Model-based Bayesian Seismic Monitoring

Nimar S Arora

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2012-125
http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-125.html

May 29, 2012

Copyright © 2012, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Model-based Bayesian Seismic Monitoring

by

Nimar S. Arora

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Stuart Russell, Chair
Professor Barbara Romanowicz
Associate Professor Dan Klein

Spring 2012

Model-based Bayesian Seismic Monitoring

Copyright 2012
by
Nimar S. Arora

Abstract

Model-based Bayesian Seismic Monitoring

by

Nimar S. Arora

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Stuart Russell, Chair

This thesis presents the underlying probabilistic model, the parameter estimation, and the inference algorithm of *NET-VISA*, Network Processing Vertically Integrated Seismic Analysis. NET-VISA is an Open Universe Probability Model (OUPM) for seismic events, the transmission of seismic waves through the earth, and their detection (or misdetection) at stations, as well as a model for spurious detections.

The probabilistic model allows for seamless integration of various disparate sources of information. Applied in the context of the International Monitoring System (IMS), a global sensor network developed for the Comprehensive Nuclear-Test-Ban Treaty (CTBT), NET-VISA achieves a reduction of around 60% in the number of missed events compared to the currently deployed system. It also finds events that are missed by the human analysts who post-process the IMS output.

Contents

List of Figures **iii**

List of Tables **v**

1 Introduction **1**

2 Related Work **5**
2.1 Signal Processing 5
2.2 Iterative Linear Least Squares 5
2.3 Global Association (GA) 8
2.3.1 False Event Identification (FEI) 9
2.4 Multiple-event Location 9
2.4.1 BAYHLoc 10
2.5 Data Association and Open Universe Probability Models 12
2.6 Tempered Inference 12

3 Model Motivation **13**

4 Generative Probabilistic Model **17**
4.1 Events 17
4.1.1 Event Rate and Time 17
4.1.2 Event Location 18
4.1.3 Event Magnitude 19
4.1.4 Overall Event Prior 20
4.2 Correct Arrivals 21
4.2.1 Detection Probability 21
4.2.2 Arrival Attributes 22
4.2.3 Overall Correct Arrivals 26
4.2.4 Model Validation 26
4.3 False Arrivals 27
4.4 Coda Arrivals 30

5 Inference **37**
5.1 Overall Algorithm . . . 37
5.2 Moves . . . 39
5.2.1 Birth Move . . . 39
5.2.2 Improve Arrival Move . . . 41
5.2.3 Improve Events Move . . . 41
5.2.4 Death Move . . . 41
5.3 Parallel Birth Moves . . . 41
5.4 Tempering . . . 42

6 Experimental Results **44**
6.1 Evaluation of seismic bulletins . . . 44
6.2 Comparison of NET-VISA and SEL3 . . . 45
6.3 Comparison of NET-VISA and LEB . . . 47
6.4 Ablation Experiments . . . 48
6.5 Data and Resources . . . 50

7 Conclusions and Further Work **54**

Bibliography **56**

List of Figures

1.1 All true (yellow stars) and predicted (red squares) events from the current automated system for a one week period 3
1.2 All true (yellow stars) and predicted (blue squares) events for a one week period 4

2.1 Example of seismic waveform, STA/LTA, and arrivals. 6

3.1 2-D world with seismic stations (triangles) 13
3.2 Posterior event location density given that one stations missed it. 14
3.3 Posterior location uncertainty with arrival time uncertainty of 0.1. 15
3.4 Posterior location uncertainty with arrival time uncertainty of 0.05. 16
3.5 Posterior location uncertainty with two stations failing to detect. 16

4.1 Estimating event rate. 18
4.2 Event location average leave-one-out log likelihood vs. bandwidth. 19
4.3 Event location log density. 20
4.4 Conditional detection probabilities for the P phase of surface events between 3 and 4 mb at Station ASAR. 23
4.5 Arrival time distribution around the IASPEI prediction for P phase arrivals at station ASAR. 25
4.6 Arrival azimuth distribution around the geographical prediction (plus SASC correction) for P phase arrivals at station ASAR. 26
4.7 Arrival slowness distribution around the IASPEI prediction (plus SASC correction) for P phase arrivals at station ASAR. 27
4.8 Arrival log amplitude distribution for the P phase of a surface event of 3–4 mb at station ASAR. 29
4.9 Arrival phase probability as function of the true phase. 30
4.10 Amplitude distribution for false arrivals at stations ASAR. 32
4.11 Phase distribution for all false arrivals. 33
4.12 The average false arrival rate per hour at all the stations. 33
4.13 Coda detection probability as a function of the triggering arrival's log amplitude. 34
4.14 Time delay for coda arrival after the triggering arrival. 34

4.15 Coda azimuth difference versus triggering arrival. 35
4.16 Coda slowness difference versus triggering arrival. 35
4.17 Coda log amplitude difference versus triggering arrival. 36
4.18 Coda phase. 36

5.1 Distance between events and inverted locations within 10 degrees and 100 seconds. The left figure shows the distance from a true event and the nearest inverted location. The right figure shows the converse, i.e. the distance from an inverted location to the nearest true event. 40

6.1 Precision-recall performance of the proposed NET-VISA and deployed SEL3 algorithms, treating the analyst-generated LEB as ground truth. 45
6.2 SEL3 (red circle), LEB (yellow circle), NET-VISA (blue square) location and posterior log odds ratio around LEB origin-id 5287957. 48
6.3 Performance after ablating arrival slowness. 50
6.4 Performance after ablating arrival azimuth. 51
6.5 Performance after ablating arrival phase. 52
6.6 Performance after ablating arrival amplitude. 52
6.7 Performance after ablating coda model. 53
6.8 Performance after ablating tempering model. 53

List of Tables

4.1 List of features used for computing the probability of detecting an arrival from an event i with magnitude e^i_m, depth e^i_d, and distance Δ_{ik} from station k. Here $\mathcal{N}(x, \mu, \sigma)$ is the standard Gaussian density with mean μ and standard deviation σ measured at x. Also, the log likelihood from using that feature. . 22
4.2 List of features used for predicting the amplitude of a phase j arrival from an event i at a station k. Also, the log likelihood after adding each feature . . . 28
4.3 Average log likelihood on held-out data of a Laplacian and a Gaussian model. 28
4.4 Correlation between the various residuals 31

5.1 Event to station distance and event depth ranges for seismic phases. 43

6.1 Recall and average error (km) subdivided by LEB event magnitude (m_b). . . 46
6.2 Recall and average error (km) subdivided by LEB event azimuth gap 46
6.3 The various bulletin events around LEB origin-id 5287957. 46
6.4 The station-phase associations of various bulletin events around LEB origin-id 5287957. 47
6.5 Precision, recall, and average error (km) of LEB and NET-VISA measured against various regional bulletins. 49
6.6 Recall and average error (km) subdivided by event magnitude (ML) in the continental U.S. 49
6.7 The definition of the various regions used for the regional evaluation and the corresponding ground truth bulletin. 50

Acknowledgments

Personal

I never would have joined the PhD program at Berkeley if my wife, Geeta, had not encouraged me, in the summer of 2005, to take the plunge. At that point I was mentally and emotionally chafing from the wear and tear of a grueling software development cycle in the replication group at Oracle – design, develop, deploy, get overwhelmed by customer bugs, redesign, and repeat. Of course, most sensible people at a similar point in their career join an MBA program or start their own company. Thankfully, sensibility is a vice that I have never been accused of.

Once I joined the AI program, my adviser Prof. Stuart Russell opened my mind to a world of opportunities. I dabbled in game theory, question answering, open universe languages, and finally I settled on nuclear explosion monitoring, which seems to be an exciting new area for AI. It was the seismologists at the Comprehensive Test Ban Treaty Organization (CTBTO) who invited us to work on the monitoring problem. They had the firm belief that recent advances in AI could transform their field. Dr. Ola Dahlman chaired the first meeting, in the spring of 2009, between seismologists and "data miners." In this meeting Stuart and I presented our ideas on applying open universe models. Our ideas were received with skepticism, but we were encouraged to try anyway.

It would have been impossible to make much headway into explosion monitoring without the help of various seismologists and other researchers in the monitoring community. Ronan Le Bras the former head of the Software Integration Unit, at the International Data Centre (IDC) Division of the CTBTO, was the guiding force behind the project. I was also helped by Robert Engdahl, David Bowers, Heidi Kuzma, Robert Pearce, Sheila Vaidya, Mark Prior, Jennifer Ylo, Misrak, Gadi, Stephen Myers, Dmitry Storchak, Istvan Bondar, and Barbara Romanowicz.

In Berkeley, I received the support of various faculty, postdocs, graduate students, and undergrads. Prof. Dan Klein's classes in AI were the most fun and educative classes that I have ever taken. Brian Milch, Jason Wolfe, Kevin Canini, Nick Hayes, Shaunak Chatterjee, Norm Aleks, Rodrigo de Salvo Braz, Erik Sudderth, and Emma Brunskill were always available for bouncing off ideas. Ariel Kleiner and Lester Mackey shared their results on false event identification. Paul Kidwell at Lawrence Livermore figured out the right set of features for computing event detection probabilities. The undergrads Matthew Cann, Hong Hu, Christopher Lin, Andrew Lee, and Ron Sun helped in various capacities. Tony Dear's implementation of multi-threaded inference helped me run my tests much faster.

Meanwhile in Oracle, my managers Alan Downing, Lik Wong, and Jagdev Dhillon allowed me to work part-time while studying at Berkeley. Without this crucial support, a PhD would have been economically unviable. Of course, it was because of a number of very

sharp developers in my group that I had the flexibility of focusing solely on design issues. And if I ever waivered in my commitments to completing my PhD I always had Mahesh Subramaniam and Jing Liu to tease and goad me on.

Finally, I'd like to thank my extended family and friends for emotionally sustaining me for the past six years.

Technical

The initial seismic model with only one seismic phase (the first arriving P phase) and arrivals lumped in 5 second buckets was presented as a poster and plenary talk in International Scientific Studies (ISS) 2009. The co-authors for this initial work were Stuart Russell and Erik Sudderth. Subsequently, a model with precise travel time for the first arriving phase and corresponding MCMC inference was presented in American Geophysical Union (AGU) 2009 as a poster and talk. My advisor, Stuart Russell, was involved in all the research presented in this thesis. Erik Sudderth was involved in the initial model design and later in various capacities.

In 2010, the model was extended to include the 14 most common seismic phases. Paul Kidwell of Lawrence Livermore joined the team at this point. He helped build a detection probability model and wrote code in R to learn the parameters of the model. He later developed a hierarchical model for learning these parameters. The resulting work was presented in European Seismological Convention (ESC) 2010 as a talk. This was also published in Arora et al. (2010a, 2011). The current work still uses the same features that Paul developed although the inference code has been rewritten in Python. The idea of a hierarchical model for learning the parameters has been adapted to other features, for example the arrival amplitude model. The resulting work with various minor refinements as well as correct handling of noise phases was presented in Science and Technology (S&T) 2011.

Tony Dear, an undergraduate at Berkeley, started helping out with the parallelization of the inference in 2011. He implemented the multi-threaded birth moves as well as the first version of the distributed inference. This addition helped scale the inference to large aftershock sequences like Tohoku 2011 and Sumatra 2004. The work on scaling inference was presented in AGU 2011 as a poster, which won the outstanding student paper award.

The coda model was added in early 2012 and was presented in European Geophysical Union (EGU) 2012 as a talk. The final addition of tempering to correctly handle seismic phase constraints as presented here has not so far been published elsewhere.

Throughout the project I received funding from CTBTO. The work described in this thesis was also presented in various workshops and talks organized by the CTBTO. Since 2011 most of the testing, specially the distributed computation, was done on Amazon EC2 which would not have been possible without generous research grants by AWS in Education.

Chapter 1

Introduction

The Comprehensive Nuclear-Test-Ban Treaty (CTBT), which bans all nuclear explosions on earth, is gaining renewed attention in light of growing worldwide interest in mitigating the risks of nuclear weapons proliferation and testing. To monitor compliance with the treaty, the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) has installed a suite of sensors of the International Monitoring System (IMS). The IMS includes waveform physical sensor stations (seismic, hydroacoustic, and infrasound) connected by a worldwide communications network to a centralized processing system in the International Data Centre (IDC) in Vienna. The IDC operates continuously and in real time, performing *station processing* (analysis and reduction of raw seismic sensor data to detect and classify signal arrivals at each station) and *network processing* (association of signals from different stations that have come from the same event). Network processing is thus a *data association* problem similar to those arising in multi-target tracking (Bar-Shalom and Fortmann, 1988). Perfect performance remains well beyond the reach of current technology: the IDC's automated system, a highly complex and well-tuned piece of software, misses nearly one third of all seismic events in the magnitude range of interest, and about half of the reported events are spurious. A large team of expert analysts post-processes the automatic bulletins to improve their accuracy to acceptable levels.

The existing network processing algorithm in use at the IDC treats the problem as multiple independent sub-problems which are solved by a pipeline of modules. The output of each module is fed into the next one without any opportunity to revisit a previous decision. For example, a seismic phase label is assigned to each arrival based on signal properties alone and this label is left unchanged even if it would make sense to choose a different phase label based on a potential event which is inferred later on in the pipeline. In reality, many of the decisions made by these modules, like assigning phase labels to arrivals, have a high degree of uncertainty with them. It is inadvisable to make a hard decision on a small piece of the input without analyzing the larger context. In this thesis, we address this shortcoming in a vertically integrated probability model for the seismic component of network processing (NET-VISA). We present a generative model which combines seismic knowledge with sta-

tistical modeling and inference. The NET-VISA model starts by generating seismic events, and then predicts the transmission and attenuation of seismic waves all the way to their detection (or misdetection) at seismic stations. Of course, this model relies on the hard decisions made by the existing signal processing software in terms of detecting arrivals. A more complete model which extends all the way to seismic waveforms, SIG-VISA, is one of the future goals of the larger VISA project.

The main contribution of this thesis is an integrated probability model for network processing. Due to the integrated nature of the model, the individual components use much more information than existing methods. Most notably the location of seismic events in NET-VISA uses information from arrival time, arrival azimuth (the incoming direction of the seismic waves), arrival slowness (the apparent surface speed of the waves), arrival amplitude, as well as detections and misdetections (i.e. negative information) of all potential seismic phases at all the stations. In contrast, the standard seismic processing uses mainly arrival time, and sometimes also includes azimuth and slowness. The output of the processing produces the event longitude, latitude, depth, and time. Whereas in NET-VISA, the event magnitude, since it affects the detection probability and the arrival amplitude, is deduced simultaneously with the event location. In classical systems, event magnitude is computed after the location is already established. A subsequent fitness measure might be computed in these systems based on the event magnitude and the detections, but this is used mainly to filter out unlikely events.

The other important aspect of the probability model is the model for spurious detections. We distinguish independent false arrivals with another type of false arrival, the *coda* arrivals, that arise because of the scattering of seismic energy. These two alternate explanations allow the model to explain very complex and noisy sequences of arrivals without the need to filter out coda arrivals upfront (as is done currently at the IDC).

The secondary contribution of this thesis is an inference algorithm which is based on simple heuristic moves motivated by generic Bayesian inference on Open Universe Probability Models (OUPMs). The inference algorithm runs on a continuous stream of incoming arrivals and produces a continuous output of event bulletins. A multithreaded and distributed implementation of the algorithm is described for better scaling on large aftershock sequences. The inference enforces hard seismic constraints on the permissible phase labels by borrowing ideas from tempered probabilistic inference.

The results demonstrate that not only is NET-VISA able to outperform the current automated system at the IDC, SEL3, but it is, in fact, able to significantly reduce the detection threshold by detecting low magnitude seismic events which are missed by human analysts in the LEB bulletin. The quality of the NET-VISA bulletin is immediately evident by the distribution of events predicted by SEL3 and NET-VISA over a one week period. These events are shown in Figures 1.1 and 1.2 respectively. Also shown in the figures are the LEB events, which we will treat as the ground truth for the most part. These figures show that the NET-VISA events follow the locations of the LEB events much more closely than SEL3. For this one week period, the SEL3 bulletin produces 53.8% spurious events

and misses 30.3% of the LEB events, while NET-VISA produces 47.1% spurious events and misses only 11.8% of the LEB events. In fact, many of the "spurious events" produced by NET-VISA turn out to be weak local events which can be validated by dense regional bulletins around the world.

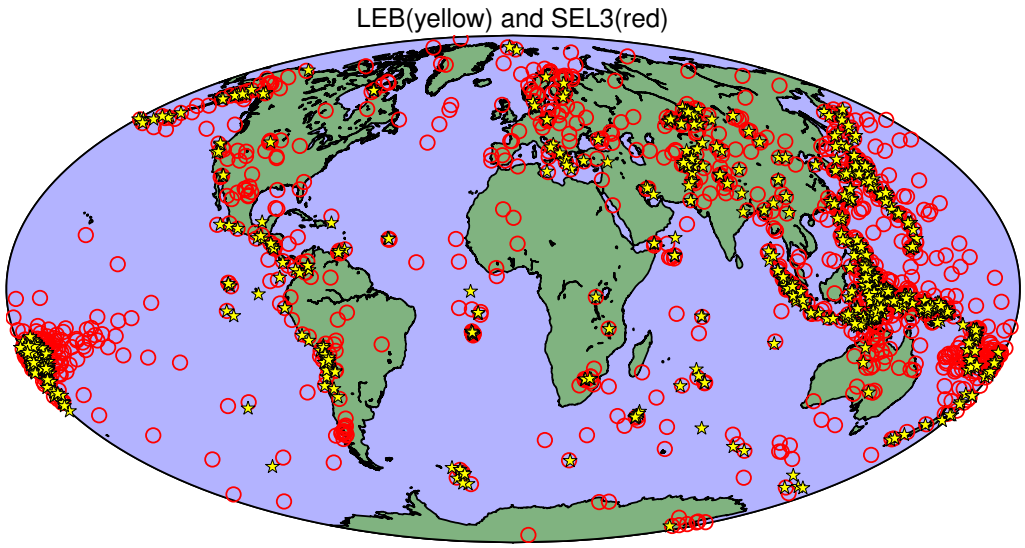

Figure 1.1: All true (yellow stars) and predicted (red squares) events from the current automated system for a one week period

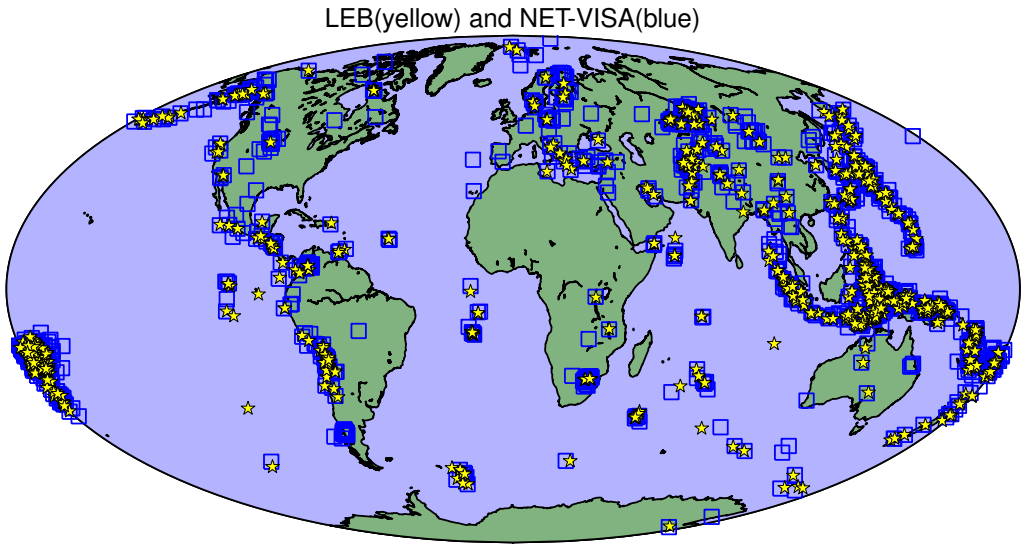

Figure 1.2: All true (yellow stars) and predicted (blue squares) events for a one week period

Chapter 2

Related Work

2.1 Signal Processing

Signal processing is the first step of all seismic analysis. The main objective is to reduce a seismic waveform into a series of parameterized arrivals that can be used by network processing to form events. The IDC signal detection algorithm looks for spikes in the seismic waveform by dividing the short-term average (STA) of the signal amplitude by the long-term average (LTA). Whenever this ratio exceeds a predetermined threshold, a seismic *arrival* is declared as being detected. Further processing is applied to compute the precise parameters of the arrival. The main parameters that are considered here are arrival time, azimuth (direction of the seismic wave), slowness (apparent surface speed of the wave), and amplitude (the height of the first spike).

