Open-Universe Probability Models for Signal-Based Nuclear Monitoring

Dave Moore

October 9, 2013

Abstract

This report introduces the problem of nuclear test monitoring, and describes the formulation of a Bayesian model of seismic waveforms along with a prototype implementation of learning and inference in that model. Preliminary results are briefly presented, and challenges discussed, along with a proposed timeline of future research efforts.

1 Background on Nuclear Monitoring

Consider a network of sensors around the world, recording vibrations in the ground, water, and air (seismic, hydroacoustic, and infrasound waves, respectively) as well as concentrations of radioactive particles (radionuclides). The output of this network is the set of observed (continuous) waveforms and radionuclide measurements recorded by the stations. The *nuclear monitoring problem* is to extract, from the network output over a particular window of time, a comprehensive list of nuclear weapons tests that occurred during that time window, specifying for each event a surface location, depth, time, and yield. We present a new approach to nuclear monitoring, situated in the framework of Bayesian statistical inference.

The primary existing monitoring network is the International Monitoring System (IMS), which has been established under the Comprehensive Nuclear Test Ban Treaty (CTBT) along with an International Data Center (IDC) that collects and processes the data from the IMS network. The IMS network, currently 85 percent complete, is planned to include 170 seismic stations, 11 hydroacoustic stations, 60 infrasound stations, and 80 radionuclide stations, distributed roughly uniformly across the globe[18]. Due to the current difficulty of distinguishing between earthquakes and nuclear explosions by automated means, and the sensitive politics of accusing a state of nuclear testing, current processing at the IDC is focused on producing a bulletin listing *all* seismic disturbances, of which any particularly suspicious cases may then be analyzed further by human experts. In addition, current automated systems focus almost exclusively on seismic data, with the other sensor modalities being used primarily when further analysis is desired of a specific event. In keeping with these practices, the work we describe here is primarily focused on the task of analyzing *purely seismic* data to produce a *comprehensive* worldwide bulletin of seismic events, though our approach should be readily extendable to the general problem including non-seismic sensor modalities and the task of specifically distinguishing earthquakes from nuclear explosions.

In the language of physical modeling, nuclear monitoring may be considered an *inverse problem*: given some understanding of the physical, causal processes that generate the waveforms observed from a seismic disturbance, we wish to *invert* those processes to determine, given the waveforms, the set of events that generated them. The monitoring problem itself subsumes many simpler inverse problems that have individually been the focus of considerable research in seismology, such as localizing seismic events given times-of-arrival at multiple stations [7] [21], estimating explosion yields or event magnitudes [13] [17], estimating the earth-structure properties that determine the velocity, attenuation, and distortion of seismic waves [6] [5] [10] [12], and distinguishing earthquakes from explosions [15] [22]. The major challenge of monitoring is to build a system that solves these subproblems simultaneously in a integrated, principled fashion.

1.1 Bayesian Approach

Bayesian probability theory provides a natural framework for the monitoring problem. In the Bayesian approach, we define a prior distribution on the space of possible event bulletins, along with a likelihood function (aka "forward model" or "generative process") that specifies, for any hypothesized set of events, a probability distribution on the space of signals observed at the stations in the network. Applying Bayes' rule, these two components then determine a posterior distribution on event hypotheses, conditioned on the observed data. We may then report the hypothesis with the highest posterior probability density (the MAP hypothesis), or query the posterior more specifically, e.g., "what is the probability that an event occurred yesterday in North Korea?", "conditioned on one such event occurring, what is the posterior density on its magnitude?", and so on.

A major attraction of Bayesian inference is that it provides a principled framework in which to combine information from many sensors, including natural handling of uncertainty (for example, observations from noisy stations are naturally downweighted since the observation model at those stations will be spread over a wider range of possible signals) and missing data (if a particular sensor is offline for some period, we can just choose not to condition on the data from that period). The Bayesian methodology automates many considerations that are often ignored by more ad-hoc approaches, for example, that absence of evidence *is* evidence of absence: if we would expect a hypothesized event to generate a response at a particular station and time, but no such response is observed, then this constitutes evidence against that event. Bayesian models are also *compositional*: using the framework of directed graphical models, we can build up a complex generative model by chaining together simpler models, each of which can make use of specific domain expertise. Thus we can incorporate, for example, existing geophysical models of the travel times of seismic waves, with the promise that if better models arise from future research we will be able to plug them in without needing to replace our global model, or implement an entirely new inference algorithm.

