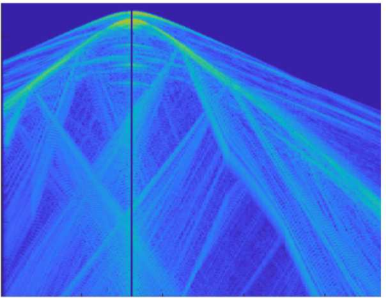
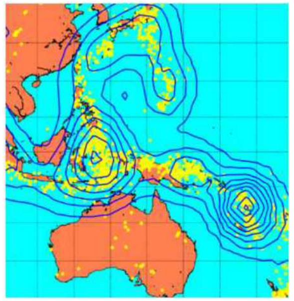
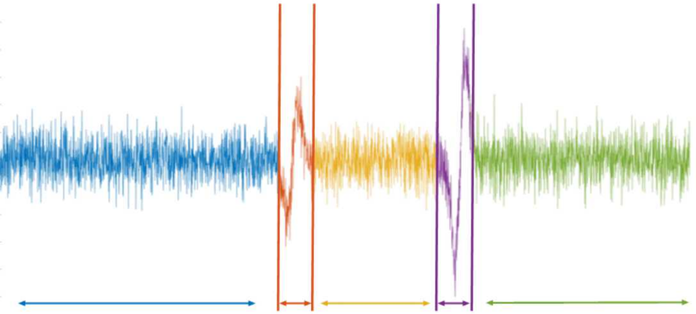
SAND2019-4228C







Feature-Based Bayesian Inference for Seismic Event Monitoring

Thomas A. Catanach

SSA Annual Meeting 2019

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND NO. 2017-XXXXP

Overview

- The Bayesian Perspective:
  - Probability distributions quantify uncertainty due to insufficient information
  - Bayesian methods for identification and estimation are critical to robust decision-making
- Target Contribution:
  - Take a Bayesian approach to waveform processing to detect and identify seismic events while integrating various sources of uncertainty to quantify confidence while identifying weak signals
  - Use a unique statistical framework and novel computational methods to make waveform-based Bayesian inference tractable
- General Approach:
  - Formulate an inference problem based upon predicting waveform features instead of the waveforms themselves
  - Simulate waveforms to build a statistical model of waveform features along with sources of feature uncertainty
  - Use Sequential Tempered Markov Chain Monte Carlo to efficiently identify events

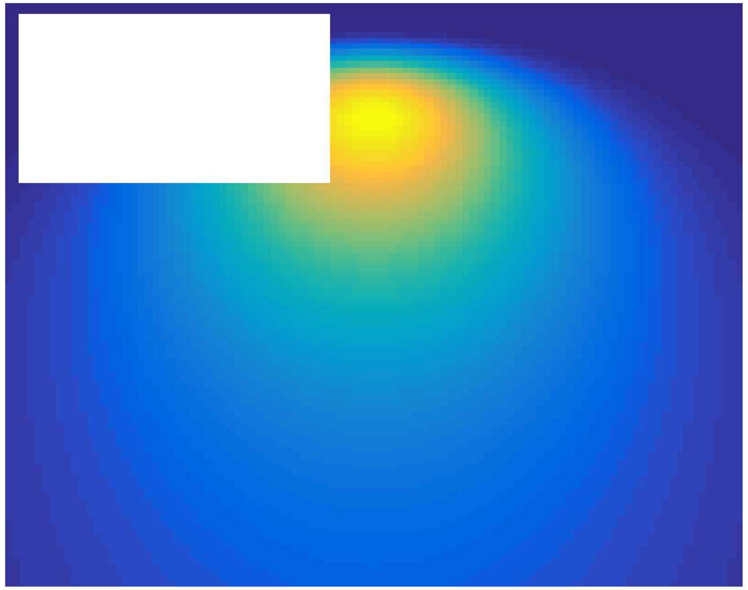
Talk Outline

1. Overview of Bayesian Inference and MCMC
2. Formulation of Seismic Monitoring as a Bayesian Inference Problem
3. Feature-Based Bayesian Inference
4. Building the Feature-Based Inference Workflow
5. Example with synthetic data
6. Future work and Conclusion