An example of the signal processing is shown in Figure 2.1. The top panel of this figure displays a filtered seismic waveform, and the lower panel shows the STA/LTA ratio. In the lower panel the blue lines mark the automatically detected arrivals while the red lines mark the analyst-detected arrivals. From this figure alone it is hard to justify many of the marked arrivals except for the largest one around 100 seconds into the waveform. In practice, the signal processing algorithm has access to multiple *channels* of seismic data. These channels could be constructed by filtering a single seismometer's output into different frequencies. Or the channels could be the output of different seismometers at the same location or, in the case of seismic arrays, over an area spanning tens of kilometers.

2.2 Iterative Linear Least Squares

Locating a seismic event given all of its associated arrival times at a network of seismic stations is a well studied, fundamental problem of seismology. Geiger's method (1910; 1912) is the basis of the standard solution for this problem. This method was implemented soon after the widespread availability of computing by Bolt (1960), Flinn (1960) and Engdahl and

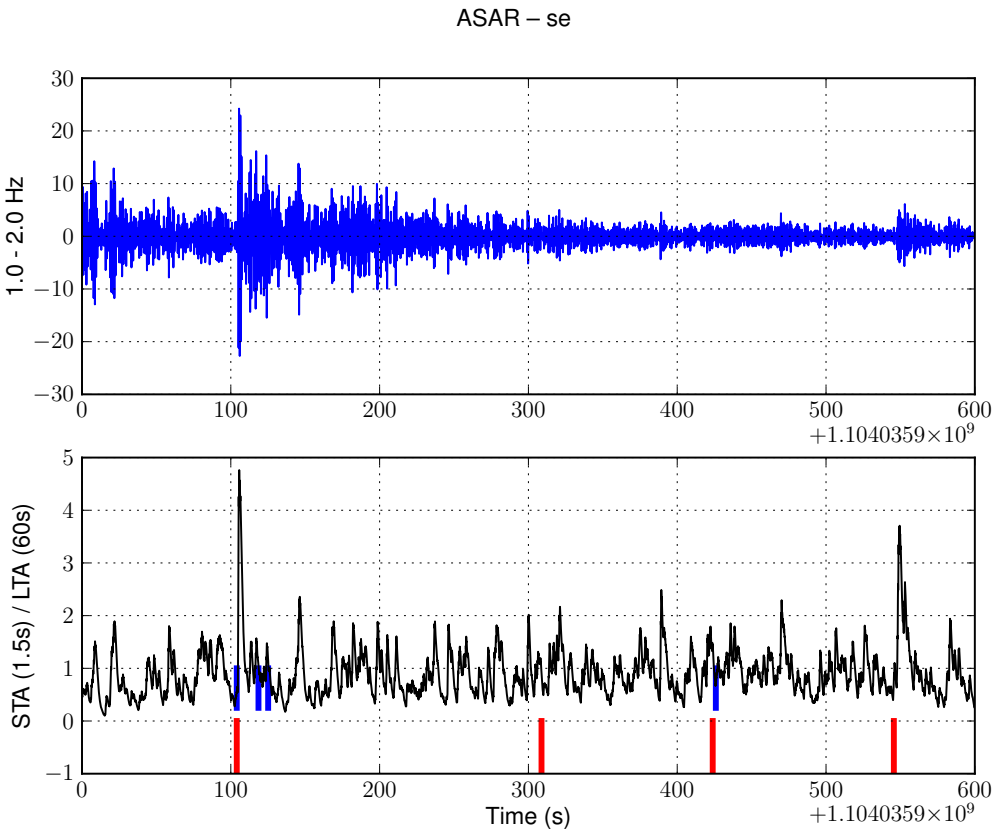

Figure 2.1: Example of seismic waveform, STA/LTA, and arrivals.

Gunst (1966).

In the following description of the method we will assume that there is only one seismic phase that is observed at each station. Let $\mathbf{e} = (e_{lon}, e_{lat}, e_d, e_t)$ be the longitude, latitude, depth, and time respectively of the event we are trying to locate. Also, let Δ_k be the distance of this event from station k, and I_T be a travel time table (a function of event depth and distance to station). Then the arrival time of the seismic waves (from the event) at the station k is given by,

$$\Lambda_t^k = e_t + I_T(e_d, \Delta_k) + r_t^k,$$

where r_t^k is an unknown time residual. The goal, roughly, is to find an event such that the time residuals at all the stations are minimized. More precisely, if $\mathbf{r_t}$ is the vector of time residuals at all the stations, then we are trying to minimize $\|\mathbf{r_t}\|$. The iterative algorithm starts with an initial event location $\mathbf{e^0}$ and attempts to minimize $\|\mathbf{r_t}\|$ by repeatedly making a small perturbation $\delta\mathbf{e^0}$ from the current location. If,

$$F_t^k(\mathbf{e^0}) = e_t^0 + I_T(e_d^0, \Delta_k^0),$$

then we can linearize F_t^k around $\mathbf{e^0}$ as follows,

$$F_t^k(\mathbf{e^0} + \delta\mathbf{e^0}) = F_t^k(\mathbf{e^0}) + \frac{\partial F_t^k(\mathbf{e^0})}{\partial \mathbf{e^0}} \cdot \delta\mathbf{e^0}.$$

Let A be the matrix of partial derivatives of F_t^k at all stations (i.e. A is a K x 4 matrix, where K is the number of stations). The perturbation that minimizes $\|\mathbf{r_t}\|$ under this linear assumption is given by the solution of,

$$\mathbf{r_t} = \mathsf{A} \cdot \delta\mathbf{e^0}. \tag{2.1}$$

As shown by Menke (1989) and Lay and Wallace (1995) we can solve Equation 2.1 by the singular value decomposition (SVD) of $\mathsf{A} = \mathsf{U} \cdot \mathsf{W} \cdot \mathsf{V}^\mathsf{T}$ as,

$$\mathsf{V} \cdot \mathsf{W}^{-1} \cdot \mathsf{U}^\mathsf{T} \cdot \mathbf{r_t} = \delta\mathbf{e^0}.$$

An alternate approach is to multiply both sides of Equation 2.1 by A^T to get,

$$\mathsf{A}^\mathsf{T} \cdot \mathbf{r_t} = \mathsf{A}^\mathsf{T} \cdot \mathsf{A} \cdot \delta\mathbf{e^0},$$

which can be solved by inverting $\mathsf{A}^\mathsf{T}\mathsf{A}$,

$$(\mathsf{A}^\mathsf{T} \cdot \mathsf{A})^{-1} \cdot \mathsf{A}^\mathsf{T} \cdot \mathbf{r_t} = \delta\mathbf{e^0}.$$

Of course, if the matrix A is singular or ill-conditioned then a better approach is to add a damping term λ as described in Ballard and Reeves (2002) to get,

$$(\mathsf{A}^\mathsf{T} \cdot \mathsf{A} + \lambda\mathsf{I})^{-1} \cdot \mathsf{A}^\mathsf{T} \cdot \mathbf{r_t} = \delta\mathbf{e^0}.$$

Instead of minimizing $\|\mathbf{r_t}\|$, we could minimize $\|{\sigma_\mathbf{t}}^{-1} \cdot \mathbf{r_t}\|$, where $\sigma_\mathbf{t}$ is a diagonal matrix of standard deviations for each station. This is equivalent to maximizing the likelihood of a generative model where,

$$r_t^k \overset{iid}{\sim} \mathcal{N}(0, \sigma_t^k), \quad k = 1 \ldots K.$$

In other words, the residuals of all the stations are drawn independently from a zero-mean Gaussian with station-specific variance. The corresponding form of Equation 2.1 that needs to be solved is,

$${\sigma_\mathbf{t}}^{-1} \cdot \mathbf{r_t} = \mathsf{A} \cdot \delta\mathbf{e^0}. \tag{2.2}$$

Equation 2.2 can also be used to estimate the confidence ellipse of the final solution as shown by Flinn (1965).

The above description is the crux of the iterative linear least squares method of Geiger. It has been extended over time by incorporating evidence resulting from more phases and additional measurements. Initially, the Jeffreys-Bullen tables for P and S phases (Jeffreys and Bullen, 1940) were mainly used. Later, with the advent of the IASPEI 1-D model (Kennett

and Engdahl, 1991) many more phases were used in event location. The widespread adoption of seismic arrays as well as 3-component stations (with seismometers facing the north, east, and vertical directions) allowed for the additional use of azimuth and slowness data as shown in the work of Roberts et al. (1989) and Magotra et al. (1987) for example. Similar to the travel time residuals at all the stations, the residuals in the azimuth and slowness are assumed to be independent.

2.3 Global Association (GA)

The problem that Geiger's method doesn't address is determining the set of arrivals generated by the same event and phase labels for all these arrivals. The Global Association (GA) software (Le Bras et al., 1994a,b), currently deployed by the IDC, attempts to solve just this problem. In addition to associating arrivals and locating events, it also performs a series of post-processing on the generated event bulletin to compute for example the magnitude of the event and a probabilistic score of whether the event is genuine.

GA processing proceeds in a series of pipelined operations. After the signal processing has computed the basic parameters of the arrival – time, amplitude, azimuth, and slowness, the first step is to assign a phase label to the arrivals. The phase label is assigned using a number of fixed rules and also a neural network. The important point is that the phase label is not assigned based on a potential event hypothesis but the arrival alone. As will be shown later (see Figure 4.9), this initial label can be highly inaccurate.

An earlier preprocessing step has already identified a set of grid locations on the surface of the earth, such that overlapping circles around these locations would cover the entire earth. These circles are called cells. Additional cell locations at depth are chosen in areas of known deep seismicity. For each cell, the preprocessing computes the probability of detection at each station. This probability is computed by simulating events within the cell. The magnitudes of these simulated events in turn are drawn from a user-specified exponential distribution. For each cell a list of "first-arrival stations" is computed.

The association step proceeds in two phases. In the first phase each of the precomputed cell is considered a potential event location. A driver, or the first arrival of the station, is searched on stations in the precomputed list of "first arrival stations". Subsequently, corroborating arrivals are searched in other stations. Preliminary confirmation tests are performed on this event based on the weighted count of corroborating observations – time, azimuth, and slowness. The first phase can create multiple incompatible events, i.e. events sharing the same arrivals or multiple arrivals at a station for an event with the same phase label. These incompatibilities are resolved in the second phase. In this phase, events may be split or merged as needed to avoid inconsistencies. A network probability test further attempts to remove unlikely event hypotheses. At this point any outliers in terms of residuals are removed from the event and the location is recalculated using Geiger's method and all the currently associated arrivals. Further confirmation tests are performed using the newly

computed event location as well as additional features like magnitude and the error ellipse to determine if the event is indeed real.

The next step in the pipeline removes conflicts between events, for example an arrival being shared by two different events. These conflicts can also arise as a result of the association step being performed in parallel on different time intervals. The conflicts are resolved by assigning the arrival to the better event. Various tests are available to determine the best event, for example by counting the number of arrivals or the size of the error ellipse.

Finally a refinement phase adds late-arriving phases or other phases missed by the processing so far. The event location is recomputed and many of the earlier confirmation tests are repeated.

2.3.1 False Event Identification (FEI)

The FEI software of Mackey et al. (2009) uses a Support Vector Machine (SVM) (Vapnik, 1998) to classify true and false events. This has been deployed at the IDC since 2012 to help the analysts quickly determine if an event should be discarded. The SVM is trained on many features of the event and the associated arrivals, for example the number of arrivals, the error ellipse etc. The performance of FEI is, of course, limited by the performance of GA (or SEL3, the name given to the output of GA) since FEI doesn't propose new events. These results can be seen in Figure 6.1 as the curve titled, "SEL3 extrapolation".

2.4 Multiple-event Location

Multiple-event location (Douglas, 1967) is another approach like FEI which seeks to improve an existing seismic bulletin. It doesn't create new events like GA, but instead relocates the events in the bulletin. The underlying idea is that although the travel times of seismic waves are unknown and can vary considerably even for events at similar depths and distances, the travel time from a particular region of the earth to a station tend to be quite similar. This similarity can be exploited to constrain the location of events clustered in a region. However, these constraints tend to lead to better relative rather than absolute locations. Various approaches are employed to improve the absolute locations as well, see for example Jordan and Sverdrup (1981), Pavlis and Booker (1983), and Dewey (1972).

Waldhauser and Ellsworth (2000) dispense with the need to compute the arrival times by directly using waveform correlations to determine the differences in travel times between stations. This method can lead to much more accurate relative location, but its applicability is limited to a narrower geographic region where the correlation is expected to be higher.

2.4.1 BAYHLoc

A more direct approach to exploit the similarity in travel time of events from the same region is to develop a model of spatially varying traveltime corrections. This is the approach taken by Schultz et al. (1998) and Myers and Schultz (2000). However, it requires a carefully curated dataset with accurate phase labels to train the corrections. A more sophisticated statistical technique is employed by Myers et al. (2007) also known as BAYHLoc, wherein instead of learning the correction terms, a prior is placed on them, and they are inferred directly from the data. A prior is also placed on the true phase label so that the phase labels can be corrected, if needed, or a phase could be dropped altogether. Another important idea of this paper is to use Markov Chain Monte Carlo (MCMC) (Andrieu et al., 2003) inference to locate the event and its associated location uncertainty, rather than the iterative linear least squares method described earlier.

Since BAYHLoc has a probabilistic model which is similar in some aspects to our model we will describe it in some detail. The following notation will help in the exposition.

- e_l^i is the location(longitude and latitude) of the ith event. For $1 \leq i \leq |e|$.
- e_t^i is the time of the ith event.
- e_d^i is the depth of the ith event.
- e is the set of all e_l^i, e_d^i and e_t^i.
- a_t^{ijk} is the arrival time of the jth phase from event i at station k as reported in the original event bulletin. For $1 \leq j \leq J$ and $1 \leq k \leq K$. Assume that $a_t^{ijk} = 0$ if the bulletin doesn't associate the jth phase for event i at station k.
- a is the set of all a_t^{ijk}.
- Λ_h^{ijk} is the true phase for arrival a^{ijk} (assuming $a_t^{ijk} \neq 0$). If $\Lambda_h^{ijk} = 0$ then a^{ijk} is not a valid arrival for event i.
- Λ_t^{ijk} is the true arrival time for event i's phase Λ_h^{ijk} at station k (assuming $\Lambda_h^{ijk} \neq 0$).
- Λ is the set of all Λ_h^{ijk} and Λ_t^{ijk}.
- Δ_{ik} is the distance between the ith event and the kth station.
- $I_T^j(d, \Delta)$ is the travel time table estimate for the travel time of the jth phase for an event at depth d and distance Δ from a station.

The BAYHLoc model describes the joint distribution $p(e, \Lambda, a)$ and uses this to compute the posterior $p(e, \Lambda \mid a)$. The various parts of the joint distribution are as follows.

$$\Lambda_t^{ijk} = e_t^i + I_T^{\Lambda_h^{ijk}}(e_d^i, \Delta_{ik}) + \delta^{ijk}.$$

Where δ^{ijk} is a path-specific travel time correction term given by,

$$\delta^{ijk} = \alpha^j + \beta^j \Delta_{ik}.$$

Here α^j and β^j are phase-specific corrections to the travel time curve. The prior distribution of these two terms for each phase are drawn from independent Gaussians.

$$a_t^{ijk} = \Lambda_t^{ijk} + \epsilon^{ijk}.$$

Here ϵ^{ijk} captures both the path-specific residual corrections as well as the arrival time pick errors. This term is drawn from a normal distribution whose precision (inverse variance) is factored into event-, phase-, and station-specific terms.

$$\begin{aligned} \epsilon^{ijk} &\sim \mathcal{N}(0, \sqrt{V^{ijk}}) \\ V^{ijk^{-1}} &= \phi_e^i \phi_h^j \phi_s^k \end{aligned}$$

All three of these factors are drawn from gamma distributions. The phase-specific factor is specified by a gamma with given hyperparameters ν_h^j and λ_h^j,

$$\phi_h^j \sim \Gamma(\nu_h^j, \lambda_h^j).$$

The other two factors of the precision are also gamma distributions but follow a constraint that the mode is one.

$$\begin{aligned} \phi_h^j &\sim \Gamma(\nu_e^i(\lambda_e^i), \lambda_e^i). \\ \phi_h^j &\sim \Gamma(\nu_s^k(\lambda_s^k), \lambda_s^k). \end{aligned}$$

The rate parameters λ_e^i and λ_s^k are drawn from vague exponential priors. The only remaining parts of the model are the priors on event location and time which are drawn independently from Gaussian distributions.

For inference, BAYHLoc combines a number of different MCMC moves. The event location e_l^i and e_d^i are updated by a Metropolis-Hastings (MH) move which draws a new location and depth from a proposal distribution centered around the current location. This update is accepted or rejected using the usual MH rule. The event time e_t^i as well as α^j and β^j can be sampled using the more efficient Gibbs sampler. This relies on the fact that each of these terms has a Gaussian prior and likelihood which makes it possible to compute the conditional posterior in closed form. Finally the precision factors ϕ_e^i, ϕ_h^j, ϕ_s^k as well as the rate parameters λ_e^i and λ_s^k are updated using slice sampling (Neal, 2003).

2.5 Data Association and Open Universe Probability Models

Data association is the general problem of associating observations to unknown objects. For example in the case of multi-target tracking (Blackman, 1986; Bar-Shalom and Fortmann, 1988) it is the task of associating observed radar blips to unobserved aircraft. Probability models for such tasks have been proposed as early as in the work of Sittler (1964). In the context of seismology, the problem is to associate the observed seismic arrivals with unknown seismic events.

Open Universe Probability Models (OUPMs) provide a rigorous statistical basis for describing data association probability models, see for example Milch et al. (2005a). OUPMs can describe situations with unknown number of objects with relational or identity uncertainty and as such are an extension of Bayes Nets (Pearl, 1988). Various languages and generic inference algorithms have been proposed for OUPMS. IBAL (Pfeffer, 2001, 2007) was the first such language based on a lisp-like syntax. Its inference engine is based mainly on forward sampling. Bayesian Logic (BLOG) (Milch et al., 2005a,b; Milch and Russell, 2006) is a language based on first-order logic and with a more declarative syntax. The inference engine of BLOG combines forward sampling with the ability to provide custom proposal distributions. Church (Goodman et al., 2008) is another lisp-like language with a forward sampling inference engine. More efficient inference by using Gibbs sampling (Geman and Geman, 1984) in the context of OUPMs has been described in Arora et al. (2010b,c).