Bayesian approaches to inverse problems in seismology have already demonstrated considerable success, for example in event localization [16] and in detection-based monitoring [1] (more below). There is also a rich literature on Bayesian methods in signal processing [3].

One particular subtlety in the Bayesian approach to monitoring is the requirement for *open universe* probability models, i.e., distributions over properties of objects (in our case, seismic events) in which the number of objects is not fixed a priori; equivalently, distributions over a space with no fixed dimension. It's not difficult to construct such distributions, but they do raise special challenges for inference. Specific models and inference techniques are discussed in more detail below.

1.2 Detection-based vs Signal-based Monitoring

Current monitoring systems typically include an initial preprocessing step in which the raw, continuous signal data at each station are thresholded and converted into a set of discrete "detections", each indicating the possible arrival of a seismic wave at a particular station at a particular time. This greatly simplifies the task of a global monitoring algorithm, which can now get by on modeling and processing this reduced discrete representation instead of the (much larger and messier) original signal data. All global monitoring systems we are aware of operate in this way, including the Global Association (GA) algorithm used in production at the IDC, as well as the Bayesian system NET-VISA, developed in earlier work by Nimar Arora at UC Berkeley [1].

Unfortunately, pre-processing signal data into detections throws away a lot of information that might otherwise be quite helpful for a global monitoring system. The goal of our current work is to avoid this loss, by developing a Bayesian monitoring system that operates directly on raw seismic signals. Such a signal-based approach has many potential benefits, including:

- **Sub-threshold arrivals.** Small blips that do not meet the threshold to trigger a detection, because they may individually be explainable as noise, can nonetheless provide statistical evidence for an event if they are present consistently across stations. Such blips are not visible to a detection-based system, but a signal-based system can aggregate weak evidence across stations and use it to support an event hypothesis.

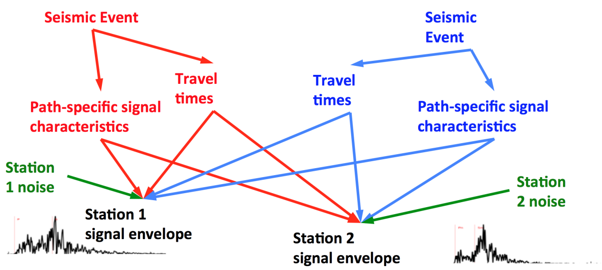

Figure 1: A high-level illustration of the SIG-VISA generative signal model, in a world with two events and two detecting stations.

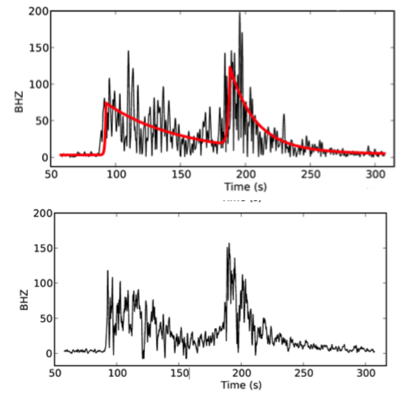

Figure 2: Top: an observed waveform showing P and S phase arrival, with red line indicating template fit(estimated modulation signal not shown). Bottom: a synthetic waveform sampled from the generative model, conditioned on the template fit shown at top but with random modulation and background noise.

- **Waveform shape information.** A signal-based system can estimate specific properties of the waveform shape, for example, the rates of signal onset (from initial arrival to peak) and decay (following the peak). Research from seismology suggests that these shape properties may be informative of other variables such as event-station distance, event magnitude, wave type of the arriving phase (P, S, Love, Rayleigh, etc), and event source type (earthquake versus explosion) [13] [4]. Comparison of energy content and waveform shape across the frequency spectrum can also yield additional information regarding these and other variables.

- **Overlapping arrivals.** If phases from two events (or two phases from the same event) arrive at similar times, a detection-based system might process them as a single detection, potentially confusing the global processing. A signal-based system handles this naturally by simply modeling the sum of the waveforms from both arrivals.

