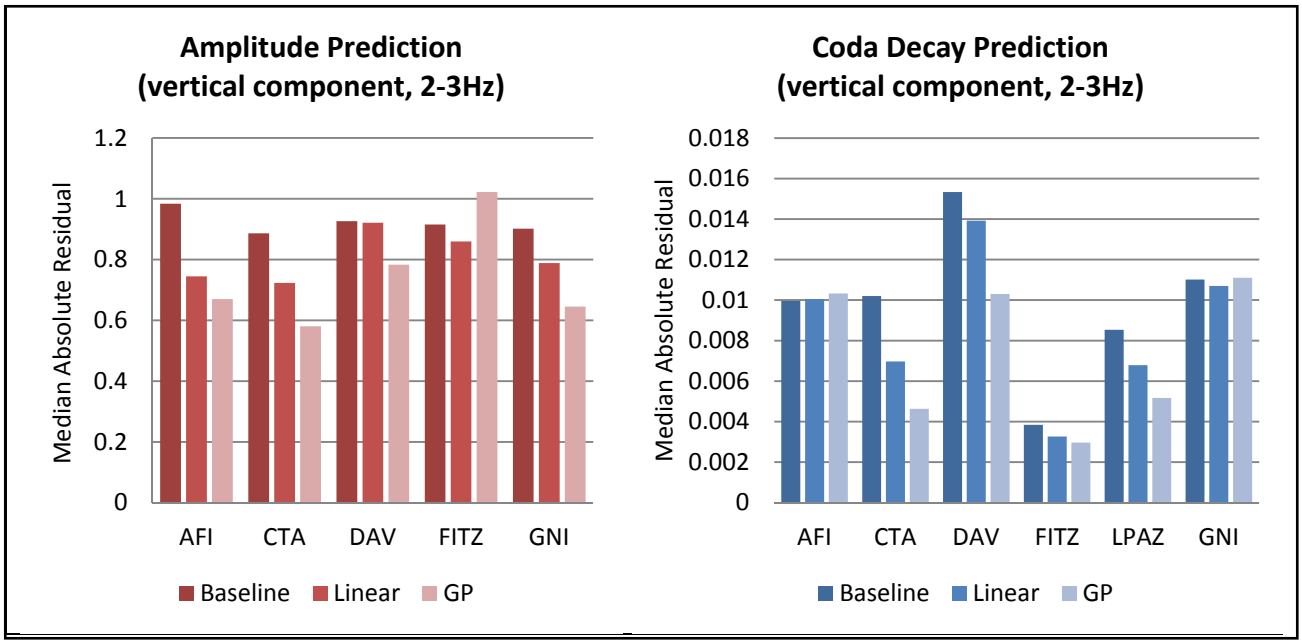

Figure 3. A comparison of models for P wave coda amplitude and decay rates at several IDC stations. Each result is a five-fold cross validation The linear model performs separate linear regressions against the event distance for regional and teleseismic (>2000km distance from station) events, while the baseline always predicts the mean value observed at the station.