2.6 Tempered Inference

In many probabilistic inference algorithms which are based on local moves hard deterministic constraints can hamper the ability of the moves to draw samples from the distribution or search for the mode. In essence, a local change to the current state can cause a global constraint to be violated and hence lead to a rejected state. A common solution to this problem is to somehow quantify the violation of the constraint and then temper it. Tempering is an idea which involves the use of a temperature parameter that is gradually decreased. It was originally used in simulated annealing (Kirkpatrick et al., 1983) which searches for the minimum of a cost function. The technique allows uphill moves (or moves that increase the violation of a global constraint) at high temperature and gradually discourages such moves at lower temperatures. In a probabilistic setting, simulated tempering (Marinari and Parisi, 1992; Geyer and Thompson, 1995) extends this idea to drawing samples from a probability distribution $p(x)$ by exponentiating the distribution to $p(x)^{T^{-1}}$, where T is the current temperature. The motivation for tempering, similar to annealing, is that at higher temperatures the probability distribution, $p(x)^{T^{-1}}$ approaches the uniform distribution, in a sense it gets heated up, and mixing is much faster.

Chapter 3

Model Motivation

In order to give some intuition into our integrated probabilistic model of seismology we will present some plots of the posterior distribution in a simplified 2-D model of the world. Assume a world in which there is exactly one seismic event which occurs at time 0 somewhere on the unit square with equal probability. The event may or may not be detected at each of four seismic stations independently. Figure 3.1 shows the 2-D world and the location of the stations are marked on it as triangles.

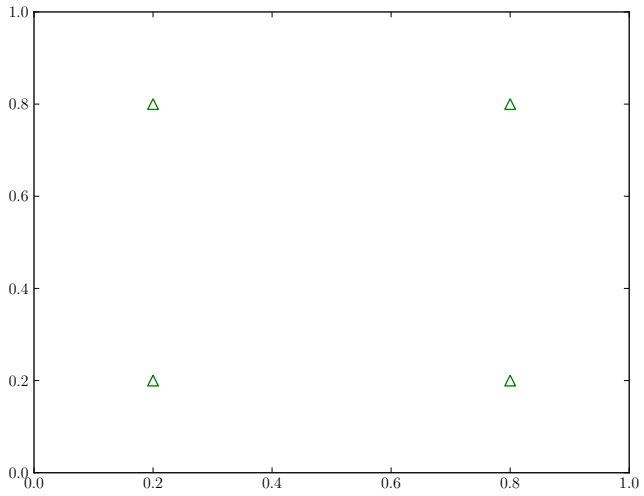

Figure 3.1: 2-D world with seismic stations (triangles)

Now, consider that the detection probability depends only on the distance between the event location and the station and this dependence is $exp(-distance)$. In this world if all we

know is that three of the stations detected the event then based entirely on the information from the 3 detections and the 1 misdetection we can get an approximate idea of the event location. Figure 3.2 shows the posterior location density assuming that the station on the top right is the only one that misses the event. This posterior is concentrated on the station at the lower left since this is farthest from the station which missed the event. Note that this posterior density is very weak. Even at its peak the posterior is only twice the uniform density.

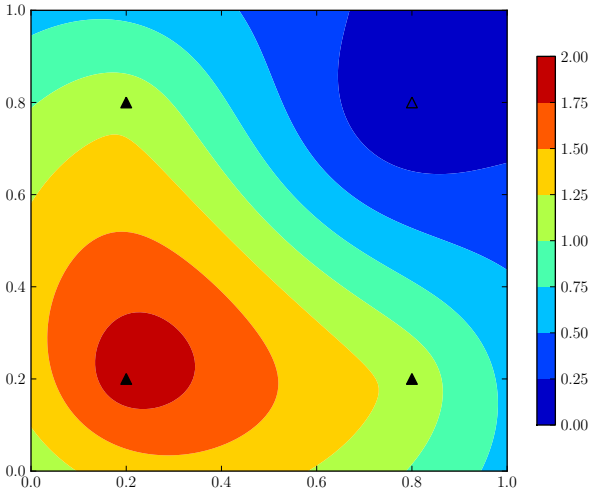

Figure 3.2: Posterior event location density given that one stations missed it.

Now, let's assume that we observe the arrival time of the seismic waves at the stations as well. This arrival time is based on a very simple travel time function plus a Gaussian residual distribution. We will assume that the seismic waves travel at a unit speed and the residual distribution has a standard deviation of 0.1. Consider again the case where exactly three stations detect the event and the arrival time is exactly equal to the theoretical time. Figure 3.3 shows this posterior distribution. The true event location is marked with an x and the misdetecting station is shown with a hollow triangle. This posterior distribution is now much more tightly concentrated and goes up to 24 times the uniform distribution. Notice also that this posterior is slightly shifted away from the true location in the opposite direction of the misdetecting station, as expected.

Now, if we had more precise travel time information, or in other words if the standard deviation of the residual was further reduced, then our posterior location density would be even further concentrated around the true location. To demonstrate this we repeated the previous experiment with half the standard deviation, i.e 0.05. This result is shown in

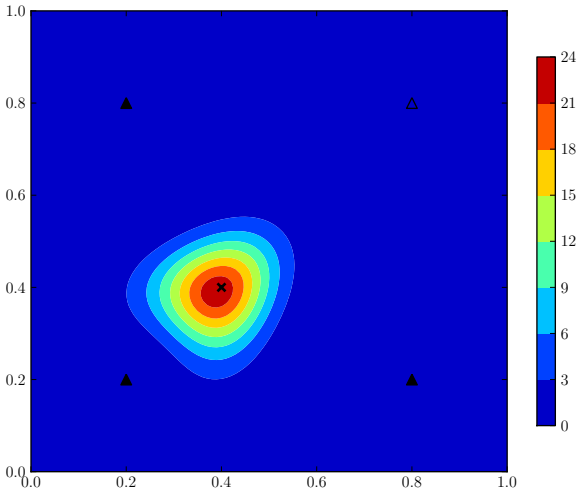

Figure 3.3: Posterior location uncertainty with arrival time uncertainty of 0.1.

Figure 3.4. The posterior now peaks at 105 times the uniform distribution.

However, if we reduce the number of detecting stations while still keeping the reduced arrival time uncertainty, the event location is no longer well constrained. In fact, as Figure 3.5 shows the posterior is now multi-modal. Also, the higher mode is the one which is farther away from the misdetecting stations.

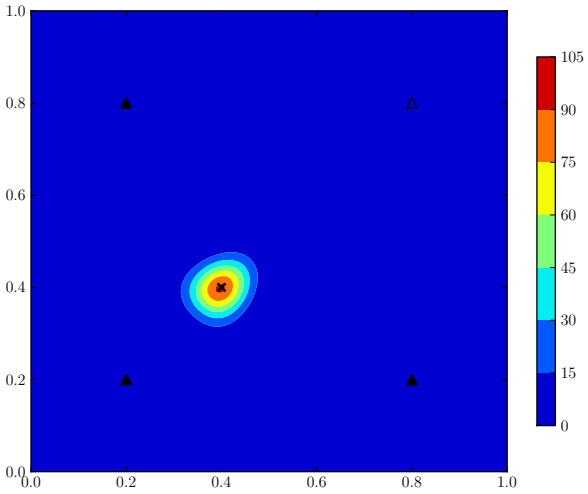

Figure 3.4: Posterior location uncertainty with arrival time uncertainty of 0.05.

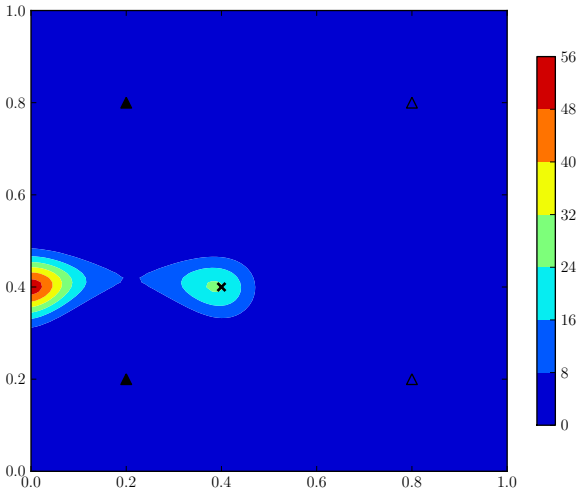

Figure 3.5: Posterior location uncertainty with two stations failing to detect.

Chapter 4

Generative Probabilistic Model

Our generative model for seismic events and arrivals follows along the lines of the aircraft detection model in (Milch et al., 2005a, Figure 3). In our model, there is an unknown number of seismic events with unknown parameters (location, time, etc.). These events produce 14 different types of seismic waves or phases. A phase from an event may or may not be detected by a station. If a phase is detected at a station, a corresponding arrival is generated. However, the parameters of the arrival may be imprecise. Additionally, an unknown number of noise arrivals are generated at each station.

4.1 Events

In the following, we only consider events with body-wave magnitude, m_b, 2 or higher. All other events are considered noise.

4.1.1 Event Rate and Time

The events are generated by a time-homogeneous Poisson process with a rate parameter, λ_e. If e is the set of events (of size $|e|$), and T is the time period under consideration (in seconds), we have

$$P_{\theta,n}(|e|) = \frac{(\lambda_e \cdot T)^{|e|} \exp(-\lambda_e \cdot T)}{|e|!}. \tag{4.1}$$

For each event, e^i, the event time, e^i_t, is uniformly distributed between 0 and T, i.e.

$$P_{\theta,t}(e^i_t) = \frac{1}{T}. \tag{4.2}$$

The parameter λ_e is estimated from the average historical event frequencies as shown in Figure 4.1. The estimated value of λ_e is 0.001266 per second or 4.6 per hour.

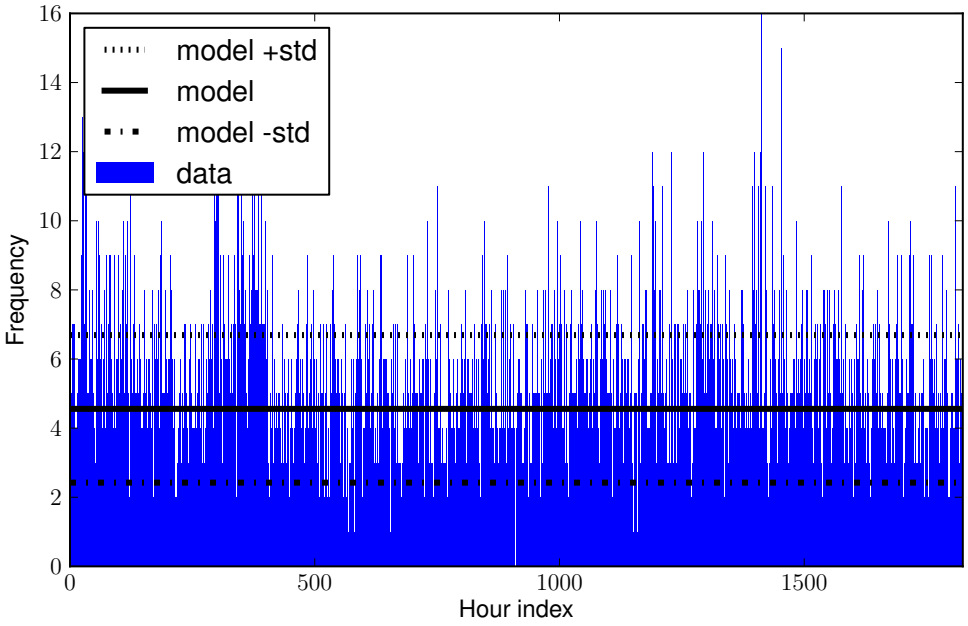

Figure 4.1: Estimating event rate.

4.1.2 Event Location

The longitude and latitude of the ith event, e_l^i are drawn from an event location density, $P_{\theta,l}(e_l)$ on the surface of the earth. This density is a mixture of a uniform density (to allow for explosions anywhere on the surface) as well as a kernel density estimate using historical event locations.

$$P_{\theta,l}(e_l) = .001\frac{1}{4\pi R^2} + .999\frac{1}{H}\sum_{h=1}^{H} K_{b,g_l^h}(e_l)\,,$$

where R is the earth's radius, H is the number of historical events, $g^1, \ldots, g^H$, and $K_{b,x}(\cdot)$ is a kernel function defined over the surface of the earth.

$$K_{b,x}(y) = \frac{1 + 1/b^2}{2\pi R^2}\frac{\exp\left(-\Delta_{xy}/b\right)}{1 + \exp\left(-\pi/b\right)}\,,$$

where $b > 0$ is the bandwidth of the kernel and Δ_{xy} is the distance between locations x and y. Both b and Δ_{xy} are expressed in radians. Notice that in the limit as $b \to \infty$, the kernel tends to the uniform density over the earth's surface.

The optimal value of b is chosen by leave-one-out cross validation (LOOCV) over a random subset of 1000 event locations in the dataset. We use grid search to find the value of b within 0 and 2 which maximizes the LOOCV value. We initially use a grid size of 0.2

and refine with a grid size of 0.05. Figure 4.2 plots the results of this search. The best value of b that is chosen is 0.7. In Figure 4.3 we show the event location prior $log(P_{\theta,l}(\cdot))$ using all of the training dataset and the optimal bandwidth.

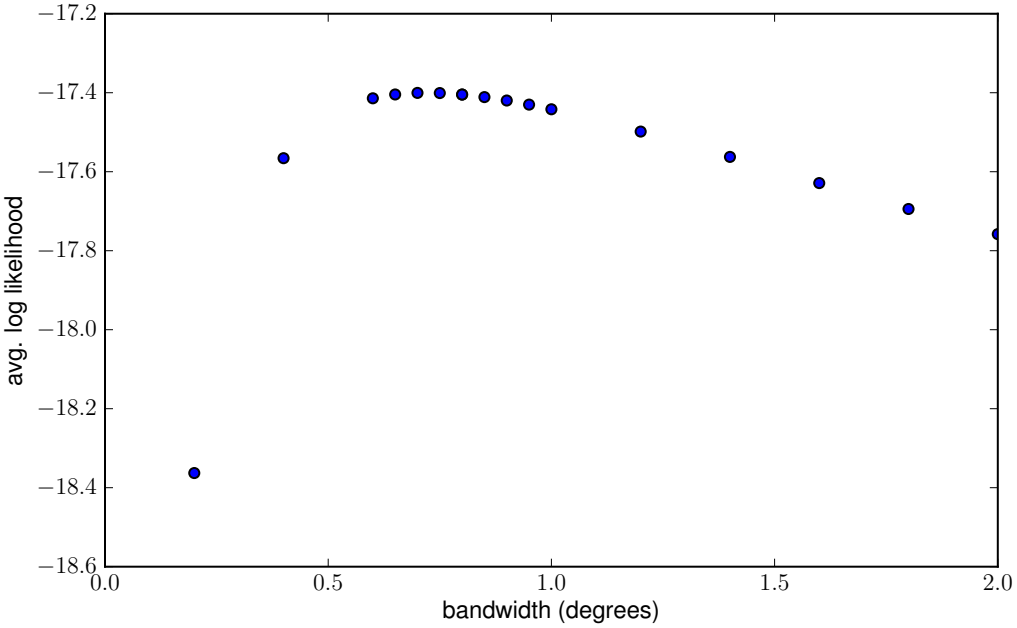

Figure 4.2: Event location average leave-one-out log likelihood vs. bandwidth.

The depth of the event, e_d^i is uniformly distributed up to a maximum depth D (700 km in our experiments),

$$P_{\theta,d}(e_d^i) = \frac{1}{D}. \tag{4.3}$$

4.1.3 Event Magnitude

The Gutenberg-Richter law (Gutenberg and Richter, 1954) posits that the number of events with magnitude m or more is ten times the number of events with magnitude $m+1$ or more. In terms of the event magnitude probability of an arbitrary event: $P_{\theta,m}(e_m^i >= m) = 10 \cdot P_{\theta,m}(e_m^i >= m+1)$. We represent this prior knowledge in our model with an exponential distribution on e_m^i with rate $\lambda_m = log(10)$:

$$P_{\theta,m}(e_m^i) = \lambda_m \exp(-\lambda_m \cdot (e_m^i - 2)). \tag{4.4}$$

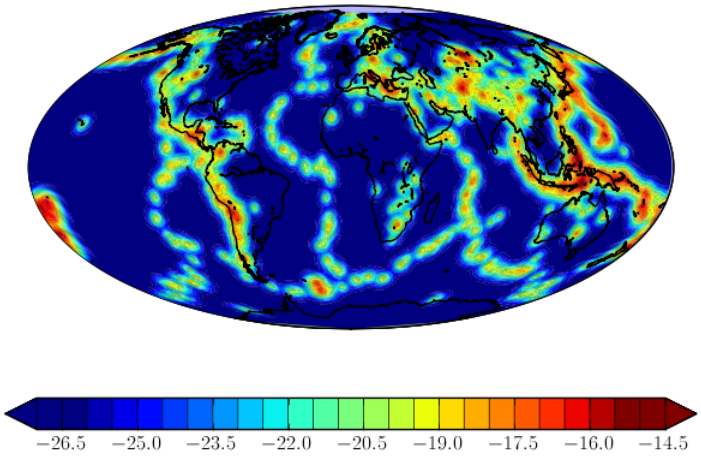

Figure 4.3: Event location log density.

4.1.4 Overall Event Prior

Under the assumption that the event location, depth, time, and magnitude are independent of each other:

$$P_\theta(e^i) = P_{\theta,l}(e^i_l)P_{\theta,d}(e^i_d)P_{\theta,t}(e^i_t)P_{\theta,m}(e^i_m)$$

Substituting from Equations 4.2, 4.3, and 4.4, we get

$$P_\theta(e^i) = P_{\theta,l}(e^i_l)\frac{1}{D}\frac{1}{T}\lambda_m \exp\left(-\lambda_m \cdot (e^i_m - 2)\right) . \tag{4.5}$$

In our model, all the events are exchangeable and are generated independently, thus

$$P_\theta(e) = P_{\theta,n}(|e|) \cdot |e|! \cdot \prod_{i=1}^{|e|} P_\theta(e^i) .$$

Note, we are overloading $P_\theta(\cdot)$ to refer to the distribution over the set of events as well as the distribution of a single event. Substituting from Equations 4.1 and 4.5,

$$P_\theta(e) = \exp\left(-\lambda_e \cdot T\right) \prod_{i=1}^{|e|} P_{\theta,l}(e^i_l)\frac{1}{D}\lambda_e\lambda_m \exp\left(-\lambda_m \cdot (e^i_m - 2)\right) . \tag{4.6}$$

We define,

$$p_\theta(e^i) = P_{\theta,l}(e^i_l)\frac{1}{D}\lambda_e\lambda_m \exp\left(-\lambda_m \cdot (e^i_m - 2)\right).$$

Thus simplifying Equation 4.6 to

$$P_\theta(e) = \exp\left(-\lambda_e \cdot T\right)\prod_{i=1}^{|e|} p_\theta(e^i). \tag{4.7}$$

4.2 Correct Arrivals

4.2.1 Detection Probability

The probability that an event i's j^{th} phase, $1 \leq j \leq J$, is detected by a station k, $1 \leq k \leq K$, depends on the phase, the station, and the event's magnitude, depth, and distance to the station. Let $P^{jk}_{\phi,d}(e_i)$ be the probability of this detection. The phase- and station-specific detection distributions, $P^{jk}_{\phi,d}(\cdot)$, were obtained using logistic regression models. In other words, the log odds of detecting an arrival is a linear function of features defined on e^i_m, e^i_d, and Δ_{ik}, the distance between event i and station k.

$$log\left(\frac{P^{jk}_{\phi,d}(e^i)}{1 - P^{jk}_{\phi,d}(e^i)}\right) = \sum_{w\in\mathcal{F}_d} \mu^{wjk}_d \cdot w(e^i_m, e^i_d, \Delta_{ik})$$

Where $\mathcal{F}_d \subseteq \mathcal{R}^3 \rightarrow \mathcal{R}$ is a set of features, and μ^{wjk}_d is the weight for the feature w. The complete set of features is defined in Table 4.1. The figure also shows some validation results for each feature. We have trained a logistic regression model on the first 1000 hours of training data for a single station, Alice Springs. Then we compute the average log likelihood of the subsequent 100 hours of training data using the learned model. This computation is repeated for increasing number of features and corresponding likelihood numbers are shown in the figure.