- **Waveform matching.** Two events occurring nearby to each other are likely to generate highly correlated waveforms at any particular station, even if the events are separated in time by many years, because their energy will have traveled via the same set of paths; earth structure is essentially static over human timescales. This allows for the possibility of event localization even from a single station: if we see an waveform that closely correlates with a previously observed event, it's likely in the same location. It also allows for improved detection sensitivity; research has shown that even arrivals below the noise threshold can still be distinguished to the extent that they generate correlations with previously recorded events [8] [20].

Of course, signal-based monitoring also requires a significant increase in model complexity to capture these and other phenomena, with corresponding complications for inference. The following sections provide a brief sketch of a prototype generative model we have developed for seismic signals, along with details of and results from an initial implementation. We refer to the system thus described as SIG-VISA, for SIGnal-based Vertically Integrated Seismological Analysis.

2 Generative Model

Figure 1 shows a sketch of the SIG-VISA generative model. The current version of the model describes the joint distribution of event hypotheses and observed signal envelopes, along with additional latent variables describing the specific disturbances produced by each arriving phase from each event at each station in the network. The model can be understood intuitively by 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; relevant characteristics include origin time, location, depth, body wave magnitude, and whether the event is a natural event or an explosion. These prior distributions are determined empirically from historical training data, as labeled by human analysts. We use largely the same set of prior distributions as described for the NET-VISA system [1].

3. Each event generates one arrival at each station for each seismic phases we expect at that station (the set of phases is computed deterministically based on the event location). The arrival time of each phase depends on the origin time of the event as well as the events location relative to the station.

4. Each arriving phase generates a phase envelope, which is taken to be the product of a *parameterized template shape* with a *stochastic modulation signal.* The template captures the general pattern of initial energy and subsequent decay, while the modulation signal allows for shorter-term fluctuations.

   Currently, we use a parameterized template following a very simple functional form, consisting of a linear onset, followed by exponential decay. This template shape is parameterized by an arrival time, an onset-to-peak time, a peak amplitude, and a decay rate following the peak. The stochastic modulation signal is taken to be a random combination of Fourier basis functions.

5. In addition to the arrivals generated by each event phase, we also sample at each station a set of *unassociated* templates, which will generate spikes in the signal not associated with any particular event. Both the number of unassociated templates at each station and their specific shape parameters (arrival time, onset time, amplitude, and decay rate) are sampled from a prior distribution.

6. Finally, the observed envelope for each frequency band at each station is modeled as the sum of all the arriving phase envelopes, plus the unassociated templates, along with a background noise process. The background noise process is taken to be an autoregressive process, with parameters estimated from training data. Figure 2 shows an example of a signal generated under the model.

Several aspects of this story bear further explanation:

- The distributions on the various template shape parameters, conditioned on the event location, are given by Gaussian process regression models trained on historical data. When possible, we also integrate physics-based models of the relevant phenomena (e.g. we use the IASPEI model of seismic travel times [11], and the Brune [2] and Mueller-Murphy [15] models of source amplitudes across the frequency spectrum), using the Gaussian process to model the residuals from the physics-based model. Gaussian process regression is chosen, first, because it yields full probability distributions rather than simple predictions (necessary for integration into a larger Bayesian model), and second, because as a *nonparametric* method it "remembers" past events from the training data, and will predict similar properties for new events that occur in nearby locations. Thus we are able to learn spatial structure for each parameter, for example, the existence of "hot spots" – regions, potentially quite far away from a station, in which that station is nonetheless particularly sensitive to events, even as other events just a few miles outside of the hot region are much more heavily attenuated – and other similar localized phenomena.

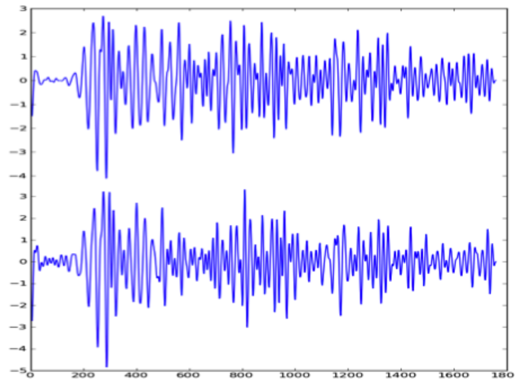

Figure 3: Correlated event waveforms (IMS events 4686108, 4689462, correlation .60).