BAYESIAN INFERENCE AND MARKOV CHAIN MONTE CARLO

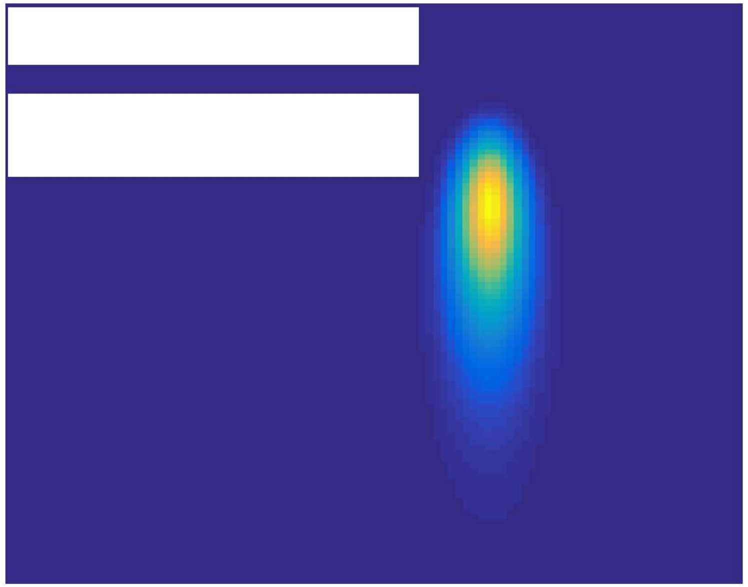
The Bayesian Inference Problem

Observations: $\mathcal{D}$



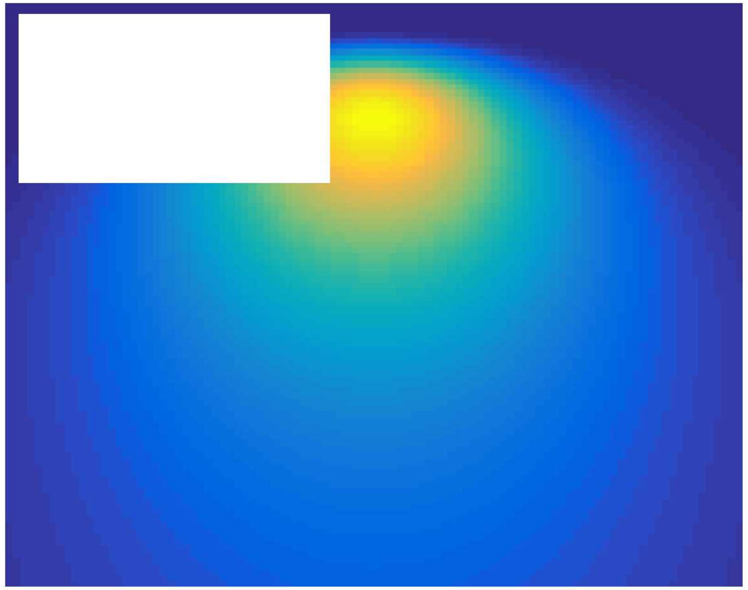
Bayes' Theorem

$$p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



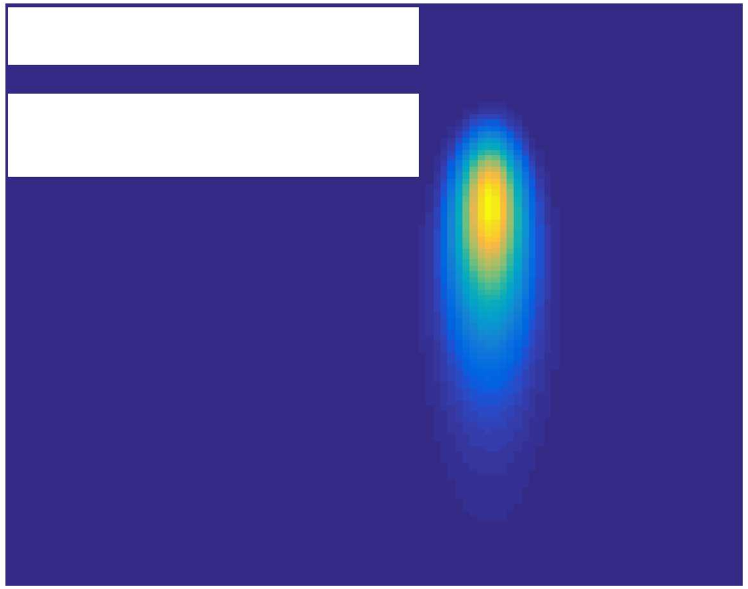
The Bayesian Inference Problem