Directly estimating the feature weights μ^{wjk}_d is not possible since many of the station-phase combinations have very little data. To deal with this data sparsity we used a hierarchical Bayesian procedure Gelman et al. (2004), which posits that the weight for a station-phase is drawn from a global prior for that phase. This global prior is in turn drawn from a non-informative prior, as follows:

$$\begin{aligned}
\mu^{wjk}_d &\sim \mathcal{N}(\mu^{wj}_d, \sigma^{wj}_d) \\
\mu^{wj}_d &\sim \mathcal{N}(0, 100) \\
(\sigma^{wj}_d)^{-2} &\sim \Gamma(0.01, 100)
\end{aligned}$$

Feature	Value	Avg. Log Likelihood
(Intercept)	1	-0.271477614686
mag	e_m^i	-0.25283924485
depth	e_d^i	-0.22786110952
dist	Δ_{ik}	-0.148838588413
dist0	$\mathcal{N}(\Delta_{ik}, 0, 5)$	-0.145619510179
dist35	$\mathcal{N}(\Delta_{ik}, 35, 20)$	-0.143358785181
dist40	$\mathcal{N}(\Delta_{ik}, 40, 20)$	-0.137420683639
dist12520	$\mathcal{N}(\Delta_{ik}, 125, 20)$	-0.136943948297
dist12540	$\mathcal{N}(\Delta_{ik}, 125, 40)$	-0.132671965356
mag6	$\mathcal{N}(e_m^i, 6, 5.5)$	-0.132297181033
mag68	$\mathcal{N}(e_m^i, 6, 8)$	-0.131333296357
md	$(7 - e_m^i) \cdot \Delta_{ik}$	-0.129367021293

Table 4.1: List of features used for computing the probability of detecting an arrival from an event i with magnitude e_m^i, depth e_d^i, and distance Δ_{ik} from station k. Here $\mathcal{N}(x, \mu, \sigma)$ is the standard Gaussian density with mean μ and standard deviation σ measured at x. Also, the log likelihood from using that feature.

Where $\mathcal{N}$ and Γ are the Gaussian and the Gamma distributions parameterized by their mean and standard deviation, and shape and scale respectively. Estimation of parameters follows a coordinate ascent procedure. For each phase j, we initialize $\mu_d^{wj} = 0$ and $\sigma_d^{wj} = 1$, and then alternately optimize $\mu_d^{wjk} : \forall w, k$, $\mu_d^{wj} : \forall w$, and $\sigma_d^{wj} : \forall w$ till convergence. In each maximization step, the optimal value of μ_d^{wjk} is computed by second order optimization, while the rest of the values have a closed form solution.

In Figure 4.4 we show the model prediction for one phase at a station.

4.2.2 Arrival Attributes

If event i's j^{th} phase is detected by a station k, we define Λ^{ijk} as the corresponding arrival, otherwise $\Lambda^{ijk} = \zeta$. Our model specifies probability distribution for the attributes of this arrival: time – Λ_t^{ijk}, azimuth – Λ_z^{ijk}, slowness – Λ_s^{ijk}, amplitude – Λ_a^{ijk}, and phase – Λ_h^{ijk}. The arrival time is

$$\Lambda_t^{ijk} = e_t^i + I_T^j(e_d^i, \Delta_{ik}) + r_t^{ijk},$$

where I_T^j is the prediction from the IASPEI travel time model (a function of the event depth, and distance to the station), and r_t^{ijk} is a residual, distributed as a Laplacian,

$$r_t^{ijk} \sim Laplace(\mu_t^{jk}, b_t^{jk}),$$

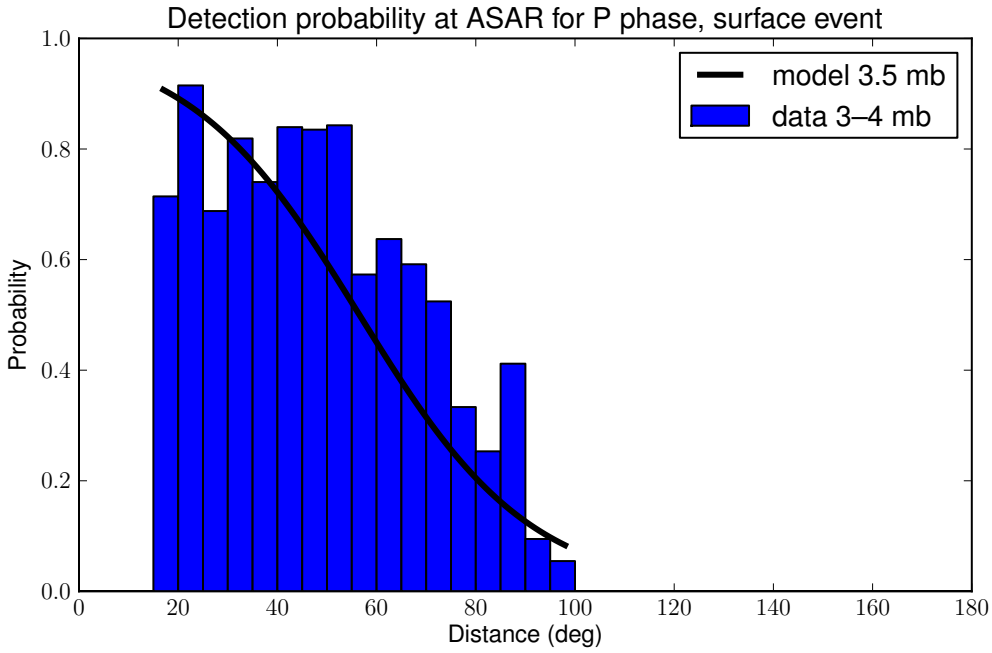

Figure 4.4: Conditional detection probabilities for the P phase of surface events between 3 and 4 mb at Station ASAR.

with mean[1] μ_t^{jk} and scale b_t^{jk}. The residual distribution accounts for the inhomogeneities in the earth's crust, which allow seismic waves to travel faster or slower than the IASPEI prediction. This distribution also accounts for any systematic biases in picking seismic onsets from waveforms. Thus, we get

$$P_{\phi,t}^{jk}(\Lambda_t^{ijk} \mid e^i) = \frac{1}{2b_t^{jk}} \exp\left(-\frac{|\Lambda_t^{ijk} - e_t^i - I_T(e_d^i, \Delta_{ik}) - \mu_t^{jk}|}{b_t^{jk}}\right).$$

Similarly, the arrival azimuth and slowness follow a Laplacian distribution.

$$P_{\phi,z}^{jk}(\Lambda_z^{ijk} \mid e^i) = \frac{1}{2b_z^{jk}} \exp\left(-\frac{|\psi(\Lambda_z^{ijk}, G_z(s_l^k, e_l^i)) - \mu_z^{jk}|}{b_z^{jk}}\right),$$
$$P_{\phi,s}^{jk}(\Lambda_s^{ijk} \mid e^i) = \frac{1}{2b_s^{jk}} \exp\left(-\frac{|\Lambda_s^{ijk} - I_s^j(e_d^i, \Delta_{ik}) - \mu_s^{jk}|}{b_s^{jk}}\right).$$

[1]The usual way to parametrize a Laplacian is by its location and scale. However, we use the terms mean and scale to avoid confusion with event locations. Fortunately, the mean of a Laplacian coincides with its location.

Where ψ computes the difference in the observed angle Λ_z^{ijk} and the angle computed from the geographical function G_z, which depends on the station location, s_l^k, and the event location, e_l^i. Also, I_s^j is the slowness value computed from the IASPEI model for phase j. It should be noted that the observed values of azimuth and slowness referred to above already include certain station-specific corrections called SASC (Slowness Azimuth Site Correction). The estimation of all the station-and-phase-specific means and scales, for example μ_t^{jk} and b_t^{jk}, is based on a hierarchical model. In this model, each mean and scale is generated from a phase-global prior, which we describe below for the arrival time.

$$\begin{aligned} r_t^{ijk} &\sim Laplace(\mu_t^{jk}, b_t^{jk}) \\ \mu_t^{jk} &\sim Laplace(\mu_t^{j}, b_t^{j}) \\ (b_t^{jk})^{-1} &\sim \Gamma(1, \beta_t^{j}) \\ \mu_t^{j} &\sim Laplace(0, 100) \\ (b_t^{j})^{-1} &\sim \Gamma(.01, 100) \\ (\beta_t^{j})^{-1} &\sim \Gamma(.01, 100) \end{aligned}$$

For each phase j, we start by initializing $\mu_t^{jk} = 0$, $b_t^{jk} = 1$ for all stations k, $\mu_t^j = 0$, $b_t^j = 1$, and $\beta_t^j = 1$. Next, we iteratively optimize the values of μ_t^{jk}, b_t^{jk} for each station, and the globals μ_t^j, b_t^j, and β_t^j. This is repeated till convergence. Each of these optimization steps has a simple closed-form solution. A very similar procedure is adopted for estimating the azimuth parameters, μ_z^{jk} and b_z^{jk}, and the slowness parameters, μ_s^{jk}, and b_s^{jk}. An example for each of these three types of distributions is shown in Figures 4.5, 4.6, and 4.7.

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

$$\begin{aligned} r_a^{ijk} &= log(\Lambda_a^{ijk}) - \sum_{w \in \mathcal{F}_a} \mu_a^{wjk} \cdot w(e_m^i, e_d^i, I_T^j(e_d^i, \Delta_{ik})), \\ r_a^{ijk} &\sim \mathcal{N}(0, \sigma_a^{jk}), \end{aligned}$$

where $\mathcal{F}_a$ is a set of feature functions (see Table 4.2), and μ_a^{wjk} is the weight for feature w. This implies that

$$P_{\phi,a}^{jk}(\Lambda_a^{ijk} \mid e^i) = \frac{1}{\sqrt{2\pi}\sigma_a^{jk}} \exp\left(-\frac{(r_a^{ijk})^2}{2\sigma_a^{jk^2}}\right) \frac{1}{\Lambda_a^{ijk}} \ .$$

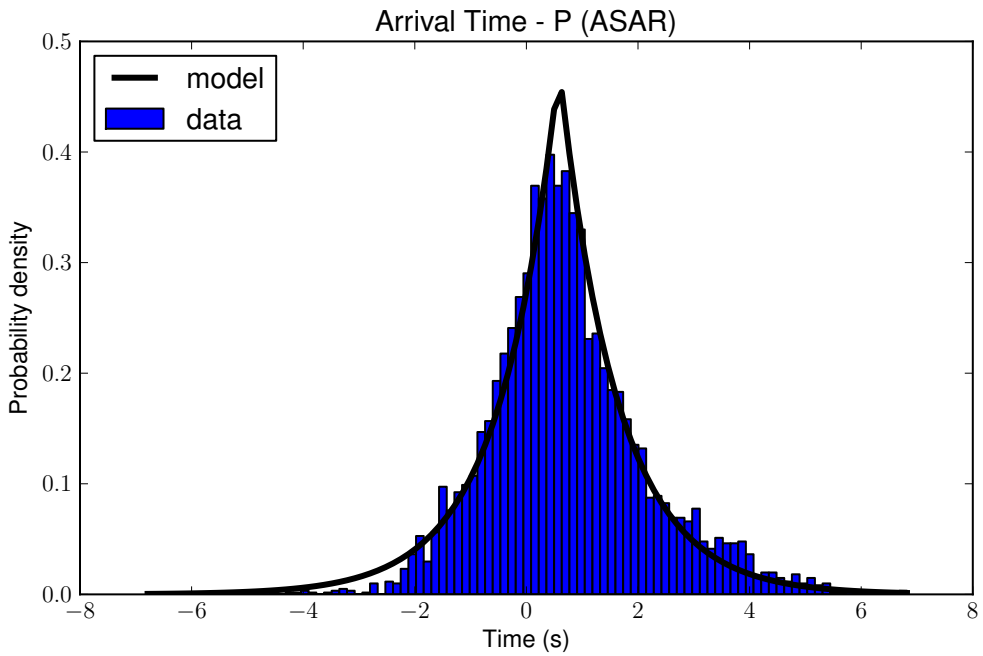

Figure 4.5: Arrival time distribution around the IASPEI prediction for P phase arrivals at station ASAR.

In order to estimate the feature weights, we use a hierarchical model, as before, which assumes that for each phase the feature weights at a station are drawn from a global prior.

$$\begin{aligned}
\mu_a^{wjk} &\sim \mathcal{N}(\mu_a^{wj}, \sigma_a^{wj}) \\
(\sigma_a^{jk})^{-2} &\sim \Gamma(100, \beta_a^j) \\
\mu_a^{wj} &\sim \mathcal{N}(0, 100) \\
(\sigma_a^{wj})^{-2} &\sim \Gamma(.01, 100) \\
(\beta_a^j)^{-1} &\sim \Gamma(.01, 100)
\end{aligned}$$

Maximum a-posteriori (MAP) inference of these weights is along the lines of the method already described for other parameters. Figure 4.8 has an example of a learned model. Table 4.2 shows the successive improvement in log likelihood on held-out data after adding each feature to the model.

Finally, the phase label Λ_h^{ijk}, which is automatically assigned to the arrival, follows a multinomial distribution whose parameters depends only on the true phase, j:

$$P_{\phi,h}^{jk}(\Lambda_h^{ijk} \mid e^i) = p_h^j(a_h^{ijk}).$$

Learning the multinomial distributions p_h^j is a simple matter of counting with add-one smoothing. The learned distribution is plotted as a heat map in Figure 4.9.

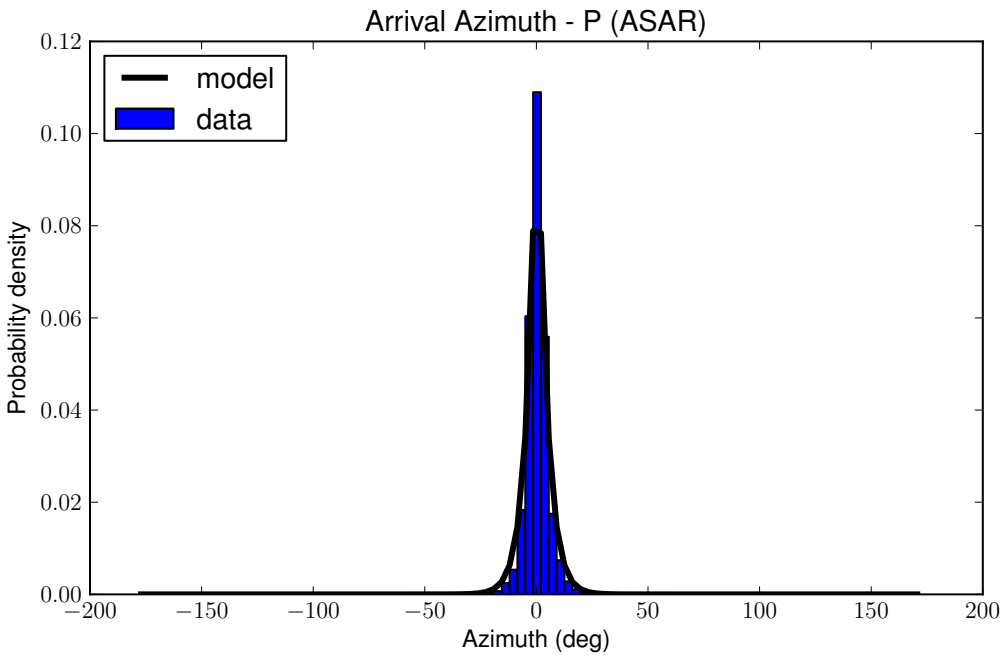

Figure 4.6: Arrival azimuth distribution around the geographical prediction (plus SASC correction) for P phase arrivals at station ASAR.

4.2.3 Overall Correct Arrivals

We assume that all the attributes of an arrival are independent of each other. Thus,

$$P_{\phi}^{jk}(\Lambda^{ijk}|e^i) = P_{\phi,t}^{jk}(\Lambda_t^{ijk} \mid e^i) P_{\phi,z}^{jk}(\Lambda_z^{ijk} \mid e^i) P_{\phi,s}(\Lambda_s^{ijk} \mid e^i) P_{\phi,a}^{jk}(\Lambda_a^{ijk} \mid e^i) P_{\phi,h}^{jk}(\Lambda_h^{ijk} \mid e^i).$$

Further, assuming that the arrivals are all independent, we have,

$$P_{\phi}(\Lambda \mid e) = \prod_{i=1}^{|e|} \prod_{j=1}^{J} \prod_{k=1}^{K} \left[\mathbb{1}(\Lambda^{ijk} = \zeta)(1 - P_{\phi,d}^{jk}(e^i)) + \mathbb{1}(\Lambda^{ijk} \neq \zeta) P_{\phi,d}^{jk}(e^i) P_{\phi}^{jk}(\Lambda^{ijk} \mid e^i) \right]. \tag{4.8}$$

4.2.4 Model Validation

We present some results justifying the model choices made for the arrival parameters. The arrival time, azimuth, and slowness residuals are all modeled as Laplace distributions. This is in contrast to the standard seismological practice of modeling these as Gaussian residuals. In order to compare the effectiveness of the two distributions at modeling the residuals, we trained both distributions on a 1000 hour training data from station Alice Springs and

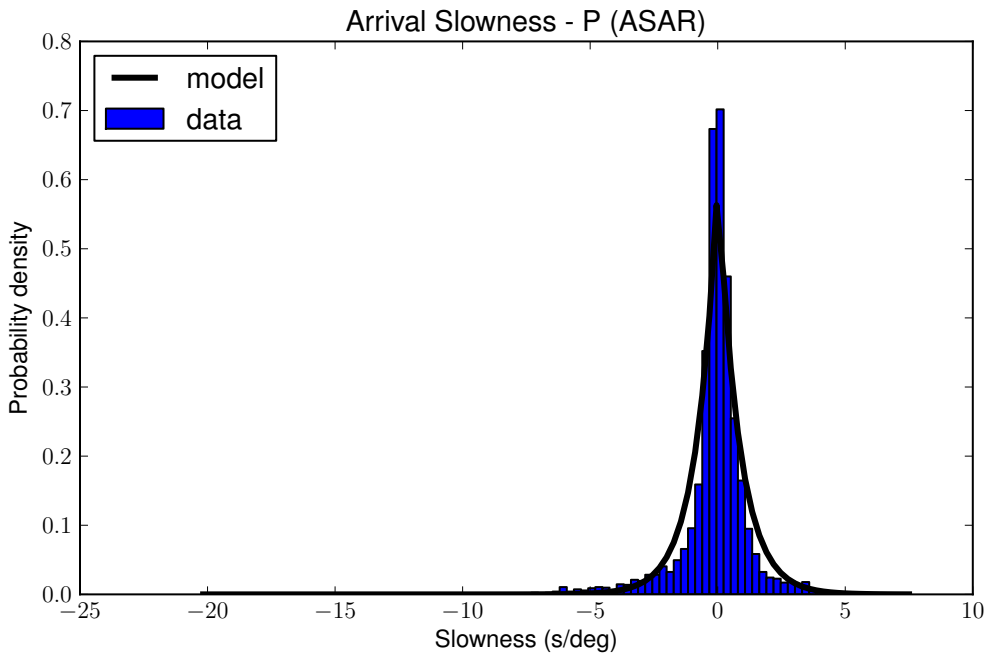

Figure 4.7: Arrival slowness distribution around the IASPEI prediction (plus SASC correction) for P phase arrivals at station ASAR.

evaluated on a separate 100 hours of data from the same station. The results of this validation are shown in Table 4.3. The table shows that the time residuals for the most common phases (P and S) are much better modeled as Laplacian. Similarly, for the azimuth and the slowness residuals the Laplace distribution explains the held-out data much better.