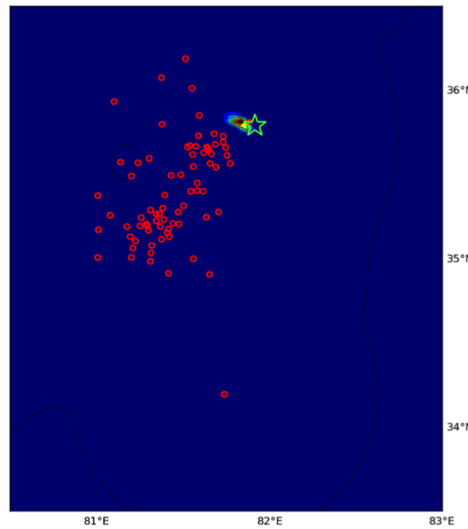

Figure 4: Posterior location density of IMS event 4689462 from a single station observation (MKAR) using a Gaussian process model of Fourier coefficients trained on data from 99 other events, shown in red. The posterior peak, obtained through grid search, is 8.4km from the true event location, marked by a green star.

The coefficients defining the shorter-term fluctuations of the modulation signal – currently, these are the amplitude and phase parameters for a set of Fourier basis vectors – are also modeled at each station by a Gaussian process conditioned on event location. This enables a waveform matching property, as we had hoped for above: under this model, two events occurring in the same location will be predicted to have the same modulations (along with the same general template shape) at any given station, and so will generate the same signals. Conversely, given observed signals from both events at a single station, the posterior distribution on location for the second event will be heavily peaked near the location of the first, previously observed event. Figures 3 and 4 demonstrate this behavior, showing how an event can be located from a single observation through the use of a Gaussian process model on Fourier coefficients.

- The historical data available for training are generated by human analysts based on initial bulletins from an automated system. Aside from not being true "ground truth" (a flaw we mostly ignore), the labeled data provide us only with details of the events themselves, and not with any information on the other latent quantities hypothesized by our model, such as the template shape parameters, modulation signal coefficients, or the number and shape of the unassociated templates. To learn models for these latent parameters, we use a rudimentary ("hard") form of the EM algorithm, in which we begin with trivial models, run an optimization routine to find the highest-probability parameters given the observed signals and ground-truth events under those models, then use the resulting parameters to train new models (e.g., Gaussian processes), and iterate until convergence.

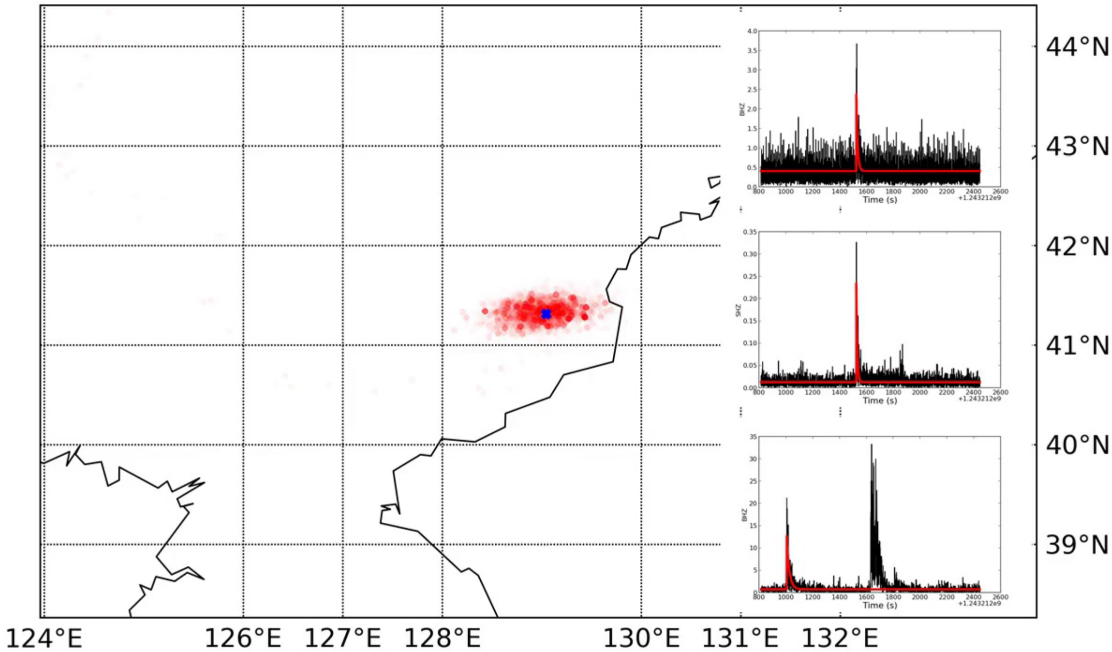

Figure 5: Posterior density obtained through MCMC for the location of the 2009 DPRK nuclear test, conditioned on observations from nine IMS stations. The ground truth location (from the human-annotated LEB bulletin) is shown in blue. Also displayed at right are posterior templates for three of the stations detecting the event.