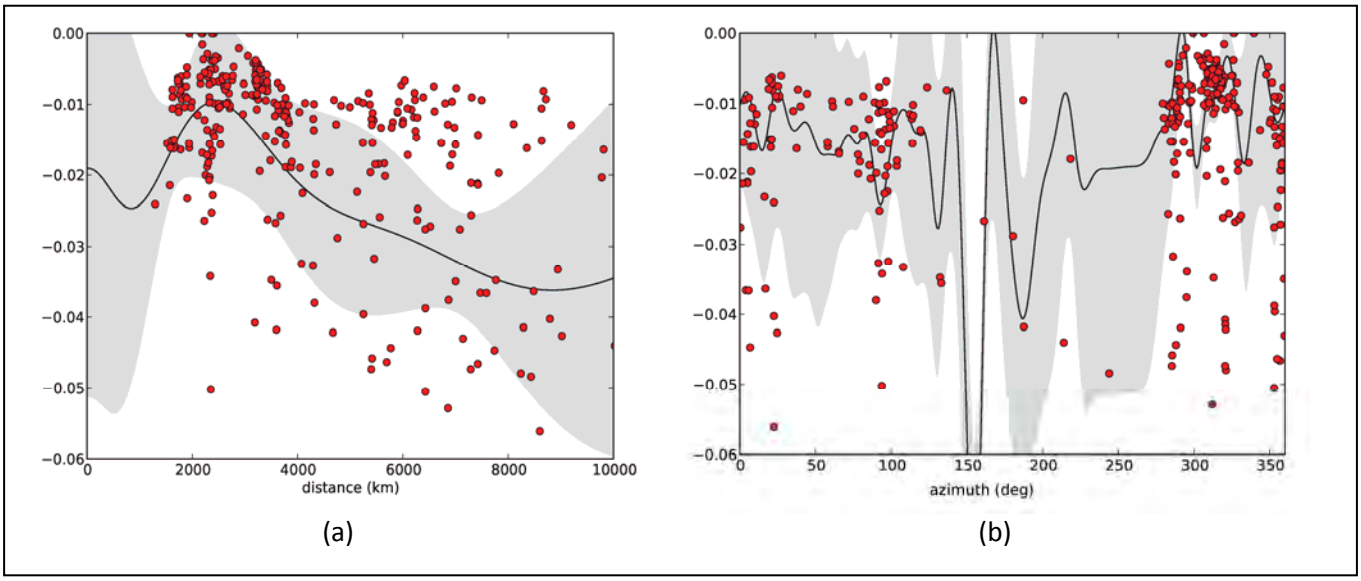

Figure 4. P coda decay rates at site CTA. (a) Gaussian process predictions for events at a range of distances, with station-to-event azimuth and event depth fixed at 0 (note that many of the points far away from the regression line have very different azimuths or depths). (b) Gaussian process predictions for a range of station-to-event azimuths, assuming depth 0 and a distance of 2700km. The shaded area illustrates one standard deviation from the mean (thus about 68% of the probability mass).