Observations: $\mathcal{D}$



Bayes' Theorem

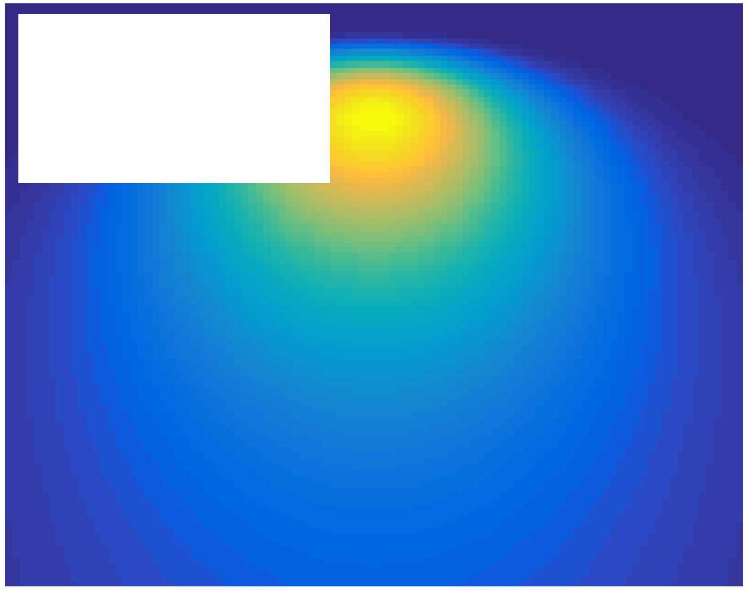
$$p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



$$p\left(\mathcal{D} \mid \mathcal{M}\right) = \underbrace{\int p\left(\mathcal{D} \mid \theta, \mathcal{M}\right) p\left(\theta \mid \mathcal{M}\right) d\theta}_{\text{Intractable}}$$