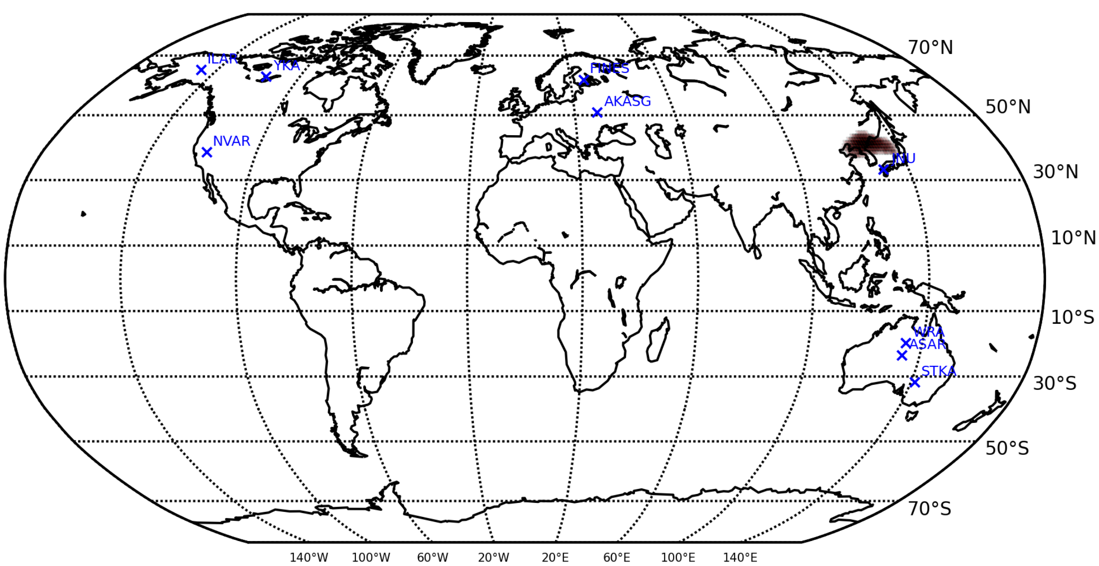

Figure 6: Hough proposal distribution for an event birth move, based on unassociated templates generated by approximately 100 steps of RJMCMC – including template birth, death, split and merge moves – on data from nine IMS stations detecting the 2009 DPRK test event. This demonstrates completely unsupervised discovery of the event's existence and location. The stations are the same as in Figure 5; the station locations are shown in blue on the map.

3 Inference and Results

As mentioned above, we are working with an open-world probability model, so we need an inference algorithm that can operate on a space of variable dimension. We use the framework of Reversible Jump Markov Chain Monte Carlo (RJMCMC) [9], which incorporating pairs of dimension-jumping moves constructed specifically to maintain detailed balance. Our current RJMCMC algorithm incorporates the following move types:

- **Template birth, death, split, merge:** these moves create and destroy new unassociated templates, and allow them to merge with (or split from) other unassociated templates as well as templates generated by events. The proposal distribution for the peak time of a newly birthed template is taken to be proportional to signal amplitude minus the envelopes of current templates, so new unassociated templates are most likely to be birthed in regions of the signal that contain large spikes not currently explained by any other templates.
- **Template-tweaking:** these are simple within-dimension, Metropolis-Hastings moves to adjust template shape parameters and modulation coefficients.
- **Event birth/death:** these moves create and destroy events. Event locations are proposed by a Hough transform (Figure 6), in which each current unassociated template votes for all bins along the space-time cone consistent with the template's arrival time, and the probability of an event proposal in a particular space-time bin is proportional to the votes it receives. When an event is birthed, it is required to generate arrivals for all appropriate phases at all stations; this may be done by proposing either the association of an existing unassociated template, or the creation of an entirely new template. Similarly, when an event is killed, the death proposal must specify, for each of its phase arrivals, whether that template is to be removed entirely or simply converted into an unassociated template.
- **Event-tweaking:** like the template-tweaking moves, these are simple within-dimension MH moves to adjust some attribute of an event (surface location, depth, magnitude, origin time, source type).