Another choice made in the model is to assume that the residuals are all independent. We have evaluated this decision based on the Pearson's correlation coefficient (Soper et al., 1917). Table 4.4 shows the correlation between the three residuals (time, azimuth, and slowness) for various phases. Since all but two of the residuals are less than 0.2 and there is no consistent pattern indicating positive or negative dependence, it appears reasonable to assume independence.

4.3 False Arrivals

The station, k, also generates a set ξ^k of false arrivals, i.e. not associated to any event phase, described by the distribution P_φ^k. The number of false arrivals is generated by a

Feature	Value	Avg. Log Likelihood
(Intercept)	1	-1.65893054005
mag	e^i_m	-1.50107477111
depth	e^i_d	-1.38073535802
ttime	$I^j_T(e^i_d, \Delta_{ik})$	-1.32228249442
ttime0	$\mathcal{N}(I^j_T(e^i_d, \Delta_{ik}), 0, 50)$	-1.29296470633

Table 4.2: List of features used for predicting the amplitude of a phase j arrival from an event i at a station k. Also, the log likelihood after adding each feature

Phase	Time Residual		Azimuth Residual		Slowness Residual	
	Laplace	Gaussian	Laplace	Gaussian	Laplace	Gaussian
P	-1.6405	-1.6698	-4.0487	-14.4774	-1.7369	-1.9064
Pn	-2.0521	-2.2165	-4.4192	-7.0387	-4.1790	-5.1656
PKP	-1.7046	-1.6426	-15.1104	-83.7955	-1.0028	-1.2878
S	-3.3560	-3.3879	-3.2432	-3.2826	-2.8765	-2.9791
PcP	-1.5537	-1.4841	-7.8644	-23.9002	-1.2650	-1.3540
ScP	-2.3160	-2.4292	-8.1886	-25.2298	-2.2920	-2.1730

Table 4.3: Average log likelihood on held-out data of a Laplacian and a Gaussian model.

time-homogeneous Poisson process with rate λ^k_f:

$$P^k_{\varphi,n}(|\xi^k|) = \frac{(\lambda^k_f \cdot T)^{|\xi^k|} \exp\left(-\lambda^k_f \cdot T\right)}{|\xi^k|!} \ .$$

The values of λ^k_f are displayed in Figure 4.12. If ξ^{kl} is one of this set of false arrivals, its time ξ^{kl}_t, azimuth ξ^{kl}_z, and slowness ξ^{kl}_s are generated uniformly over their respective ranges,

$$P^k_{\varphi,t}(\xi^{kl}_t) = \tfrac{1}{T}, \quad P^k_{\varphi,z}(\xi^{kl}_z) = \tfrac{1}{M_z}, \quad P^k_{\varphi,s}(\xi^{kl}_s) = \tfrac{1}{M_s},$$

where M_z and M_s are the range of values for azimuth and slowness respectively. The log amplitude of the false arrival is generated from a mixture of two Gaussians, which is estimated from the data with a standard Expectation Maximization (EM) procedure, and a flat uniform prior with probability 0.1. The resulting distribution, $p^k_{\varphi,a}(\cdot)$, at one station is displayed in Figure 4.10.

$$P^k_{\varphi,a}(\xi^{kl}_a) = p^k_{\varphi,a}(\log(\xi^{kl}_a))\frac{1}{\xi^{kl}_a}$$

Finally, the phase label ξ^{kl}_h assigned to the false arrival follows a multinomial distribution, $P^k_{\varphi,h}(\xi^{kl}_h)$ (See Figure 4.11) learned by add-one smoothing. Overall, assuming the false arrival

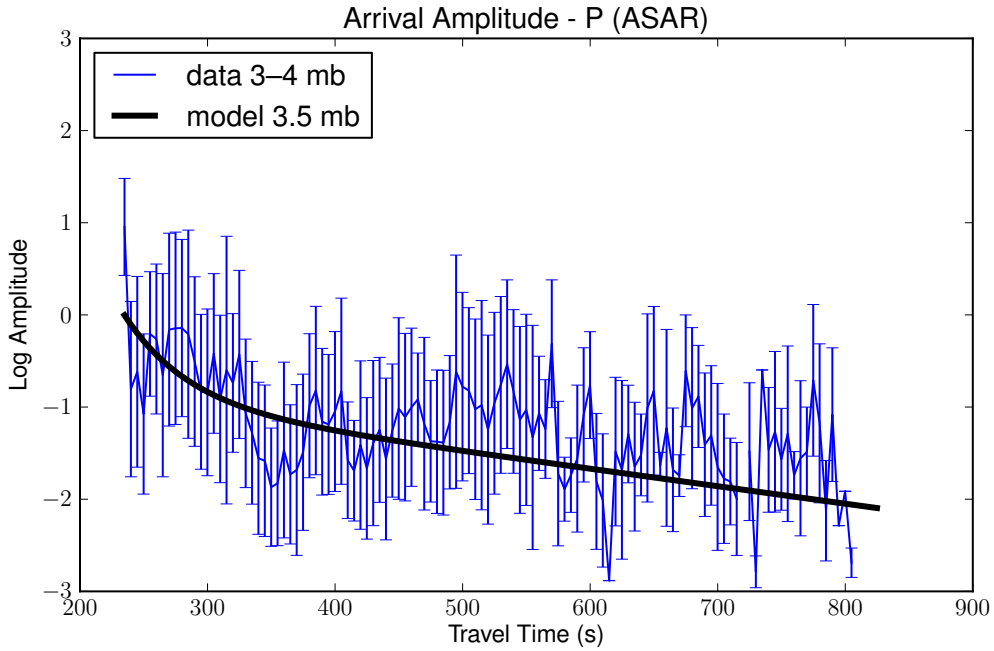

Figure 4.8: Arrival log amplitude distribution for the P phase of a surface event of 3–4 mb at station ASAR.

attributes are independently generated

$$P_\varphi^k(\xi^{kl}) = P_{\varphi,t}^k(\xi_t^{kl}) P_{\varphi,z}^k(\xi_z^{kl}) P_{\varphi,s}^k(\xi_s^{kl}) P_{\varphi,a}^k(\xi_a^{kl}) P_{\varphi,h}^k(\xi_h^{kl}).$$

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

$$\begin{aligned} P_\varphi^k(\xi^k) &= P_{\varphi,n}^k(|\xi^k|) \cdot |\xi^k|! \prod_{l=1}^{l=|\xi^k|} P_\varphi^k(\xi^{kl}) \\ &= \exp\left(-\lambda_f^k \cdot T\right) \prod_{l=1}^{l=|\xi^k|} \frac{\lambda_f^k}{M_z M_s} P_{\varphi,a}^k(\xi_a^{kl}) P_{\varphi,h}^k(\xi_h^{kl}) \,. \end{aligned}$$

Where, as before, we have overloaded P_φ^k to refer to a distribution over a set of arrivals as well as a single arrival. We define

$$p_\varphi^k(\xi^{kl}) = \frac{\lambda_f^k}{M_z M_s} P_{\varphi,a}^k(\xi_a^{kl}) P_{\varphi,h}^k(\xi_h^{kl}).$$

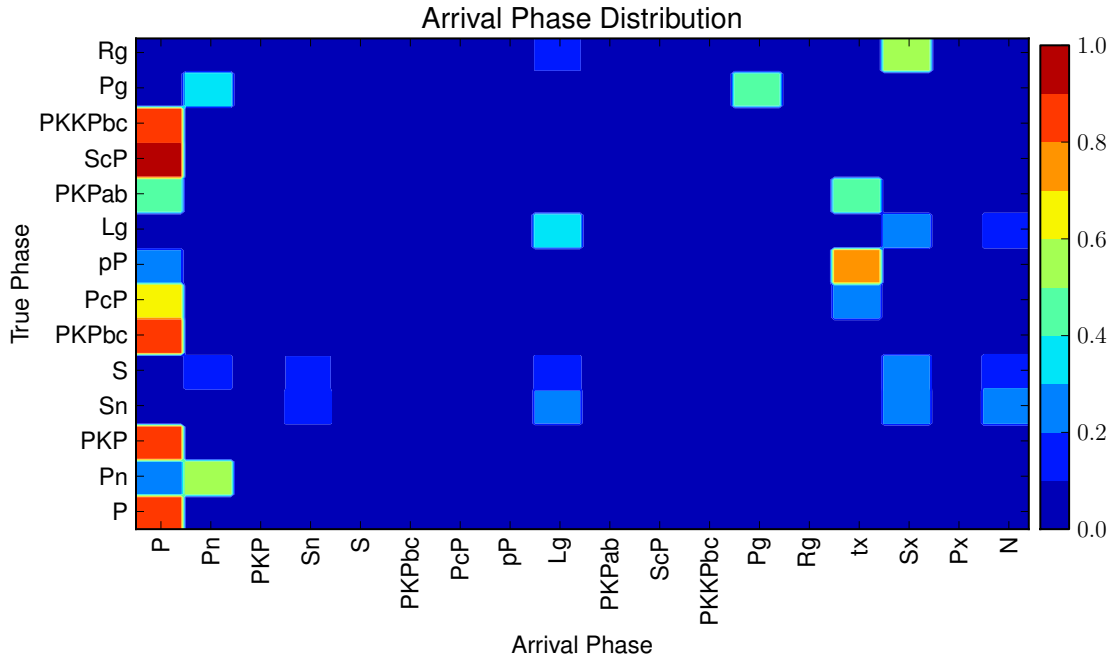

Figure 4.9: Arrival phase probability as function of the true phase.

Now, assuming that the false arrivals at different stations are independent of each other,

$$P_\varphi(f) = \exp\left(-\sum_{k=1}^{K} \lambda_f^k \cdot T\right) \prod_{k=1}^{K} \prod_{l=1}^{l=|\xi^k|} p_\varphi^k(\xi^{kl}). \tag{4.9}$$

4.4 Coda Arrivals

In our model, the scattered energy which follows a true or false arrival may trigger subsequent *coda* arrivals. For example the extra arrivals marked in Figure 2.1. It might appear that such arrivals are just ordinary false arrivals. However, unlike false arrivals which are generated from a mostly uniform prior, the coda arrivals are strongly correlated with the triggering arrival. If the coda arrivals are not modeled explicitly then our inference will end up hypothesizing additional spurious events as the most likely explanation for many of the coda arrivals.

We model the parameters of the coda arrival with a distribution P_γ, which is a function of the parameters of the previous triggering arrival. Whether the triggering arrival was a false arrival, or caused by an event, or itself triggered by another arrival, is immaterial. We define η^a as the previous arrival at the same station as arrival a, or ζ if there is no such previous arrival. Now, the probability that an arrival a is a coda arrival is given by $P_{\gamma,d}(\eta^a)$, which is a

Phase	Time and Slowness	Time and Azimuth	Slowness and Azimuth
P	0.0079	0.0893	0.1971
Pn	0.1015	0.1284	0.0842
PKP	0.0166	0.1419	0.1034
S	0.1222	0.3634	0.0693
PcP	-0.0905	-0.1624	0.6670
ScP	0.0645	-0.1292	0.1062

Table 4.4: Correlation between the various residuals

function of the amplitude of the previous arrival η^a, and we define $P_{\gamma,d}(\zeta) = 0$. We estimate $P_{\gamma,d}$ with a non-parametric model by discretizing the previous arrival's log amplitude into buckets of size .25 between -4 and 10. This distribution is displayed in Figure 4.13. Any points outside these extreme values are mapped to the nearest bucket.

One problem that arises while training the coda arrivals is that the IDC analysts don't annotate coda arrivals in the LEB bulletin making it hard to estimate the parameters of P_γ. Our solution is to heuristically annotate some of the unassociated arrivals as coda arrivals, and use this annotation to learn P_γ, and also P_φ. Our procedure is to search the training data for any unassociated arrivals within 30 seconds of a prior arrival at the same station and with an azimuth and slowness within 50 degrees and 10 seconds per degree respectively of the prior arrival's values, and to mark such arrivals as coda.

We model the distribution of the time delay between the coda arrival and the triggering arrival as a Gamma distribution, see Figure 4.14.

$$\begin{aligned} a_t - \eta_t^a &\sim \Gamma(\rho_t, \nu_t), \text{i.e.} \\ P_{\gamma,t}(a_t \mid \eta^a) &= \frac{1}{\Gamma(\rho_t)\nu_t^{\rho_t}}(a_t - \eta_t^a)^{\rho_t - 1} \exp(-\frac{a_t - \eta_t^a}{\nu_t}). \end{aligned}$$

The difference in azimuth, slowness, and log amplitude of the coda versus the triggering arrival are all modeled as Laplace distributions, see Figures 4.15, 4.16, and 4.17.

$$\begin{aligned} P_{\gamma,z}(a_z \mid \eta^a) &= \frac{1}{2\nu_z} \exp\left(-\frac{|\psi(a_z, \eta_z^a) - \rho_z|}{\nu_z}\right), \\ P_{\gamma,s}(a_s \mid \eta^a) &= \frac{1}{2\nu_s} \exp\left(-\frac{|a_s - \eta_s^a - \rho_s|}{\nu_s}\right), \text{and} \\ P_{\gamma,a}(a_a \mid \eta^a) &= \frac{1}{2\nu_a} \exp\left(-\frac{|\log(a_a) - \log(\eta_a^a) - \rho_a|}{\nu_a}\right) \frac{1}{a_a}, \end{aligned}$$

Finally, the coda phase is a multinomial distribution, $P_{\gamma,h}$ (this doesn't depend on the previous arrival's phase), see Figure 4.18. It is worth noting that the coda model is not station-specific, hence data sparsity is not a concern.

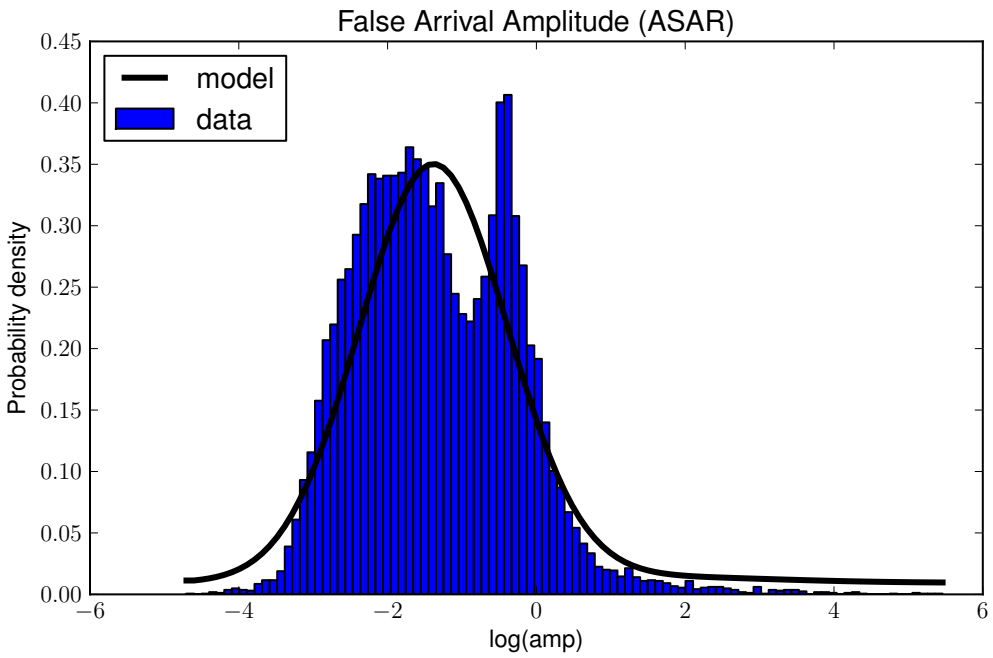

Figure 4.10: Amplitude distribution for false arrivals at stations ASAR.

Overall, assuming that all the parameters of the coda arrival are independent,

$$P_\gamma(a) = P_{\gamma,t}(a_t \mid \eta^a)\, P_{\gamma,z}(a_z \mid \eta^a)\, P_{\gamma,s}(a_s \mid \eta^a)\, P_{\gamma,a}(a_a \mid \eta^a)\, P_{\gamma,h}(a_h).$$

Further, if κ is the set of all coda arrivals, and A is the set of all observed arrivals then assuming that the coda arrivals are independent of each other,

$$P_\gamma(\kappa|A) = \prod_{a\in\kappa} P_{\gamma,d}(\eta^a) P_\gamma(a) \prod_{a\in A-\kappa} (1 - P_{\gamma,d}(\eta^a)). \tag{4.10}$$

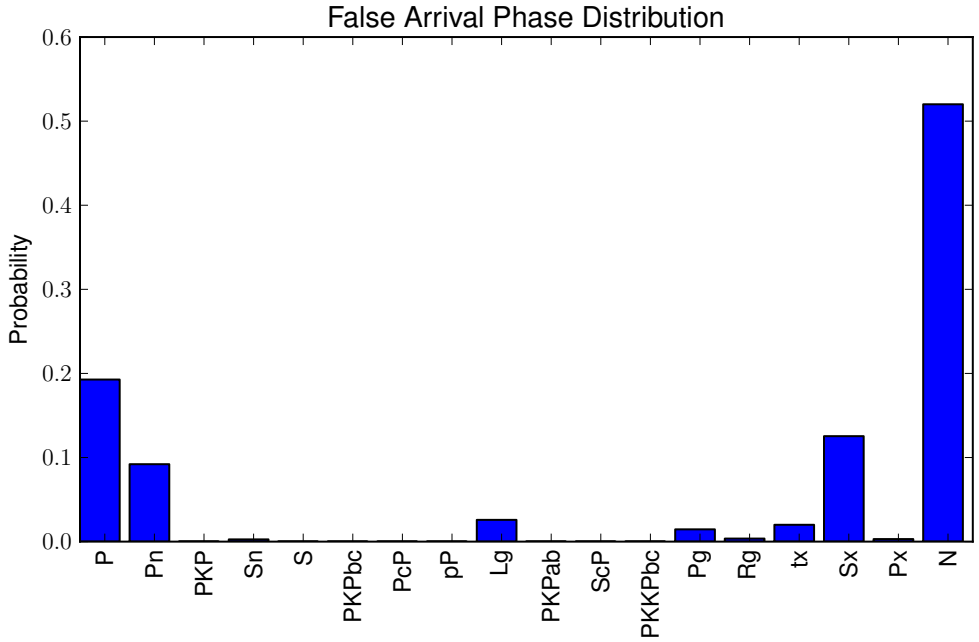

Figure 4.11: Phase distribution for all false arrivals.

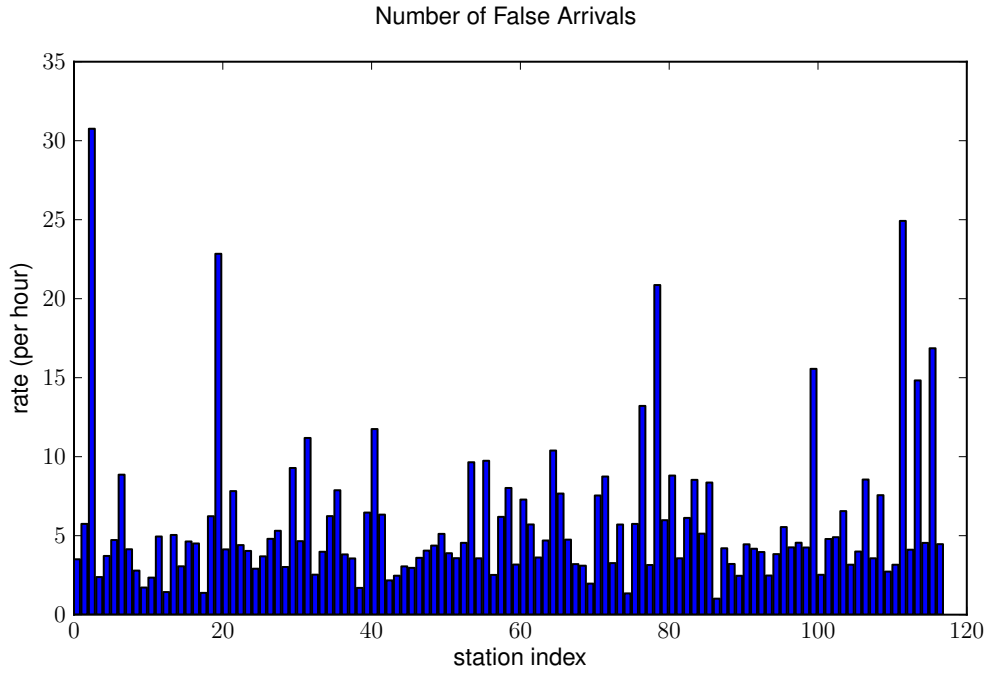

Figure 4.12: The average false arrival rate per hour at all the stations.