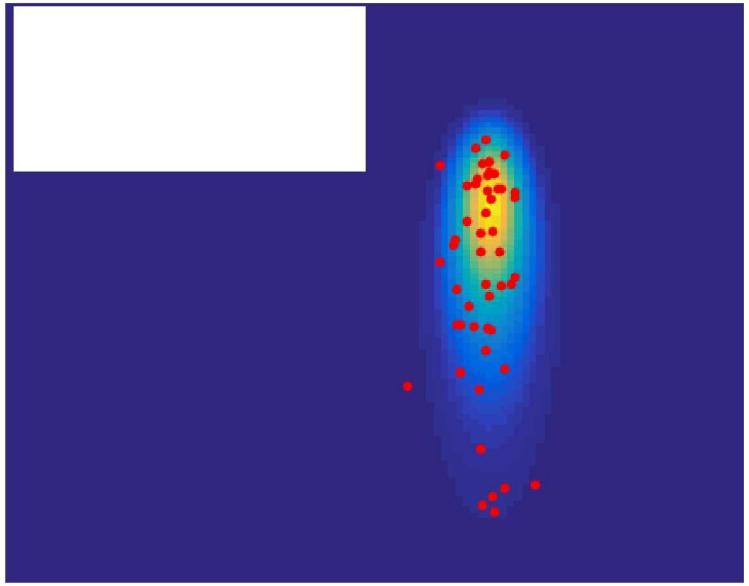
The Bayesian Inference Problem

Observations: $\mathcal{D}$



Bayes' Theorem

$$p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$

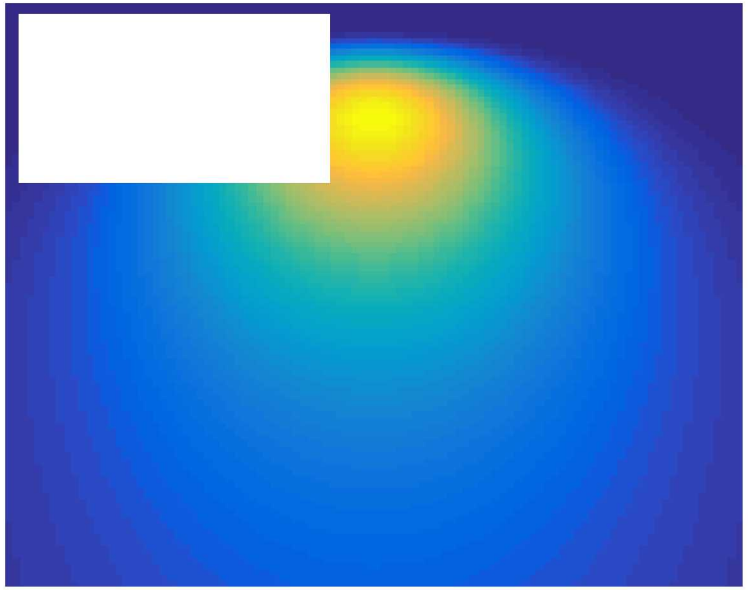


Posterior Estimation:

$$\mathbb{E}\left[g\left(\theta\right) \mid \mathcal{D}, \mathcal{M}\right] = \int g\left(\theta\right) p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) d\theta \approx \frac{1}{N} \sum_{i=1}^{N} g\left(\theta_i\right)$$

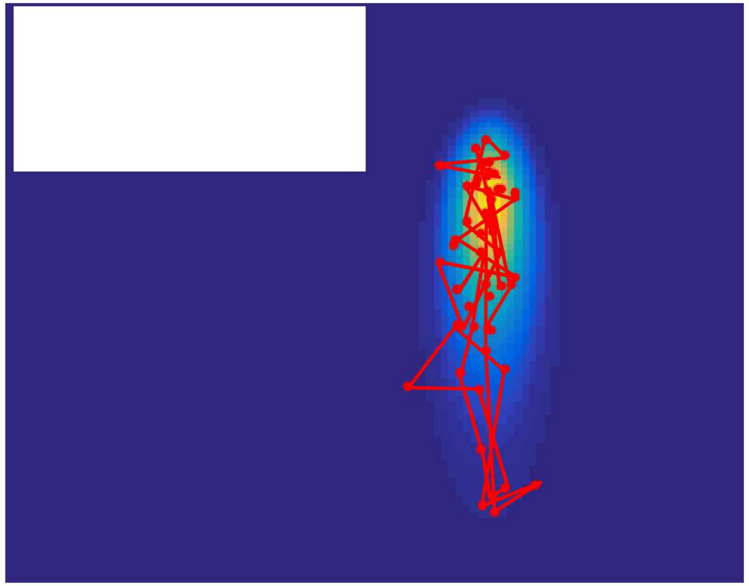
The Bayesian Inference Problem

Observations: $\mathcal{D}$



Bayes' Theorem

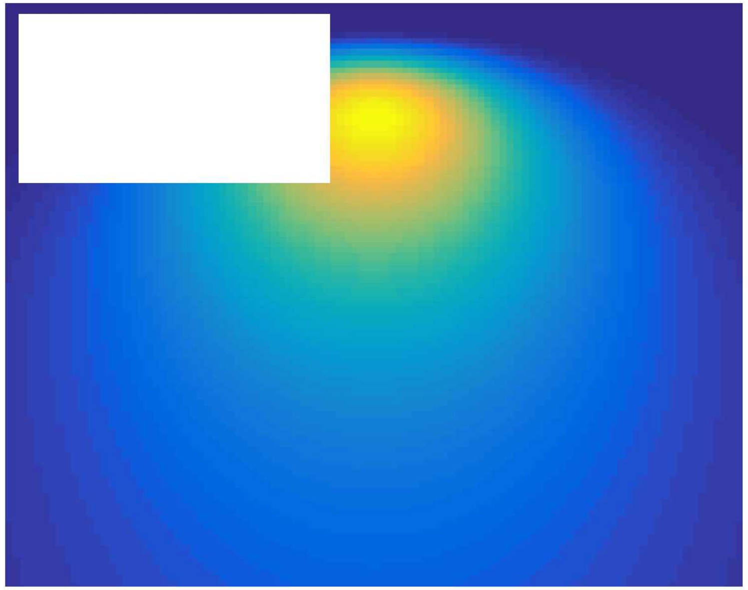
$$p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



Exploration of the space by proposal distribution	+	Accept/Reject correction	=	Metropolis-Hastings MCMC