We have seen some promising results from this initial stab at an inference algorithm. For example, Figure 5 shows localization of the North Korean nuclear test that occured on May 25, 2009, in which the posterior mean is within 2km of the "ground truth" location, though admittedly with quite a wide region of uncertainty (partially attributable to the small number of detecting stations used). Note that this particular experiment used a fixed hypothesis of a single event; the event birth/death moves were disabled so there was no attempt to create additional events.

3.1 Challenges of Inference

Successful inference in a model with this level of complexity raises many challenges. The most obvious challenge is tractability: proposal distributions and acceptance ratios must be efficiently computable, so that the MCMC chain can progress; we also need the chain to converge to a stationary distribution in a reasonable number of steps. Some specific issues include:

- **Efficient evaluation of Gaussian process posteriors**. The nice properties of Gaussian process regression come at a computational cost. A naïve calculation of the GP posterior distribution (mean and variance) at a particular point involves a vector multiplication with the training set covariance matrix, requiring time $O(n^2)$ in the number of training points. Since we must often do this calculation for several GP models to calculate a single acceptance ratio, and many stations have tens of thousands of historical data points, the situation quickly becomes untenable. For GP models such as ours in which the covariances are highly local, this complexity can be improved somewhat by the use of sparse matrices, but still scales at least linearly with the training set.

  To take better advantage of this local-covariance property – which states, for example, that the distribution over travel-times for an event in Japan should only depend on historical events nearby in Japan, and shouldn't be slowed down if we've happen to have also observed millions of events in Australia – we developed a data structure and associated algorithm, called a *product tree*, which performs the necessary matrix computations while ignoring irrelevant training points, by operating on a spatial, hierarchical tree representation of the training data. [14] Unfortunately, our current implementation does not have favorable cache behavior, so although it requires significantly fewer FLOPS than a sparse matrix multiplication to reach the same result, it is not faster in real-world usage. We hope to address this with further optimizations, and perhaps explore an alternative approach using locality sensitive hashing.

- **Local Maxima**. The use of a modulation signal allows an envelope to be highly optimizated to fit the observed signal at its current time; this same property makes it exceedingly difficult to accept a move that *changes* the time (or other significant shape parameters) of a well-optimized envelope, even if the original time was not optimal. We are working to address this through joint proposals that also modify the modulation signal to adjust for the change in arrival time (or other shape parameter). It may also prove necessary to implement some form of tempering to allow the MCMC chain to move more easily between separate local maxima.

- **Waveform-based event proposals**. Suppose that we observe signals from an event for which we have a nearby doublet event in our training set. As described above, the waveform-matching behavior of the model will result in high posterior probability immediately surrounding the location of the matching event in the training set; however, the vast majority of the probability space will be relatively flat, until the event gets close enough (within a few km) to match its doublet. Thus, the waveform-matching behavior of the model is mostly ineffective unless accompanied by a proposal that can recognize a signal that correlates highly with a previous observation, and directly propose the location of that observation. This is not difficult in principle, but cross-correlating all incoming signals against a large library of historical data in reasonable time will require a nontrivial engineering effort.

4 Other challenges

Aside from the aforementioned challenges with inference, we face several other challenges and directions for future work, some related specifically to our particular model and others more general.

- **Non-Fourier modulation signals.** Although Fourier coefficients have served as a useful proof of concept representation for the modulation signal, they also have serious drawbacks. First of all, they are inherently periodic, and so can only represent a fixed length of signal before repeating (which makes no sense in the context of our model); this requires us to choose a fixed length of modulation which will necessarily be too short for slowly-decaying signals, since we will stop modulating them before they reach the noise floor, but too long for quickly-decaying signals, since some of their Fourier coefficients will depend on signal behavior below the noise floor. A related issue is that a Fourier representation is non-local in the time domain; when a signal goes below the noise floor we would like to have higher confidence in our estimates of the first part of the modulation signal, above the noise floor, than in the second part below, but a Fourier representation makes no such distinctions. We intend to explore wavelet representations as one potential alternative, though as yet we have no experience with their use.