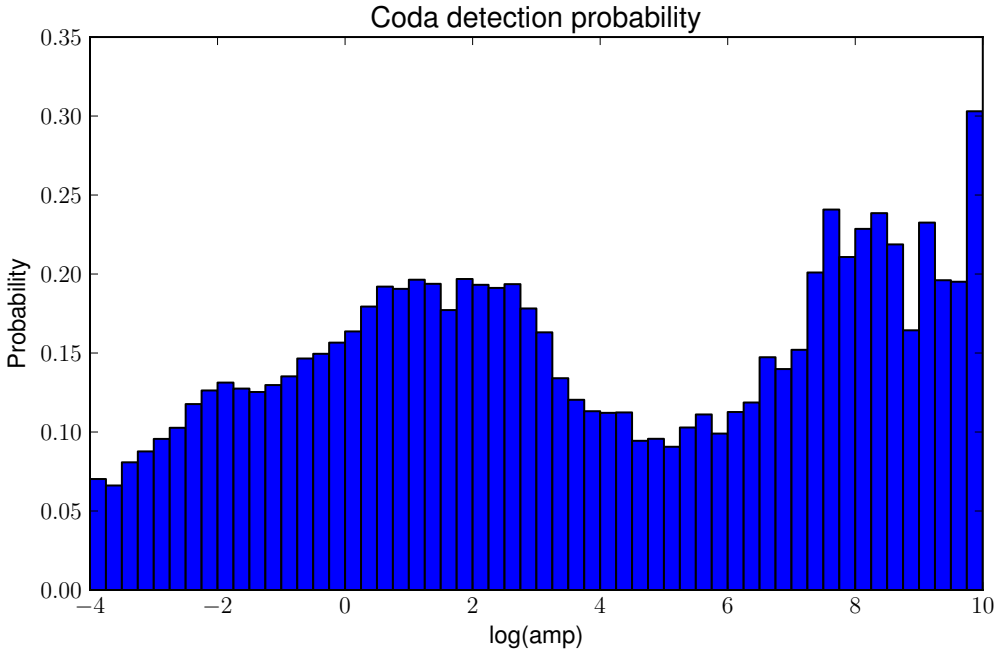

Figure 4.13: Coda detection probability as a function of the triggering arrival's log amplitude.

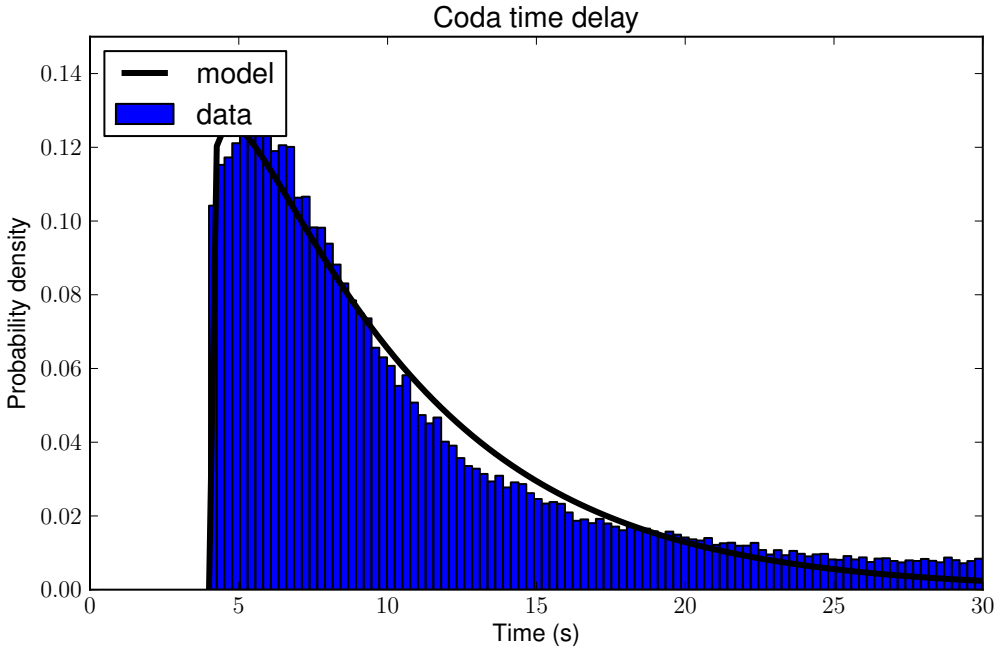

Figure 4.14: Time delay for coda arrival after the triggering arrival.

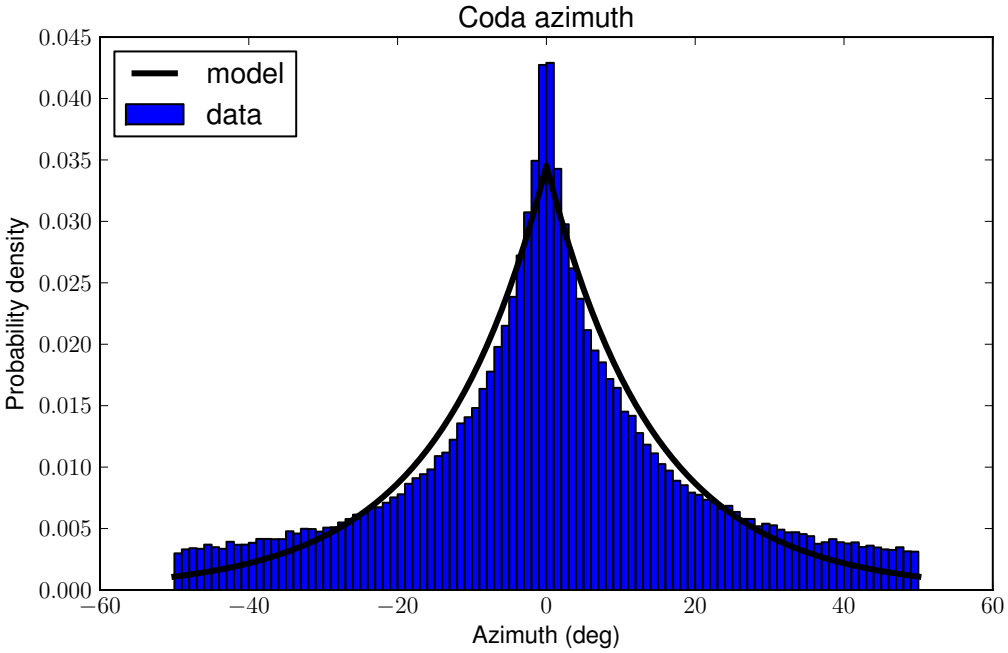

Figure 4.15: Coda azimuth difference versus triggering arrival.

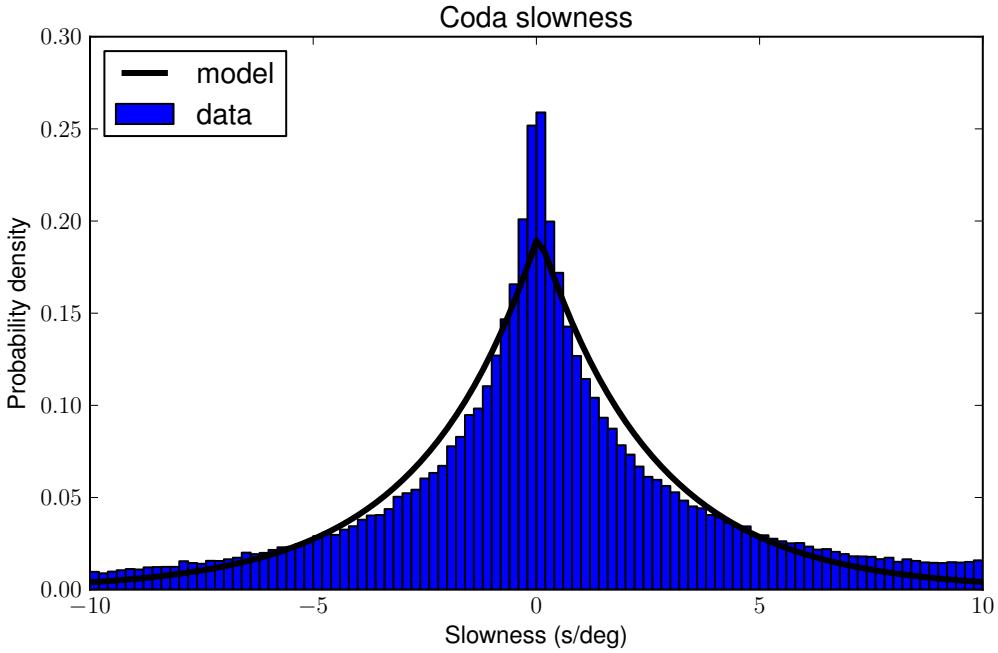

Figure 4.16: Coda slowness difference versus triggering arrival.

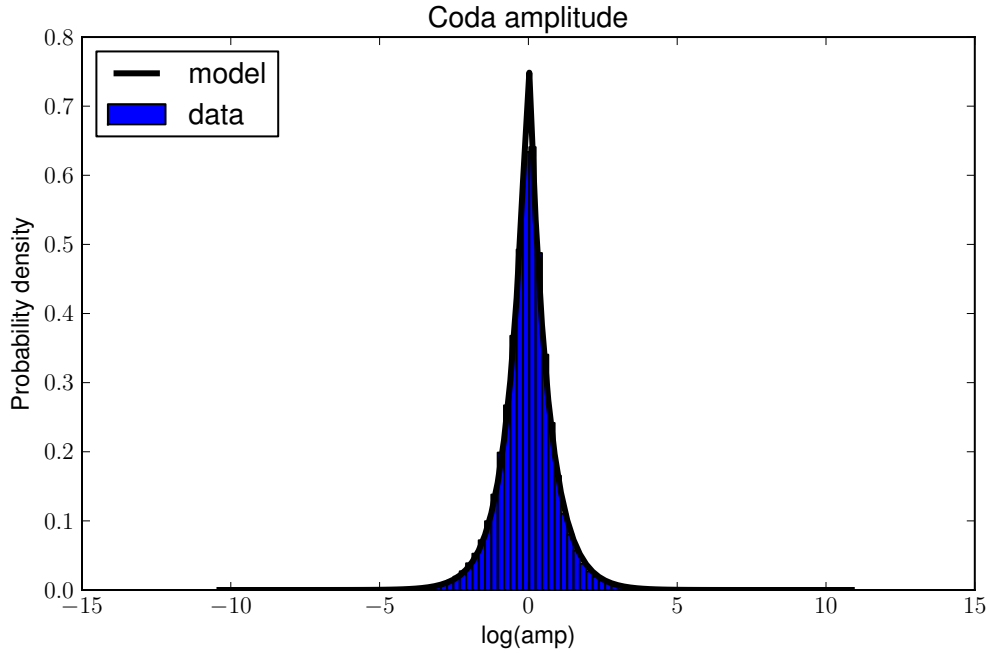

Figure 4.17: Coda log amplitude difference versus triggering arrival.

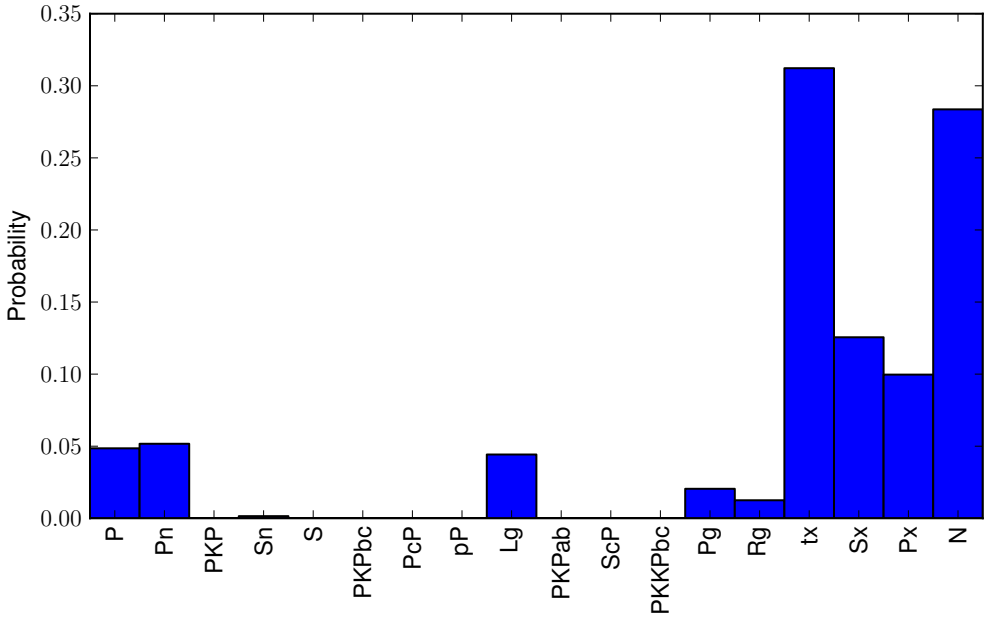

Figure 4.18: Coda phase.

Chapter 5

Inference

Combining the model components developed in the preceding section, the overall probability of any hypothesized sequence of events e, true arrivals Λ, false arrivals ξ, and coda arrivals κ for a given set of observed arrivals A, where A^k is the set of arrivals at station k is

$$P(e, \Lambda, \xi, \kappa, A) = P_\theta(e) P_\phi(\Lambda \mid e) P_\varphi(\xi) P_\gamma(\kappa \mid A) \mathbb{1}(A = \Lambda \oplus \xi \oplus \kappa). \tag{5.1}$$

Where the last term is required to ensure that any arrival is in exactly one of the three sets Λ, ξ, and κ. Other obvious consistency requirements such as requiring $\xi^k \subseteq A^k$ are left out for brevity. We will attempt to find the most likely explanation consistent with the observations,

$$\arg\max_{e,\Lambda,\xi,\kappa} P(e, \Lambda, \xi, \kappa, A).$$

5.1 Overall Algorithm

Since arrivals from real seismic sensors are observed incrementally and roughly in time-ascending order, our inference algorithm also produces an incremental hypothesis which advances with time. Our algorithm can be seen as a form of greedy search, in which the current hypothesis is improved via a set of local moves.

Let M_T denote the maximum travel time for any phase. Initially, we start with an event-window of size W from $t_0 = 0$ to $t_1 = W$, and an arrival-window of size $W + M_T$ from $t_0 = 0$ to $t_2 = W + M_T$. Then we perform a series of greedy moves which add or update events in the event-window, delete existing events, or classify (as true arrival, false arrival, or coda arrival) the arrivals in the arrival window. Next, the windows are moved forward by a step size S. At this point, events older than $t_0 - M_T$ become stable: none of the moves modify either the event or arrivals associated with them. These events are then output. While in theory this algorithm never needs to terminate, our experiments continue until the test dataset is fully consumed.

Our initial hypothesis is that all new arrivals added to the arrival window are false arrivals. We refine this by classifying any arrival a (at station k) as a coda arrival if

$$P_{\gamma,d}(\eta^a)P_\gamma(a) > (1 - P_{\gamma,d}(\eta^a))p^k_\varphi(a).$$

This default classification for an arrival is retained whenever it is no longer associated with an event. For convenience we define

$$\Upsilon^k(a) = \max(P_{\gamma,d}(\eta^a)P_\gamma(a), (1 - P_{\gamma,d}(\eta^a))p^k_\varphi(a)).. \tag{5.2}$$

Next, the birth move generates new events in the event-window. These events are added to the hypothesis with $\Lambda^{ijk} = \zeta$ for each new event i. Subsequently, we repeat the following N times: improve-arrival move for each arrival in the arrival-window, and improve-event move for each event in the event window. Finally, the death move kills some of the events and we repeat one round of improve-arrivals and improve-events move. We describe these steps algorithmically below. The individual moves will be described in the next section.

1. $t_0 = 0; t_1 = W; t_2 = W + M_T$
2. repeat while $t_0 <$ max time
   (a) give a default classification to arrivals in t_0 to t_2
   (b) add events from birth-move(t_0, t_1, $\{a : t_0 \leq a_t \leq t_2\}$)
   (c) repeat N times
      i. for each arrival a, s.t. $t_0 \leq a_t \leq t_2$, improve-arrival(a)
      ii. for each event e^i, s.t. $t_0 \leq e^i_t \leq t_1$, improve-event(e^i)
   (d) for all events e^i, death-event(e^i)
   (e) for each arrival a, s.t. $t_0 \leq a_t \leq t_2$, improve-arrival(a)
   (f) for each event e^i, s.t. $t_0 \leq e^i_t \leq t_1$, improve-event(e^i)
   (g) $t_0 += S; t_1 += S; t_2 += S$
   (h) output e^i, Λ^{ijk} for all e^i s.t. $e^i_t < t_0 - M_T$
3. output any remaining e^i

In order to simplify the computations needed to compare alternate hypotheses, we decompose the overall probability of Eq. 5.1 into the contribution from each event. We define the score S_e of an event as the probability ratio of two hypotheses: one in which the event exists, and another in which the event doesn't exist and all of its associated arrivals have the default classification (false or coda). If an event has score less than 1, an alternative hypothesis in which the event is deleted clearly has higher probability. Critically, this event score is

unaffected by other events in the current hypothesis. From Eqs. 4.7, 4.8, 4.9, 4.10, and 5.2 we have

$$S_e(e^i) = p_\theta(e^i) \prod_{j=1}^{J} \prod_{k=1}^{K} \left[\mathbb{1}(\Lambda^{ijk} = \zeta)(1 - P_{\phi,d}^{jk}(e^i)) + \frac{\mathbb{1}(\Lambda^{ijk} \neq \zeta) P_{\phi,d}^{jk}(e^i) P_{\phi}^{jk}(\Lambda^{ijk} \mid e^i)}{\Upsilon^k(\Lambda^{ijk})} \right].$$

Note that the final fraction above is a likelihood ratio comparing interpretations of the same arrival as either the arrival of event i's j^{th} phase at station k, or the as a false arrival or a coda arrival. We can further decompose the score into scores S_d for each arrival. The score of Λ^{ijk}, defined when $\Lambda^{ijk} \neq \zeta$, is the ratio of the probabilities of the hypothesis where the arrival is associated with phase j of event i at station k versus the default classification.

$$S_d^{jk}(\Lambda^{ijk} \mid e^i) = \frac{P_{\phi,d}^{jk}(e^i)}{1 - P_{\phi,d}^{jk}(e^i)} \frac{P_{\phi}^{jk}(\Lambda^{ijk} \mid e^i)}{\Upsilon^k(\Lambda^{ijk})} \ .$$

By definition, any arrival with score less than 1 is more likely to be a false or coda arrival. Also, the score of an individual arrival is independent of other arrivals and events in the hypothesis. These scores play a key role in the following local search moves.