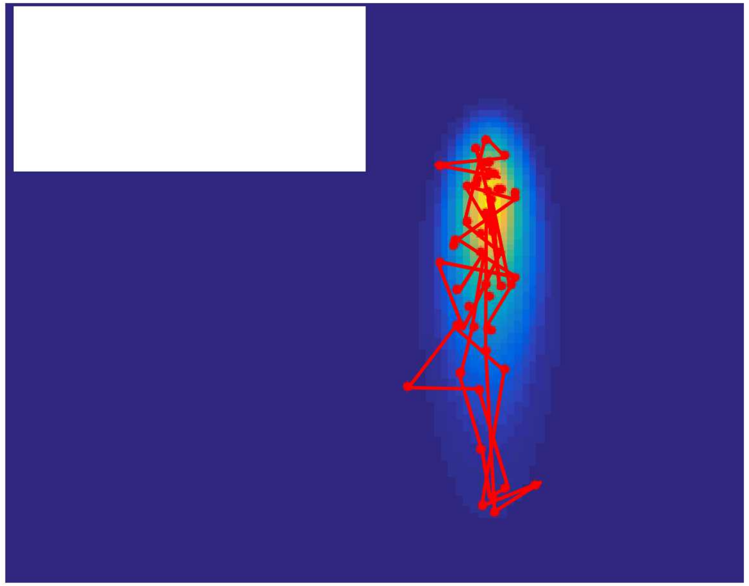
The Bayesian Inference Problem

Observations: $\mathcal{D}$



Bayes' Theorem

$$p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



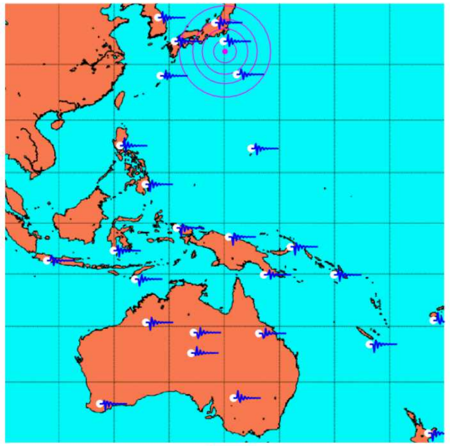
Posterior Estimation: $$\mathbb{E}\left[g\left(\theta\right) \mid \mathcal{D}, \mathcal{M}\right] = \int g\left(\theta\right) p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) d\theta \approx \frac{1}{N}\sum_{i=1}^{N} g\left(\theta_i\right)$$

Effective Number of Samples: $$ESS\left[g\left(\theta_{1:N}\right)\right] = \frac{var\left[g\left(\theta\right)\right]}{var\left[\frac{1}{N}\sum_{i=1}^{N} g\left(\theta_i\right)\right]}$$

BAYESIAN INFERENCE FOR SEISMIC EVENT LOCALIZATION

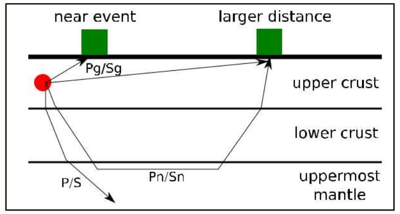
The Bayesian Framework

Data: $\mathcal{D}$



Likelihood: $p(\mathcal{D} \mid \theta)$

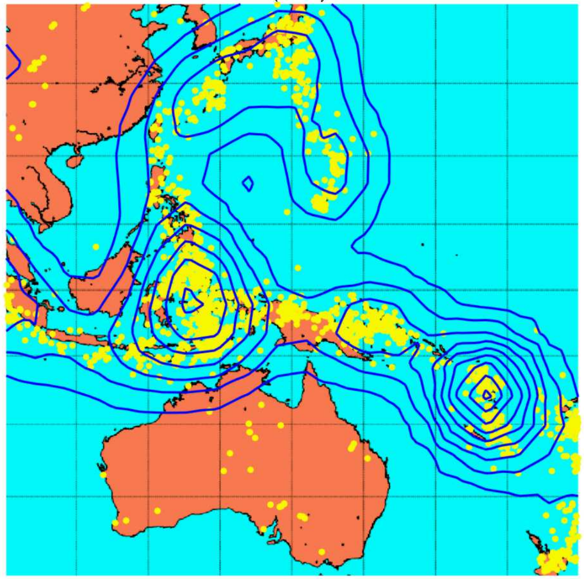
Physics Model
Sensor Model
Uncertainty Model