- **Non-Gaussian observation models.** The standard formulation of Gaussian process regression assumes a Gaussian observation model, which dovetails nicely with other assumptions to yield an analytic Gaussian posterior. However, not all of our parameters of interest follow Gaussian distributions; for example, seismic travel time residuals are known to follow a Laplacian distribution (with single-exponential tails heavier than a Gaussian's double-exponential tail) [1] and our models should accomodate that. Unfortunately, non-Gaussian observation models remove the analytic tractability of GP regression and allow for only an approximation to the posterior [19]. More work is required to understand whether this can be made tractable for our needs.

- **Implementation in BLOG.** Bayesian LOGic (BLOG) is a language for open-university probability models, capable of performing automatic learning and inference given only a declarative model specification. Currently the general-purposes inference techniques available in BLOG are too slow for a model as complex as SIG-VISA, but a full BLOG implementation of SIG-VISA would still be useful; it would both provide a formal specification of the SIG-VISA model and create a difficult "stress test" motivating further improvements in general-purpose BLOG inference.
- **Interpreting high-dimensional posteriors.** This last challenge is one of the simplest and yet most general, concerning the interpretation of a high-dimensional joint distribution over a varying set of objects. Given a set of MCMC posterior samples, how do we summarize the information they contain to best understand the results of the inference? Note that the MAP estimate is not necessarily well-defined on a space of variable dimension, as probability densities are sensitive to normalizing constants which may vary with the dimension. Some of the relevant issues involve identity uncertainty: if an event birth move and corresponding death move each succeeded many times in the same region, how do we decide whether to report one event in that region, existing with, say, 50% probability, versus the possible existence of many events, each with a much smaller individual probability of existing? Can we identify sets of consistent or opposing hypotheses, e.g. "some samples hypothesize events A and C, and others B and D, but no sample ever hypothesizes A and B together"? Are there other ways of visualizing or interpreting the results of inference?

5 Proposed Milestones for Future Research

- **November 2013.** Evaluate current prototype model (single frequency band, no array stations, no waveform matching) for detection and localization performance from the full worldwide monitoring network, first using synthetic signals, then real observed signals. For observed signals, compare to detection-based models (GA, NET-VISA) and to a human analyst bulletin (LEB).
- **December 2013.** Integrate model of correlated signals at array stations; demonstrate improvements in sensitivity and localization from small sets of stations.
- **February 2014.** Performance improvements (improved MCMC scheduling, sparse GP training, full integration of semi-parametric GP models, multi-core parallelization, possibly Hamiltonian MCMC) to allow for tractable inference using GP models trained on 10 years of historical data from the IDC, including joint models for array stations. Goal: process an hour of worldwide network data in less than one week (thus enabling realtime inference given a 168-node cluster).
- **March 2014.** Extend the model to operate over multiple frequency bands; evaluate the impact on sensitivity, localization, magnitude and source type discrimination. Integrate non-Gaussian observation models for GP regression, e.g. of travel time residuals.
- **April 2014.** Explore extended envelope shape models (including e.g. separate terms for peak versus coda shape) and modify the model as appropriate. Incorporate LLNL 3D travel-time model. Augment parametric part of GP parameter models to include regional binning and earth-structure features, for more accurate prediction of envelope shapes.
- **May 2014.** Comprehensive evaluation against NET-VISA and LEB on full two-week test set, including signal models at array stations and multiple frequency bands. Initial exploration of non-Fourier modulation representations, and implementation of a candidate (wavelet?) representation.
- **September 2014.** Implement waveform-correlation-based event proposal distribution, along with MCMC moves for the new wiggle representation. Evaluate sensitivity and location improvements from the waveform-matching model.

- **October 2014.** Additional model and/or inference improvements (motivated by experiment): possibly parallel/tempered MCMC, or heteroskedastic GP regression / GP regression with uncertain inputs.
- **December 2014.** Demonstrate state-of-the-art monitoring performance using full model on global network data. Full specification of the SIG-VISA seismic model in (extended) BLOG syntax.
- **Spring 2015.** Additional refinements to the SIG-VISA model and system, including initial prototype models for hydroacoustic and infrasound data. Identify and implement extensions to the BLOG formalism necessary to support (in principle) the SIG-VISA model.
- **Fall 2015-Spring 2016.** Augment BLOG to support efficient learning and inference in the SIG-VISA model.