5.2 Moves

Among the moves, the birth move is the only one which is not a greedy move: the proposed event will almost always have a score $S_e(e^i) < 1$ until some number of arrivals are assigned by subsequent moves. The overall structure of these moves could be easily converted to an MCMC or simulated annealing algorithm. However, in our experiments this search outperformed simple MCMC methods in terms of speed and accuracy.

5.2.1 Birth Move

The birth move proposes events within a given time range, based on a list of arrivals. It starts off by inverting each of these arrivals to obtain an initial candidate list of event locations and times. The ability to invert an event follows from the fact that the slowness of an arrival is a monotonic function of distance (with fixed depth). If one assumes that an arrival is the P phase of a surface event one can obtain a distance estimate from the slowness which combined with the arrival azimuth and time gives an estimate for an event location and time. In Figure 5.1 we show the statistics of the distance between the inverted locations obtained from all arrivals in a 1 week period and the corresponding ground truth events during the same time period.

Next, we attempt to construct the best possible event within a 5 degree and 50 second ball around each of the candidate location in steps of 2.5 degrees, magnitudes $3, 4$, surface depth and using all available detections. The best such event is further optimized using the

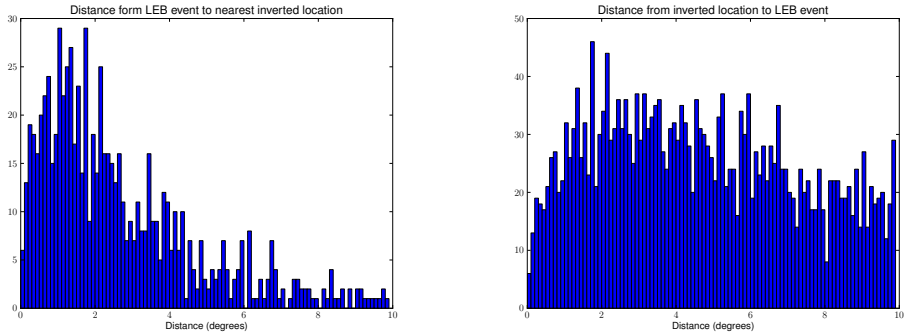

Figure 5.1: Distance between events and inverted locations within 10 degrees and 100 seconds. The left figure shows the distance from a true event and the nearest inverted location. The right figure shows the converse, i.e. the distance from an inverted location to the nearest true event.

improve-arrival and improve-event moves and then set aside. This process is repeated as long as the best event has a score greater than 1. An event is not allowed to use arrivals associated to events found earlier in this process. Finally, all these set aside events sans the arrivals associated to them are returned. The more precise details of this are given algorithmically below.

1. given t_0, t_1, and arrivals A
2. repeat for each a in A
   (a) invert a to obtain a candidate location ι_a
   (b) repeat for each location e in a ball around ι_a
      i. initialize Λ^e
      ii. repeat for each arrival a in A (let k be a's station)
         A. determine the phase j with the max score $S_d^{jk}(a \mid e)$
         B. if $S_d^{jk}(a \mid e) > S_d^{jk}(\Lambda^{jk} \mid e)$ or if $\Lambda^{jk} = \zeta$ and $S_d^{jk}(a \mid e) > 1$ then set $\Lambda^{jk} = a$.
3. let e be the event with the max score $S_e(e)$ in step 2
4. repeat 100 times
   (a) invoke improve-event(e)
   (b) invoke improve-arrival(a) for all a in A with e as the only potential event
5. if $S_e(e) > 1$ then set aside event e and remove arrivals in Λ^{ejk} from A, go to step 2
6. return set aside events

5.2.2 Improve Arrival Move

For each arrival in the arrival window, we consider all possible phases j of all events i up to M_T seconds earlier. We then associate the best event-phase for this arrival that is not already assigned to an arrival with higher score at the same station k. If this best event-phase has score $S_d^{jk}(\Lambda^{ijk} \mid e^i) < 1$, the arrival is changed to its default status (one of false or coda). In more precise terms.

1. given arrival a at station k
2. repeat for each event e
   (a) determine the phase j with the max score $S_d^{jk}(\Lambda^{ejk} \mid e)$
3. let e be the event with the max score $S_d^{jk}(\Lambda^{ejk} \mid e)$
4. if $S_d^{jk}(a \mid e) > S_d^{jk}(\Lambda^{ejk} \mid e)$ or if $\Lambda^{ejk} = \zeta$ and $S_d^{jk}(a \mid e) > 1$ then set $\Lambda^{ejk} = a$.

5.2.3 Improve Events Move

For each event e^i, we consider 100 points chosen uniformly at random in a small ball around the event (2 degrees in longitude and latitude, 100 km in depth, 5 seconds in time, and 2 units of magnitude), and choose those attributes with the highest score $S_e(e^i)$.

5.2.4 Death Move

Any event e^i with score $S_e(e^i) < 1$ is deleted, and all of its currently associated arrivals are marked as false alarms.

5.3 Parallel Birth Moves

The algorithm described so far works on a single core of one machine. This may be sufficient for most days when the number of arrivals is low. However, after a large earthquake, for example the 2004 Sumatra earthquake or the 2011 Tohoku earthquake (off the coast of Japan), the number of seismic arrivals can jump ten-fold above normal. Under these circumstances it is necessary to take advantage of parallel processing to keep up with the deluge of data. The main object of our parallelization is the birth move in Section 5.2.1 which is asymptotically the most expensive. This move is roughly quadratic in the number of arrivals since we propose an event from each arrival and we attempt to associate each of these events with all the other arrivals.

In order to parallelize the birth move we employ two different strategies. First we exploit multiple cores by implementing a multithreaded birth move. Step 2 in the birth move (see

Section 5.2.1) inverts each arrival into an event and attempts to find the best of these events. We divide the task of inverting the arrivals evenly among the threads. Each thread finds its best event in parallel and then in serial we find the best event from all the threads.

The second approach is to divide the time windows among multiple machines and have each machine compute the birth proposal events for its window in parallel, possibly using the multithreaded proposer. The birth proposers for all the windows are computed up front and stored before the rest of the algorithm is started. When the overall algorithm described in Section 5.1 invokes the birth proposer in Step 2b, the previously computed events are returned. There is a slightly subtlety in the manner in which the time windows for the birth move are divided among the available machines. Since the time taken by the birth move is quadratic in the number of arrivals, the time windows are divided in such a manner that the square of the number of arrivals in each window are roughly equal.

5.4 Tempering

The model and inference that we have described so far places no constraints on which seismic phases from an event may be detected at a station. As long as the detection probability $P^{jk}_{\phi,d}(e^i)$ is non-zero, we could allow an arrival at station k to be associated as the phase j for event e^i. In reality, the phase labels capture properties of the seismic waves which only make sense at certain event-station distances, or when the event is at a certain depth, or both. For example, a P phase is not detected at a station which is 10 degrees away from an event unless the event is more than 40 km deep. If the event was less than 40 km deep a Pn phase may be possible though. A complete list of these constraints is presented in table 5.1.

In general, the phase labels in the training data follow these constraints and the learned model assigns low probability to a phase association which violates them. The training data though is not perfect and even if there is a small probability assigned by the model, it is quite possible that such violations can make it into the resulting bulletin. To ensure that all these phase constraints are respected we could always force the detection probability to be zero whenever a constraint is violated. However, this would greatly limit the ability of the inference to search the space of hypotheses. We have chosen to gradually decrease the detection probability in each iteration of the inference for phases that violate constraints. This allows the inference flexibility in associating any phase early on, but then over successive iterations all the violations are removed. Let ϵ^{ijk} be the maximum of the amount that the distance in degrees between event i and station k exceeds the bounds for phase j or the amount that the depth of event i exceeds the bounds for phase j. We replace the original detection probability in the model by the tempered detection probability, which is defined as,

$$P^{jk}_{\phi,d'}(e^i) = P^{jk}_{\phi,d}(e^i) \cdot \exp(-\frac{\epsilon^{ijk}}{T_n})$$

Where $T_n > 0$ is the temperature in the nth iteration (1 to N) of Step 2c in the overall

algorithm (Section 5.1). If T_n is very small and approaching zero then the detection probabilities for impossible phases will approach zero and the inference will automatically avoid these phase labels. On the other hand, for larger T_n the constraints will be almost ignored. In our inference we start with $T_1 = 100$ and multiply the temperature by .6 in each iteration. After the typical $N = 20$ iterations, the temperature goes down to 0.0036.

Phase	min-distance (deg)	max-distance (deg)	min-depth (km)	max-depth (km)
P	0	98	40	800
P	17	98	0	40
Pn	2	17	0	40
PKP	114	180	0	800
Sn	2	12	0	40
S	0	30	40	800
PKPbc	149	155	0	800
PcP	10	70	0	800
pP	10	98	10	800
Lg	0	12	0	40
PKPab	145	176	0	800
ScP	10	62	0	800
PKKPbc	105	125	0	800
Pg	0	8	0	40
Rg	0	3	0	0

Table 5.1: Event to station distance and event depth ranges for seismic phases.

Chapter 6

Experimental Results

6.1 Evaluation of seismic bulletins

The accuracy of a seismic bulletin depends primarily on the accuracy of the event locations in the bulletin and the coverage of the events reported. At the same time, the bulletin must not contain too many false alarms which could render it useless. Another important factor is the quality of the arrivals associated with the events. However, the quality can be quite subjective and we will not consider this much further other than ensuring that each arrival associated to an event is within the distance and depth guidelines for the corresponding phase.

The central question in evaluation is determining whether a reported event in a predicted bulletin matches a true event in the ground truth bulletin. We have chosen the operational criteria that two events which are within 5 degrees of epicentral distance and within 50 seconds in time are a potential match. In practice, multiple reported events could match a true event and vice versa. Our approach is to find a one-to-one matching between reported and true events and mark all other events as either spurious reported events, or missed true events. Clearly, if given a choice, we would match the events which are closer to each other in terms of epicentral distances. At the same time our prime objective is to maximize the size of the matching. In other words, we compute a max-cardinality min-weight matching on the bipartite graph over events in the two bulletins with edges between reported and true events which are a potential match. The weight of an edge is the epicentral distance (in km) between the two events. Given a matching between two bulletins we will estimate the precision, the percentage of reported events which match, the recall, the percentage of true events which match, and the average error, the average distance between matched events.

6.2 Comparison of NET-VISA and SEL3

We first treat the IDC analyst-generated LEB as ground truth, and compare the performance of our NET-VISA algorithm to the currently deployed SEL3 system. Using the scores for hypothesized events, we have generated a precision-recall curve for NET-VISA, and marked SEL3 on it as a point (see Figure 6.1). Also in this figure, we show a precision-recall curve for SEL3 using scores from an SVM trained to classify true and false SEL3 events (Mackey et al., 2009), SEL3 extrapolation. As shown in the figure, NET-VISA has at least 18% more recall at the same precision as SEL3, and at least 33% more precision at the same recall as SEL3.

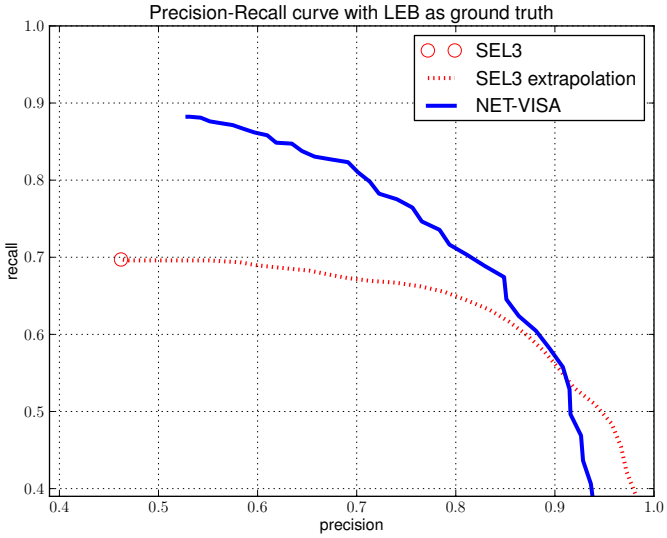

Figure 6.1: Precision-recall performance of the proposed NET-VISA and deployed SEL3 algorithms, treating the analyst-generated LEB as ground truth.

To further understand why NET-VISA is able to find events missed by SEL3 we subdivide the NET-VISA and SEL3 recall and average error by two different criteria. In Table 6.1 we subdivide by LEB event magnitude. For magnitudes up to 4, NET-VISA has nearly 20% higher recall with similar error. In Table 6.2 we subdivide by LEB event azimuth gap. The azimuth gap of an event is the largest difference between successive event-to-station azimuths for stations where the automated processing detected an arrival for the event. Large gaps indicate that the event location is under-constrained. For example, if all stations are to the southwest of an event, the gap is greater than 270 degrees and the event will be poorly localized along a line running from southwest to northeast. The results in these two tables highlight a common theme: NET-VISA performs significantly better than SEL3 whenever there is less data available. Under these circumstances the additional information in the

m_b	Count	SEL3		NET-VISA	
		Recall	Err	Recall	Err
unknown	74	64.9	101	90.5	87
2 – 3	36	50.0	186	88.9	137
3 – 4	558	66.5	104	86.6	100
> 4	164	86.6	70	92.1	65
all	832	69.7	99	88.2	94

Table 6.1: Recall and average error (km) subdivided by LEB event magnitude (m_b).

Azimuth Gap	Count	SEL3		NET-VISA	
		Recall	Err	Recall	Err
0 – 90	55	100.0	22	100.0	32
90 – 180	260	93.5	66	98.1	55
180 – 270	273	59.7	120	88.3	104
270 – 360	244	48.8	173	75.0	152
all	832	69.7	99	88.2	94

Table 6.2: Recall and average error (km) subdivided by LEB event azimuth gap

NET-VISA model – location prior, amplitude, misdetections etc. – play a critical role in determining a better location for the events.

Bulletin	lon	lat	depth (km)	m_b	time	$S_e(\cdot)$	$P_{\theta,l}(\cdot)$
LEB	177.67	-33.29	0.0	3.5	1237691617.7	13.3	-10.9
NET-VISA	-179.61	-33.50	0.7	3.3	1237691599.7	16.0	-9.9
SEL3	170.48	-32.42	0.0	3.8	1237691669.5	-1.8	-17.2

Table 6.3: The various bulletin events around LEB origin-id 5287957.

In Table 6.3 we show the location of one such under-constrained LEB event, origin-id 5287957, the nearest NET-VISA event, which is 2 degrees and 18 seconds away, and the nearest SEL3 event, which is 8 degrees and 52 seconds away. Although there are very few arrivals to help locate this event and the azimuth gap is more than 270 degrees it is worth noting that NET-VISA is able to use the event location prior, shown in the last column of Table 6.3, to choose a more likely location for the event. Further, NET-VISA associates this event to an Sn phase at station URZ (Urewera, New Zealand) without a corresponding Pn phase at the same station, something SEL3 would avoid! The associations of the events to

Bulletin	station:phase	station:phase	station:phase
LEB	URZ:Sn	ASAR:P	WRA:P
$S_d(\cdot)$	7.8	13.1	15.0
NET-VISA	URZ:Sn	ASAR:P	WRA:P
$S_d(\cdot)$	8.8	12.6	14.8
SEL3	PLCA:P	ASAR:P	WRA:P
$S_d(\cdot)$	1.2	13.0	15.1

Table 6.4: The station-phase associations of various bulletin events around LEB origin-id 5287957.

the automatically detected arrivals are displayed in Table 6.4. Not shown in this table is a Pn arrival at station URZ that the analyst manually added to the LEB event. If it were not for this Pn arrival, the analyst would probably not have associated the Sn arrival either. Of course, the additional arrival was not available to NET-VISA or SEL3 and hence their location estimates are expected to be different from the LEB location. This particular event also demonstrates the perils of choosing the MAP event location. In Figure 6.2 we plot the contours of $max(S_e(e_l = \cdot, e_{m,t} = e^{LEB}_{m,t}), S_e(e_l = \cdot, e_{m,t} = e^{NET-VISA}_{m,t})$, or in other words the maximum score of a hypothetical event which could use the NET-VISA or the LEB event time, magnitude, and associated arrivals. As this figure suggests, the posterior event location density is multimodal. The mode picked by NET-VISA happens to be 2 degrees away from the mode near the LEB location.

6.3 Comparison of NET-VISA and LEB

Returning to Figure 6.1, the gap between SEL3 extrapolation and NET-VISA on the lower right end of the figure suggests that NET-VISA is predicting spurious events with extremely high confidence. In reality many of these events are real events which are missed in the LEB bulletin by the human analysts. To understand the true scope of the LEB bulletin we have compared it against various regional bulletins, which are based on data from many more stations than in the IMS. In Table 6.5 we restrict both the LEB and NET-VISA bulletins to different regions of the earth and evaluate against appropriate regional bulletins. This table shows that NET-VISA outperforms LEB on all of these regions, and most notable in the U.S. and Kazakhstan, where NET-VISA finds at least an additional 10% of the events reported in the regional bulletin. The results in the continental U.S. are further subdivided by the local magnitude (ML) in Table 6.6. These results show clearly that NET-VISA is able to find half of the ML 3 to 4.5 and some even weaker events while LEB finds only 7% of events with ML 3 to 4.5 and none below this.

The exact demarcations of these regions and the bulletins used are described in Table 6.7.

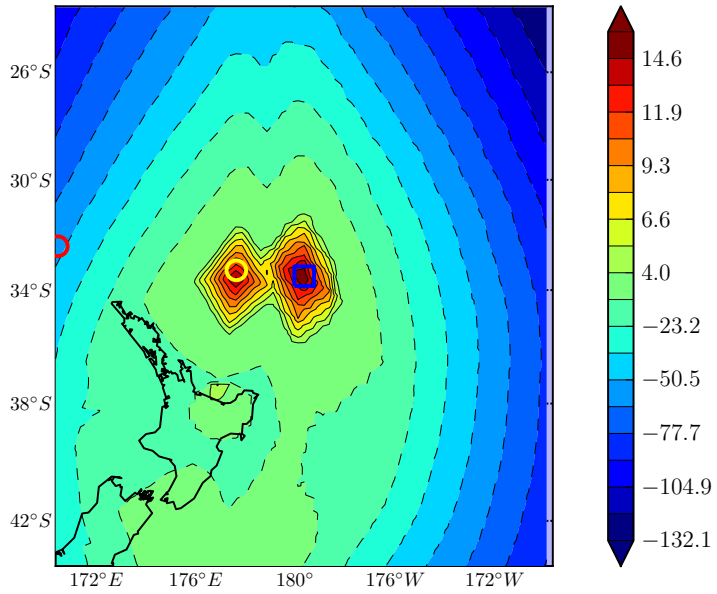

Figure 6.2: SEL3 (red circle), LEB (yellow circle), NET-VISA (blue square) location and posterior log odds ratio around LEB origin-id 5287957.

Incorporated Research Institutions for Seismology's (IRIS) primary seismic origins were used for the continental United States. For the other regions we relied upon the raw bulletins from the International Seismological Centre (ISC)(International Seismological Centre, 2010). The relevant agency codes are displayed in the table.

6.4 Ablation Experiments

In order to demonstrate the effectiveness of each component of our model and inference, we have separately ablated each component and measured the resulting performance. Figure 6.3 shows the precision-recall curve without the slowness. In other words, the slowness of the arrivals is not observed. The slowness value is still used to compute the proposed locations in the birth move. Similarly, Figure 6.4 has the curve for the model sans azimuth. These two figures demonstrate that slowness and azimuth are crucial for determining the event location. We haven't included the curve after ablating arrival time because this was extremely low – precision 10.8 % and recall 20.7%.