Bayes' Theorem:

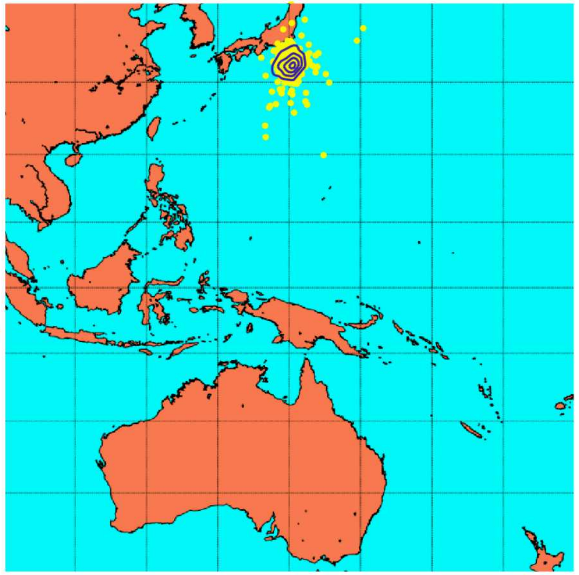
$$p(\theta \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \theta)\, p(\theta)}{p(\mathcal{D})}$$

Prior: $p(\theta)$



Knowledge about where events are likely to occur

Posterior: $p(\theta \mid \mathcal{D})$



Updated knowledge about where a specific event occurred

Background on existing Bayesian Seismic Monitoring Methods

Detection-Based

- Description
  - Each station pre-processes their observed waveforms to extract arrival “picks”
  - The likelihood of an event (or events) is based upon how well the observed arrival times correspond to arrivals from seismic waves generated by the event hypothesis
  - Arrival time and detection uncertainty can be integrated into the model
- Examples: BayesLoc[1], NET-VISA[2]
- Advantages
  - Requires only a model of travel-time and not the waveform
- Disadvantages
  - Events that produce weak signals below the pick threshold cannot be detected, even when many sensors are combined

[1]Myers, S. C., Gardar Johannesson, and Robert J. Mellors. “BayesLoc: A robust location program for multiple seismic events given an imperfect earth model and error-corrupted seismic data” (2011)

[2]Arora, Nimar S., Stuart Russell, and Erik Sudderth. "NET-VISA: Network processing vertically integrated seismic analysis" (2013)

Background on existing Bayesian Seismic Monitoring Methods

Signal-Based

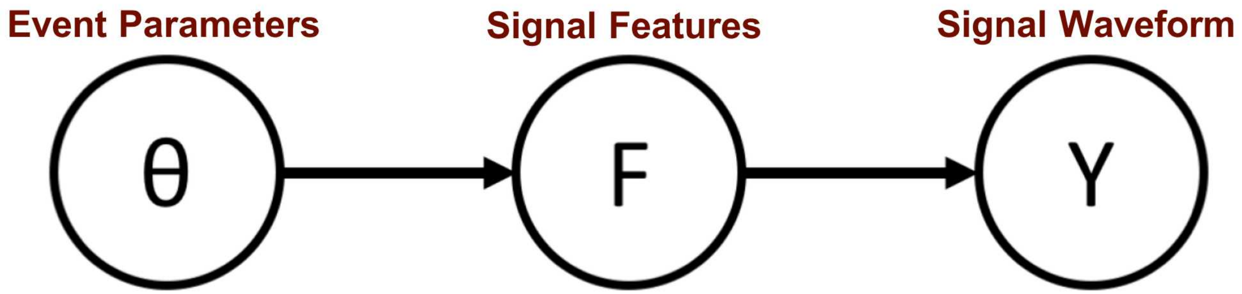
- Description
  - The likelihood of a candidate event (or events) is based upon comparing predicted waveforms given the event hypothesis, noise process, and other modeled uncertainty to the observed waveforms
- Example: SIG-VISA[3]
- Advantages
  - Can integrate many sensors to detect low magnitude signals
  - Waveform characteristics can contain useful information for event identification
- Disadvantages
  - Requires learning and evaluating a generative model of the full waveform to compute the likelihood of the observed signal