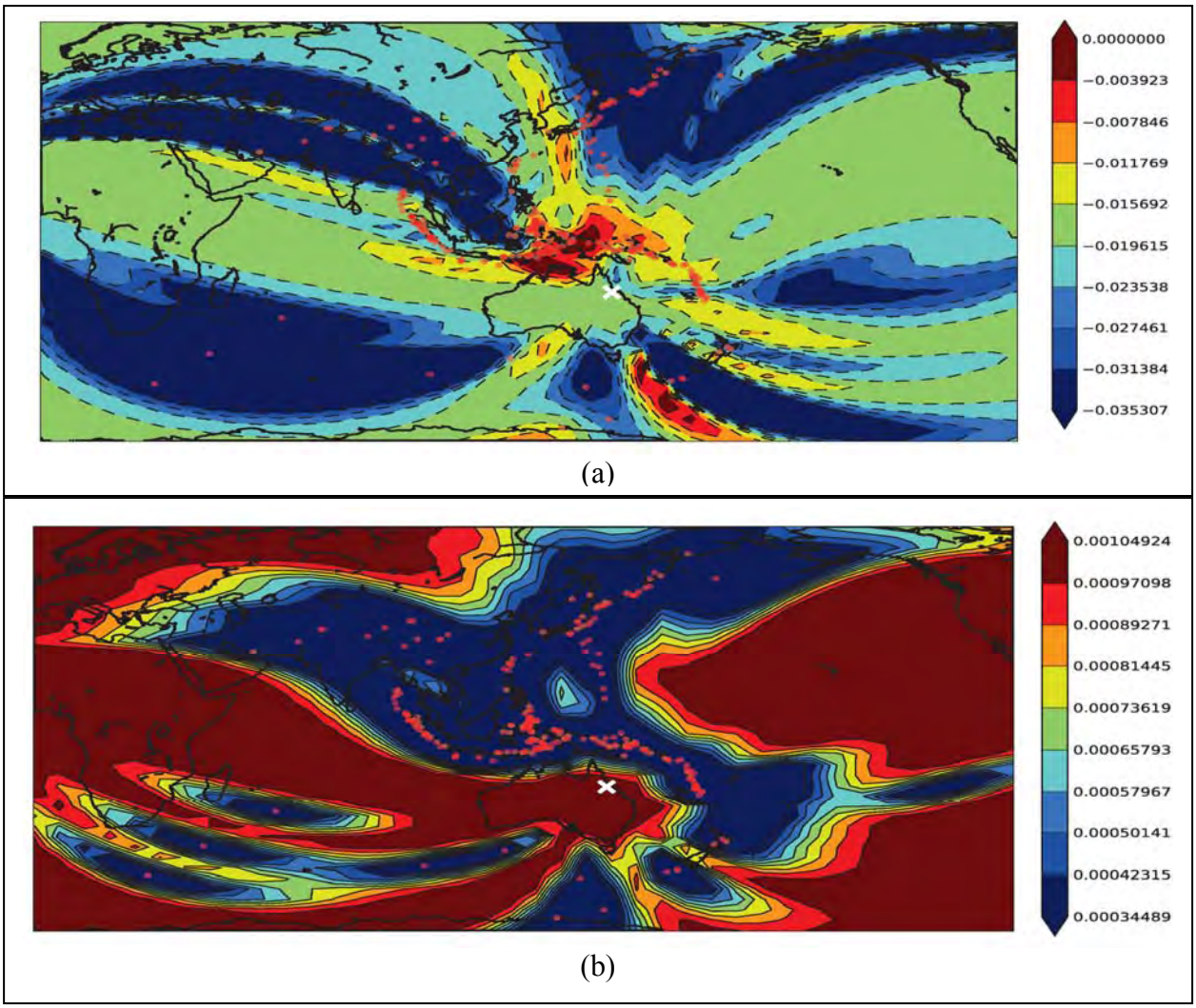

Figure 5. (a) Predicted P wave coda decay for surface events detected at site CTA (site location is labeled with a white 'x'). (b) Variance of P wave coda decay. Both maps show the 313 events from the training set plotted in red.

Inference and Event Localization

Although we have focused primarily on the task of building a generative model, performing efficient inference within such a model is one of the major challenges in signal-based Bayesian monitoring. We plan to take on this challenge directly in future work. Our current prototype implementation uses a grid search over event locations, integrating out the template parameters using Monte Carlo sampling and evaluating the likelihood of the sampled envelopes relative to the observed envelope under the autoregressive noise and modulation models, to compute the density for a single event location using evidence from a small number of stations. An example is shown in Figure 6.

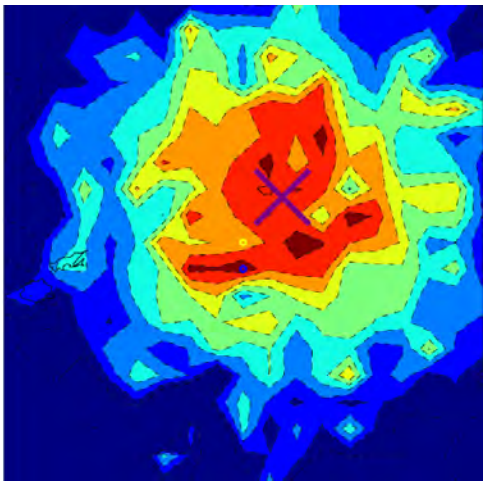

Figure 6. Posterior location density for IDC evid 5373954 using waveforms from sites AFI and URZ. The posterior mode is marked by a blue circle, the LEB location is labeled with a yellow circle, and the purple X indicates the location of station AFI. The width of the plot is 20 degrees, and the distance between the posterior mode and LEB location is approximately 1.2 degrees.

Waveform Matching via the Modulation Model

An interesting extension to the basic SIG-VISA system described above involves replacing the randomly sampled modulation signal with a signal conditioned on the event location. This addition models the established tendency of nearby events to produce highly correlated waveforms due to path-specific scattering, which has been exploited by waveform matching techniques (Thorbjanrardottir et al., 1987; Harris, 1991; Withers et al., 1999, Slinkard et al., 2011). Including such a constraint in our model causes waveform matching behavior to emerge naturally from the probabilistic inference process. Such behavior is especially valuable for a monitoring system dealing with aftershock sequences consisting of a large number of events occurring within a small area. More generally, waveform matching improves localization and helps in avoiding association errors, thus improving overall monitoring performance.

To obtain the desired functionality, we need a generative model for the modulated signal that produces similar modulations when two events are nearby. In an ideal world, events in the *same* location with the *same* source mechanism would produce *identical* modulated signals at a given station; hence, one could consider learning a deterministic spatial model of modulations. Absent any information about the source mechanism, we should still allow for variations in modulation even from events in the same location. One simple mechanism which implements this functionality is to model the modulation signal as an autoregressive process as above, but with a spatial prior distribution over the AR process parameters (e.g., a Gaussian process learned from historical data, similar to the use of GP regression for template parameters described above) instead of assuming a single constant set of parameters for all events observed at a station. Since the parameters of an AR process determine its spectral characteristics, this technique allows the model to learn the general spectral properties of events occurring in different locations. Figure 7(a) shows the power spectral density of three event arrivals, two from nearby locations and one further away, suggesting that such spectral properties have the potential to distinguish event locations. The spectra from the two "close" events (18km apart) are strikingly similar. By way of demonstration, Figure 7(b) plots the posterior location density for IDC event id 4689462, based on the waveform detected at a single station (MK31) and using a GP regression model trained on the parameters of an order-5 AR model for the raw waveforms. (In our preliminary experiments we have not yet had the chance to test this approach for modeling the modulation signal in the context of the template model described above, but expect those experiments to yield similar results.) Note that the figure shows a likelihood peak within the same cluster of events as the true event location.