References

[1] Nimar Arora, Stuart Russell, Paul Kidwell, and Erik B. Sudderth. Global seismic monitoring as probabilistic inference. In *Advances in Neural Information Processing Systems*, page 7381, 2010.

[2] James N Brune. Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research*, 75(26):4997–5009, 1970.

[3] J. V Candy. *Bayesian signal processing: classical, unscented and particle filtering methods.* Wiley-Blackwell, Oxford, 2009.

[4] Georgia B. Cua. *Creating the Virtual Seismologist: developments in ground motion characterization and seismic early warning.* phd, California Institute of Technology, 2005.

[5] M. P. Flanagan, S. C. Myers, and K. D. Koper. Regional travel-time uncertainty and seismic location improvement using a three-dimensional a priori velocity model. *Bulletin of the Seismological Society of America*, 97(3):804–825, June 2007.

[6] M. P. Flanagan, S. C. Myers, C. A. Schultz, M. E. Pasyanos, and J. Bhattacharyya. LLNLs 3-DA priori model constraints and uncertainties for improving seismic location. Technical report, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, 2000.

[7] Ludwig Geiger. Probability method for the determination of earthquake epicenters from the arrival time only. *Bulletin of St. Louis University*, 8(1):56–71, 1912.

[8] Steven J. Gibbons and Frode Ringdal. The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1):149166, 2006.

[9] David I. Hastie and Peter J. Green. Model choice using reversible jump markov chain monte carlo. *Statistica Neerlandica*, 66(3):309338, 2012.

[10] Juerg Hauser, Kathleen M. Dyer, Michael E. Pasyanos, Hilmar Bungum, Jan I. Faleide, Stephen A. Clark, and Johannes Schweitzer. A probabilistic seismic model for the european arctic. *Journal of Geophysical Research*, 116(B1), January 2011.

[11] B. L. N. Kennett and E. R. Engdahl. Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, 105(2):429465, 1991.

[12] G. Lin, C. H. Thurber, H. Zhang, E. Hauksson, P. M. Shearer, F. Waldhauser, T. M. Brocher, and J. Hardebeck. A california statewide three-dimensional seismic velocity model from both absolute and differential times. *Bulletin of the Seismological Society of America*, 100(1):225–240, January 2010.

[13] Kevin Mayeda, Abraham Hofstetter, Jennifer L. O'Boyle, and William R. Walter. Stable and transportable regional magnitudes based on coda-derived moment-rate spectra. *Bulletin of the Seismological Society of America*, 93(1):224–239, February 2003.

[14] David A. Moore and Stuart J. Russell. Product trees for gaussian process covariance in sublinear time. In *Proceedings of the UAI-2013 Workshop on Models of Spatial, Temporal, and Network Data (UAI-MSTND)*, Bellevue, WA, 2013.

[15] Richard A. Mueller and John R. Murphy. Seismic characteristics of underground nuclear detonations. *Bulletin of the Seismological Society of America*, 61(6):16751692, 1971.

[16] Stephen C. Myers, Gardar Johannesson, and William Hanley. A bayesian hierarchical method for multiple-event seismic location: Bayesian stochastic multiple-event location. *Geophysical Journal International*, 171(3):1049–1063, November 2007.

[17] Michael E. Pasyanos, William R. Walter, and Kevin Mayeda. Exploiting regional amplitude envelopes: A case study for earthquakes and explosions in the korean peninsula. *Bulletin of the Seismological Society of America*, 102(5):1938–1948, 2012.

[18] Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization. Annual report 2012. Technical report, 2012.

[19] Carl Edward Rasmussen and Christopher K. I Williams. *Gaussian processes for machine learning.* MIT Press, Cambridge, Mass., 2006.

[20] M. Slinkard, Stephen Heck, D. Carr, Regina Eckert, and C. Young. Broad area event detection using waveform correlation and distributed computing. *Proceedings of 2012 Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, page 806815, 2012.

[21] Felix Waldhauser and William L. Ellsworth. A double-difference earthquake location algorithm: Method and application to the northern hayward fault, california. *Bulletin of the Seismological Society of America*, 90(6):13531368, 2000.

[22] William R. Walter, Kevin M. Mayeda, and Howard J. Patton. Phase and spectral ratio discrimination between NTS earthquakes and explosions. part i: Empirical observations. *Bulletin of the Seismological Society of America*, 85(4):1050–1067, August 1995.