The arrival phase reported by the IDC software contains important information about the polarization of the seismic waveform which helps distinguish the P and S waves which

Region	Count	LEB			NET-VISA		
		Precision	Recall	Err	Precision	Recall	Err
Japan	1565	100.0	1.9	38	83.7	2.3	122
U.S.	132	100.0	3.0	33	78.3	13.6	106
Italy	96	50.0	1.0	49	66.7	2.1	43
Kazakhstan	65	73.3	16.9	57	70.4	30.8	67

Table 6.5: Precision, recall, and average error (km) of LEB and NET-VISA measured against various regional bulletins.

ML	Count	LEB		NET-VISA	
		Recall	Err	Recall	Err
unknown	13	23.1	35	23.1	62
1.0 – 2.0	23	0.0	-	0	-
2.0 – 2.5	48	0.0	-	4.2	176
2.5 – 3.0	35	0.0	-	17.1	131
3.0 – 4.5	13	7.7	28	53.8	84
all	132	3.0	33	13.6	106

Table 6.6: Recall and average error (km) subdivided by event magnitude (ML) in the continental U.S.

have orthogonal particle motion and also the Lg phase (see Figure 4.9 for the differences in the phase to phase generation probabilities). With the phase information ablated from the model, there is much more freedom to assign arrivals to events resulting in many spurious events. This shows up as reduced precision as we increase the event score threshold. The phase ablation results are presented in Figure 6.5. On the other hand, ablating amplitude, Figure 6.6, doesn't reduce performance as much since there is considerable uncertainty in the amplitude model (Figure 4.8). This higher uncertainty results in the amplitude providing very little information. .

Ablating the coda model, on the other hand, actually results in a slight increase of the recall from 88.2% to 89.9%. However, this comes at the expense of a large number of spurious events – the precision drops from 52.9% to 33.7%. The precision recall curve is plotted in Figure 6.7. The spurious events mainly occur in clusters around large magnitude events which tend to generate a number of coda detections.

The final ablation results are on the tempering of distance-depth constraints. In Figure 6.8 we show three different results. The baseline curve, as usual, is the full NET-VISA model and inference. The "hot" curve is the NET-VISA model running at the max temperature $T = 100$ throughout. The "cold" curve is the NET-VISA model running at the

Region	Longitudes	Latitudes	Ground truth
Japan	130 to 145	30 to 45	JMA bulletin (ISC)
U.S.	-125 to -70	25 to 50	IRIS
Italy	6 to 19	36 to 48	ROM bulletin (ISC)
Kazakhstan	46 to 86	40 to 55	NNC bulletin (ISC)

Table 6.7: The definition of the various regions used for the regional evaluation and the corresponding ground truth bulletin.

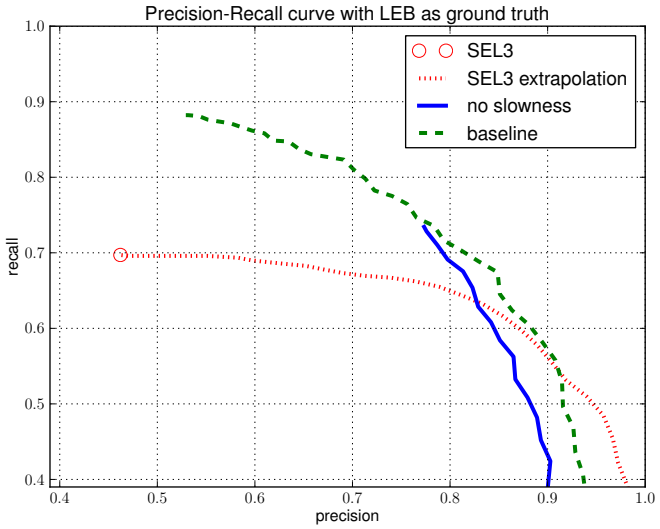

Figure 6.3: Performance after ablating arrival slowness.

minimum temperature $T = .003$. At the high temperature there is greater flexibility in assigning phases to events and this leads to slightly better performance, and conversely the run at the cold temperature has slightly worse performance than the baseline. However, in the high temperature run only 35% of the inferred events respect the constraints, whereas 99% of the events from the other two runs respect all the constraints.

6.5 Data and Resources

All the results in this thesis are based on 3 months of IMS data (March 22 to June 20, 2009), which was made available by the CTBTO through vDEC, Virtual Data Exploitation Center (Vaidya et al., 2009). The 3 months of data was divided into 7 days of validation data (March 22 to March 29, 2009) on which all of the results were produced and 2.5 months

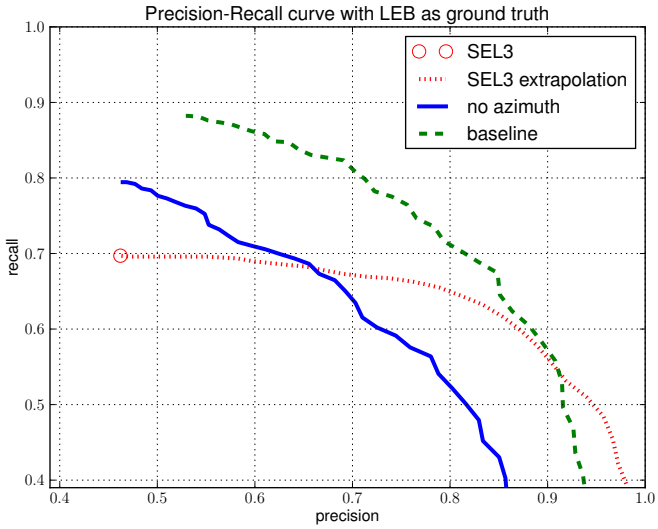

Figure 6.4: Performance after ablating arrival azimuth.

of training data (April 5 to June 20, 2009).

The inference used a window size, W, of 30 minutes, a step size, S, of 15 minutes, and $N = 20$ iterations. There were a total of 832 LEB events in the validation data and roughly $120,000$ arrivals. Inference on a single 8-core machine running at 2.5 GHz took about 21 hours. On a cluster of 20 machines, each with 8 cores, the birth move completed in 1.2 hours and consumed a total of 19.5 node-hours on the cluster. The rest of the inference finished up in 3.5 hours in serial. Estimating model parameters on the training data took about 2 hours. All the performance results in this paper were obtained on the Amazon EC2 cluster which was made possible by an AWS in Education grant award.

The regional events were obtained from two sources. The IRIS primary origins were downloaded from SeismiQuery (`http://www.iris.edu/SeismiQuery/sq-events.htm`) and the ISC raw bulletins from `http://www.isc.ac.uk/iscbulletin/search/catalogue/`. The IRIS primary origins in turn are derived from various sources – USGS ENS alert, USGS QED, USGS Weekly PDE, USGS Monthly PDE, GCMT, and ISC.

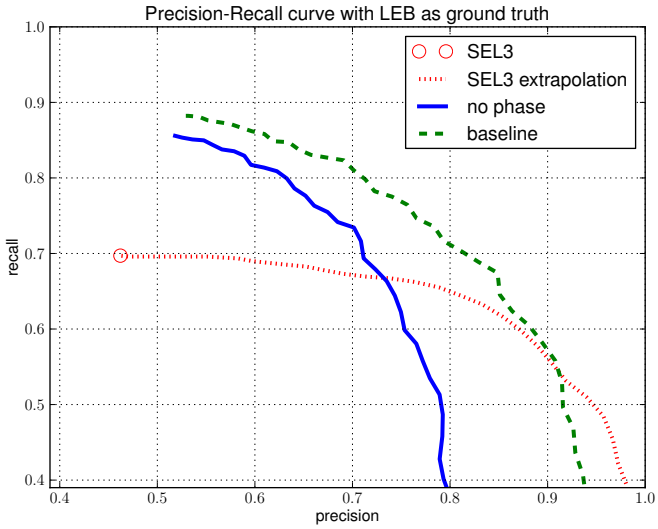

Figure 6.5: Performance after ablating arrival phase.

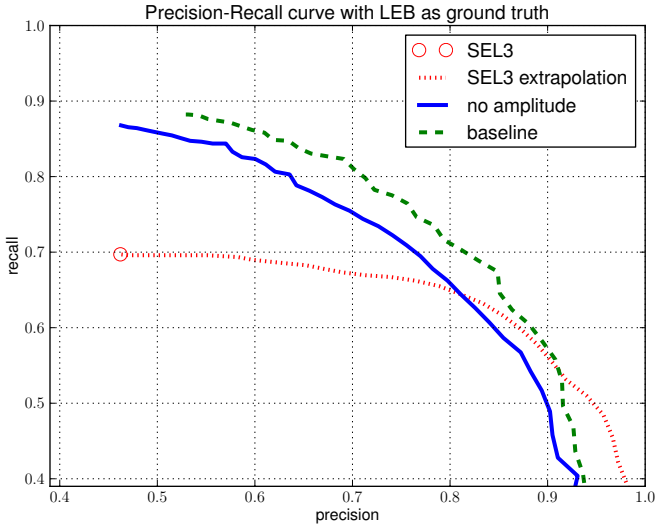

Figure 6.6: Performance after ablating arrival amplitude.

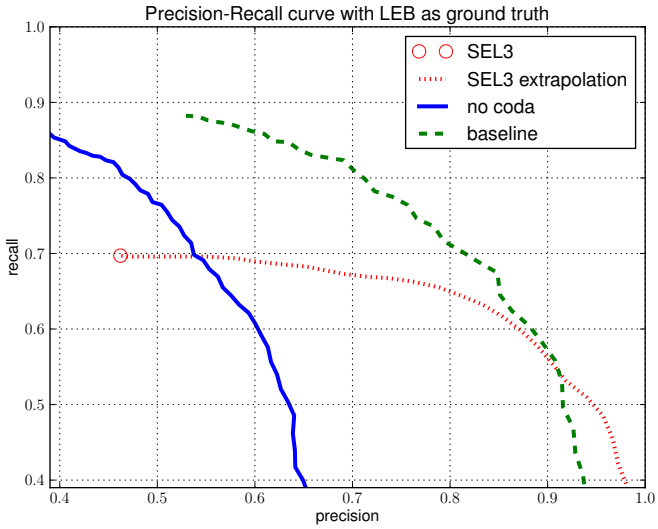

Figure 6.7: Performance after ablating coda model.

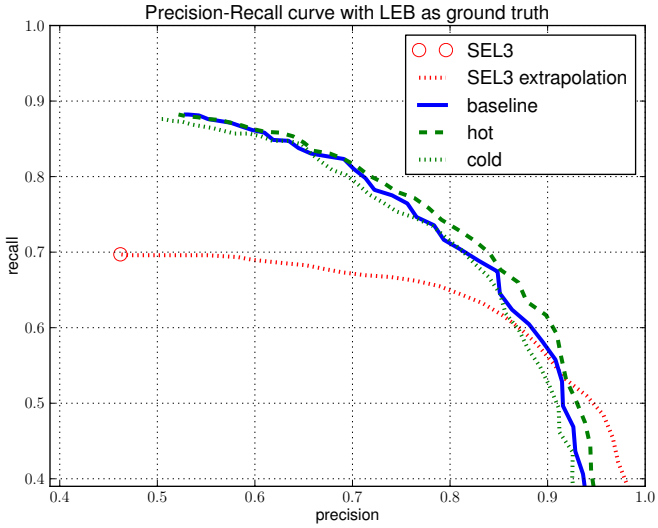

Figure 6.8: Performance after ablating tempering model.

Chapter 7

Conclusions and Further Work

Standard seismological processing proceeds in multiple steps. Phase identification, arrival association, event location, event quality determination, and magnitude estimation. In sharp contrast, we have proposed a probabilistic model which incorporates all of these separate processes in a single vertically integrated model. There is no need for an iterative linearized least squares algorithm for event location. Our simple, parallelizable inference algorithm takes the complete model into account when proposing and locating events. Although probabilistic models for event location have been proposed before, ours is the first one which combines association and location. In doing so, we have incorporated detection probabilities and various other aspects such as historical seismicity, arrival amplitude and phase into a holistic inference.

A vertically integrated model for seismology may be a new idea, but the individual components of this model are based on standard seismological practice. For instance, we use the IASPEI travel time tables and corrections for the earth's ellipticity as well as SASC corrections for azimuth and slowness. All the seismic phases reported in the NET-VISA bulletin comply with the accepted distance and depth ranges for those phases. The approach of using tempering to gradually incorporate deterministic constraints allows us to take advantage of seismic knowledge without being hampered by it.

Our results demonstrate a significant improvement on classical systems. The NET-VISA system can not only reduce the human analyst effort required to achieve a given level of accuracy, but can also lower the magnitude threshold for reliable detection. Given that the difficulty of seismic monitoring was cited as one of the principal reasons for non-ratification of the CTBT by the United States Senate in 1999, one hopes that improvements in monitoring may increase the chances of final ratification and entry into force.

In future, we see various extensions to our work. The parameters of our model can be estimated continuously without the need for a separate training and test set. The generative model can be extended to include waveforms directly rather than the arrival parameters, so that there is no need for explicitly picking arrivals. The general approach of combining association and location can be extended to other similar problems such as locating microseismic

events in oil and gas drilling.

Bibliography

C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50:5–43, 2003.

Nimar Arora, Stuart Russell, Paul Kidwell, and Erik Sudderth. Global seismic monitoring as probabilistic inference. In *Advances in Neural Information Processing Systems (NIPS) 23*, pages 73–81. MIT Press, 2010a. URL `http://books.nips.cc/papers/files/nips23/NIPS2010_0891.pdf`.

Nimar Arora, Stuart Russell, Paul Kidwell, and Erik Sudderth. Global seismic monitoring: A bayesian approach. In *AAAI*, 2011.

Nimar S Arora, Rodrigo de Salvo Braz, Erik Sudderth, and Stuart J. Russell. Gibbs sampling in open-universe stochastic languages. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, 2010b. URL `http://event.cwi.nl/uai2010/papers/UAI2010_0202.pdf`.

Nimar S. Arora, Stuart J. Russell, and Erik Sudderth. Automatic inference in BLOG. In *Statistical Relational AI (StarAI) Workshop. American Association of Artificial Intelligence (AAAI)*, 2010c.

S. Ballard and P. Reeves. Improved seismic event location resolution using a damped least squares algorithm. In *24th Annual Seismic Research Review Symposium*, 2002.

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

S. S. Blackman. *Multiple-Target Tracking with Radar Applications*. Artech House, Dedham, MA, 1986.

B. A. Bolt. The revision of earthquake epicentres, focal depths and origin times using a high-speed computer. *Geophysics Journal Royal Astronomical Society*, 3:433 – 440, 1960.

J.W. Dewey. Seismicity and tectonics of western venezuela. *Bulletin of the Seismological Society of America*, 62:1711–1751, 1972.

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

E. R. Engdahl and R. H. Gunst. Use of a high speed computer for the preliminary determination of earthquake hypocenters. *Bulletin of the Seismological Society of America*, 56: 325 – 336, 1966.

E. A. Flinn. Local earthquake location with an electronic computer. *Bulletin of the Seismological Society of America*, 59:1365 – 1398, 1960.

E. A. Flinn. Confidence regions and error determinations for seismic event location. *Review Geophysics*, 3:157 – 185, 1965.

L. Geiger. Herdbestimmung bei erdbeden ans den ankunftzeiten. *K. Gessel. Wiss. Goett.*, 4:331–349, 1910.

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

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, 2004.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6): 721–741, November 1984.

Charles J. Geyer and Elizabeth A. Thompson. Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association*, 90 (431):909–920, September 1995.

Noah Goodman, Vikash Mansinghka, Daniel Roy, Keith Bonawitz, and Joshua Tenenbaum. Church: a language for generative models. In *UAI*, 2008.

B. Gutenberg and C.F. Richter. *Seismicity of the Earth and Associated Phenomena*. Princeton University Press, Princeton, N.J., 2nd edition, 1954.

International Seismological Centre. On-line Bulletin. `http://www.isc.ac.uk`, 2010. URL `http://www.isc.ac.uk`.

H. Jeffreys and K. E. Bullen. Seismological tables. *British Association for the Advancement of Science*, 1940.

T.H. Jordan and K.A. Sverdrup. Teleseismic location techniques and their application to earthquake clusters in the south-central pacific. *Bulletin of the Seismological Society of America*, 71:1105–1130, 1981.

B. L. N. Kennett and E. R. Engdahl. Traveltimes for global earthquake location and phase identification. *Geophysics Journal International*, 105:429 – 465, 1991.

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, may 1983.

T. Lay and T. C. Wallace. *Modern Global Seismology*. Academic Press, 1995.

R. Le Bras, H. Swanger, T. Sereno, G. Beall, and R. Jenkins. Global association. Technical Report ADA304805, Science Applications International Corp, San Diego, CA, Nov 1994a. URL `http://handle.dtic.mil/100.2/ADA304805`.

R. Le Bras, H. Swanger, T. Sereno, G. Beall, R. Jenkins, and W. Nagy. Global association: Design document and user's manual. Technical Report SAIC-94/1142, Science Applications International Corp, San Diego, CA, 1994b.

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

N. Magotra, N. Ahmed, and E. Chael. Seismic event detection and source location using single station, three component data. *Bulletin of the Seismological Society of America*, 77: 958 – 971, 1987.

E. Marinari and G. Parisi. Simulated tempering: A new monte carlo scheme. *Europhysics Letters*, 19:451–458, 1992.

W. Menke. *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, 1989.

Brian Milch and Stuart Russell. General-purpose MCMC inference over relational structures. In *UAI*, pages 349–358, Arlington, Virginia, 2006. AUAI Press.

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

Brian Milch, Bhaskara Marthi, David Sontag, Stuart Russell, Daniel L. Ong, and Andrey Kolobov. Approximate inference for infinite contingent Bayesian networks. In *In Proc. 10th AISTATS*, pages 238–245, 2005b.

S.C. Myers and C.A. Schultz. Improving sparse network seismic location with bayesian kriging and teleseismically constrained calibration events. *Bulletin of the Seismological Society of America*, 90:199–211, 2000.

Stephen C. Myers, Gardar Johannesson, and William Hanley. A Bayesian hierarchical method for multiple-event seismic location. *Geophysical Journal International*, 171:1049–1063, 2007.

Radford M. Neal. Slice sampling. *Annals of Statistics*, 31(3):705–767, 2003.

G.L. Pavlis and J.R. Booker. Progressive multiple event location (pmel). *Bulletin of the Seimological Society of America*, 73:1753–1777, 1983.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco, revised edition, 1988.

Avi Pfeffer. IBAL: A probabilistic rational programming language. In *In Proc. 17th IJCAI*, pages 733–740. Morgan Kaufmann Publishers, 2001.

Avi Pfeffer. The design and implementation of IBAL: A general purpose probabilistic language. In *Statistical Relational Learning*. MIT Press, 2007.

R. G. Roberts, A. Christoffersson, and F. Cassidy. Real-time event detection, phase identification and source location estimation using single station three-component seismic data. *Geophysical Journal*, 97:471 – 480, 1989.

C.A. Schultz, S.C. Myers, J. Hipp, and C. Young. Nonstationary bayesian kriging: a predictive technique to generate corrections for detection, location, and discrimination. *Bulletin of the Seismological Society of America*, 88:1275–1288, 1998.

R. W. Sittler. An optimal data association problem in surveillance theory. *IEEE Trans. Military Electronics*, MIL-8:125–139, 1964.

H.E. Soper, A.W. Young, B.M. Cave, A. Lee, and K. Pearson. On the distribution of the correlation coefficient in small samples appendix ii to the papers of "student" and r. a. fisher. a co-operative study. *Biometrika*, 11:328–413, 1917.

Sheila Vaidya, Robert Engdahl, Ronan Le Bras, Karl Koch, and Ola Dahlman. Strategic initiative in support of CTBT data processing: vDEC (virtual Data Exploitation Centre). In *CTBTO International Scientific Studies*, 2009.

V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

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