A second mechanism we are considering is to model the modulation signal as a random linear combination of Fourier or wavelet basis functions, now with a spatial prior distribution (e.g., a Gaussian process) on the basis coefficients. Since this approach models the modulation signal as a deterministic function of the regression outputs (as opposed to the autoregressive approach of the previous paragraph, which introduces additional randomness in the sample path of the AR process), we believe it has the potential to yield sharper location estimates, though our experiments are still in the preliminary stages. Figure 8 shows a pair of synthetic signals generated using this approach.

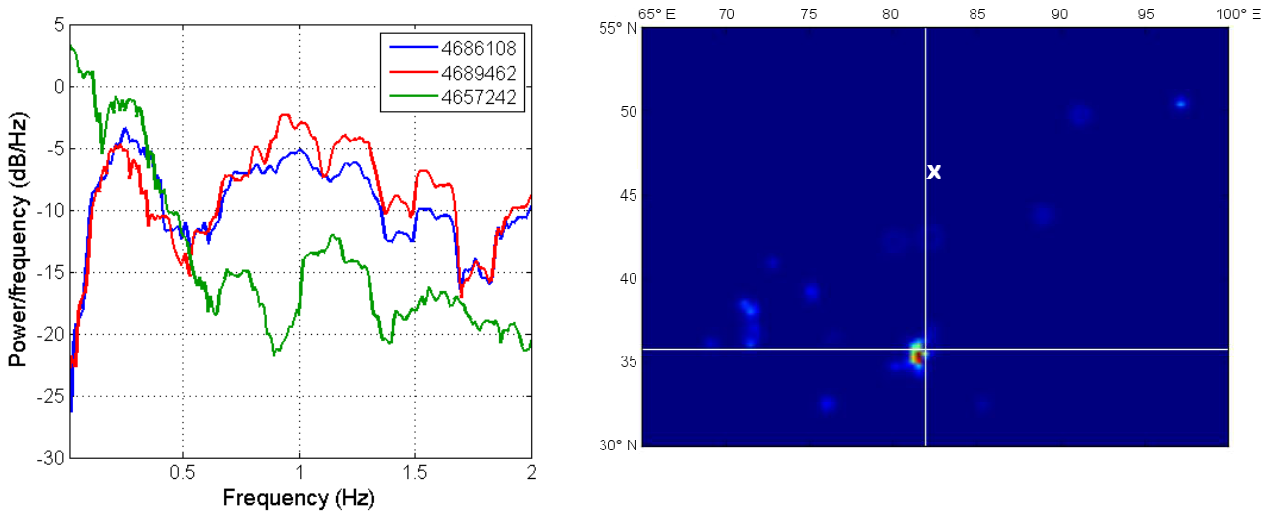

Figure 7. (a) Power spectral densities of raw waveforms (40s after Pn arrival) recorded by the vertical component of IDC station MK31 from the Makanchi array. The events with IDC evids 4689462 and 4686108 are located at distance 18km from each other, and their waveforms have a normalized cross-correlation peak of 0.59. Event 46572742, which is included for contrast, is at a distance of approximately 70km and has a cross-correlation peak of approximately 0.07 vs each of the other two events. (b) Posterior location density for IDC evid 4689462, under a GP regression model for the AR coefficients trained on raw waveforms from 99 other event arrivals during a two-week time period at MK31 (the station location is market with a purple X). The intersection of the two white lines is the true event location.

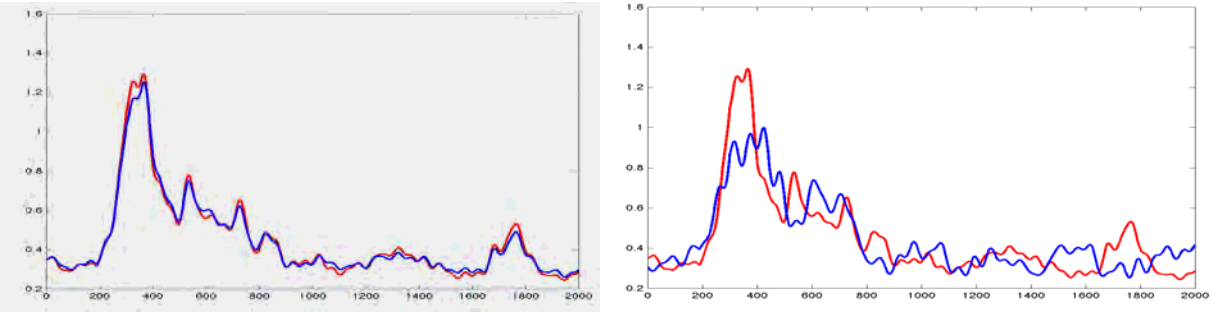

Figure 8. Illustration of a deterministic location-based prior for a synthetic Fourier basis envelope model. (a) A reference envelope (red) and a sampled envelope (blue) from the Gaussian-process generative model for a nearby event location. (b) The same reference envelope (red) and a sampled envelope (blue) from a distant event location. Here, "nearby" and "distant" are relative to the Gaussian process correlation distance.

CONCLUSIONS AND RECOMMENDATIONS

Our work builds upon recent successes in Bayesian monitoring and aims to continue the progress by extending a generative probability model to waveform envelopes. Here we have described such a model, which is implemented in a prototype system. Areas for future work include refinement of the model and an extension to array stations, more complete incorporation of a waveform matching component, and progress towards efficient inference techniques.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of DTRA for this work; as well as the CTBTO for providing data as well as access to their vDEC system where several of these experiments were performed. We also thank Megan Slinkard and Regina Eckert of Sandia National Labs for sharing with us their data on correlated arrival waveforms.

REFERENCES

Aki, K. (1969). Analysis of the seismic coda of local earthquakes as scattered waves. *J. Geophys. Res*. **74**, 615–631.

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

Box, G., G. Jenkins, and G. Reinsel (1994). *Time Series Analysis: Forecasting and Control*. Prentice-Hall.

Brune, J. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes. *J. Geophys. Res.* **75**, 4997–5009.

Dempster, A.P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1), 1–38.

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

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

Hanks, T., and H. Kanamori (1979). A moment magnitude scale. *J. Geophys. Res.* **84**, 2348–2350.

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

Huseby, E.S., Ruud, B.O., and Dainty, A.M. (1998). Robust and reliable epicenter determinations: Envelope processing of local network data. *Bull. Seismol. Soc. Am*. **81**(1), 284–190.

Kanamori, H. (1977). The energy release in great earthquakes. *J. Geophys. Res.* **82**, 2981–2987.

Mayeda, K., A. Hofstetter, J. L.O'Boyle, W. R. Walter (2003). Stable and transportable regional magnitudes based on coda-derived moment-rate spectra. *Bull. Seismol. Soc. Am.* **93**, 224–239.

Myers S.C., S.L. Beck, G. Zandt, and T.C. Wallace (1998). Lithospheric-scale structure across the Bolivian Andes from tomographic images of velocity and attenuation for P and S waves, *J. Geophys. Res*. **103**, 21,233–21,252.

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.

Rasmussen, C. and C. Williams (2006). *Gaussian Processes for Machine Learning*. MIT Press.

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), 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.

Slinkard, Megan E., Dorthe B. Carr, Stephen L. Heck, and Christopher J. Young (2011). Towards an automated waveform correlation detector system, in *Proceedings of the 2011 Monitoring Research Review*: *Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-11-04823, Vol. 1, pp. 392–399.

Sipkin, S.A. (2003). A correction to body-wave magnitude m_b based on moment magnitude M_W. *Seismol. Res. Lett*. **74**(6), 739–742.

Thorbjarnardottir, B. S. and J. C. Pechmann (1987). Constraints on relative earthquake locations from cross-correlation of waveforms. *Bull. Seismol. Soc. Am*. **77**, 1626–1634.

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**(6), 1353–1368.

Withers, M., R. Aster, and C. Young (1999). An automated local and regional seismic event detection and location system using waveform correlation. *Bull. Seism. Soc. Am*. **89**, 657–669.