[3] Moore, David A., and Stuart J. Russell. "Signal-based Bayesian seismic monitoring" (2017)

FEATURE-BASED INFERENCE

Defining feature-based inference

Graphical Model:



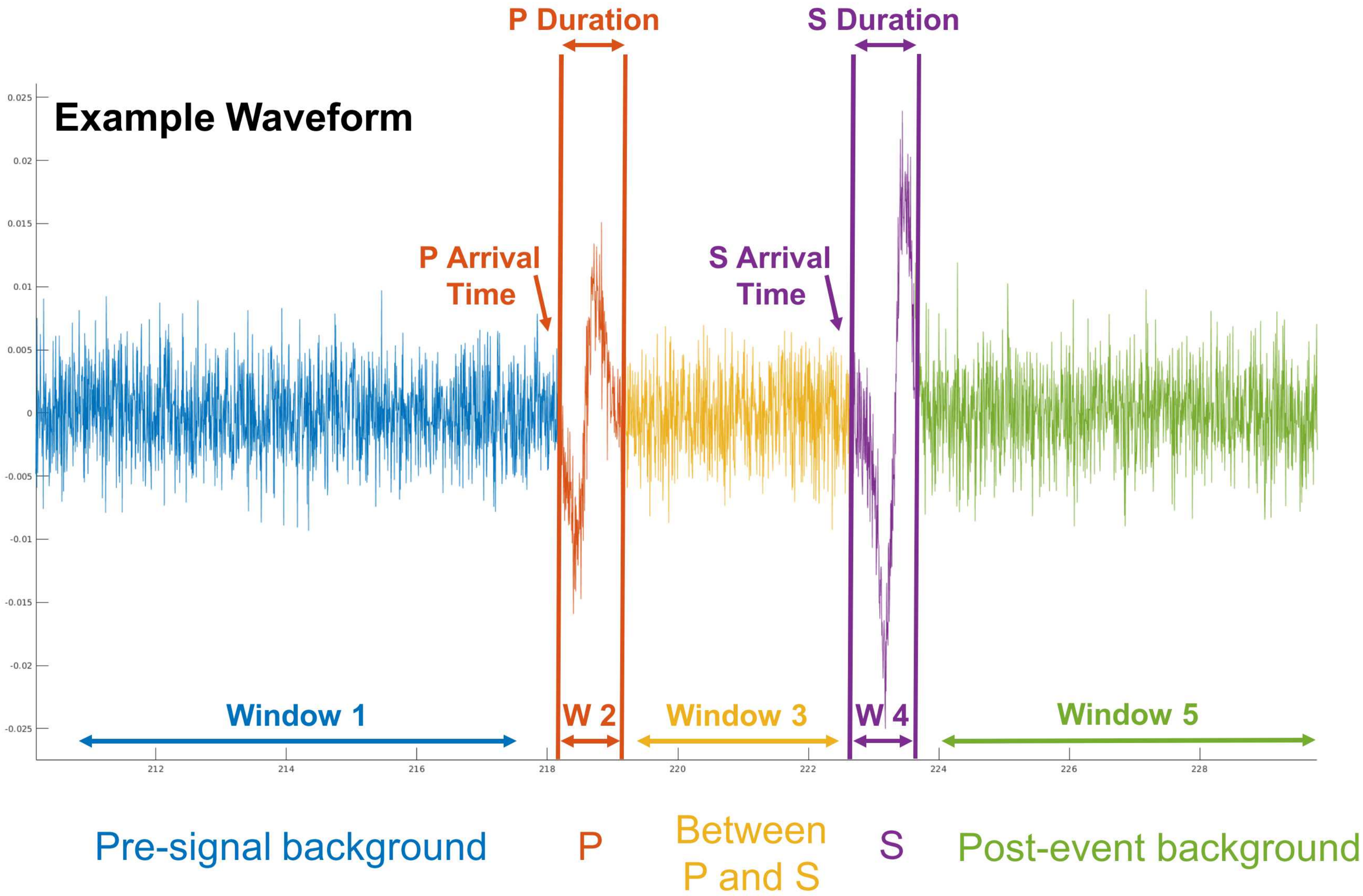
Bayesian Inference:

- Feature-based inference requires building statistical models for the likelihood of a signal given certain features and the likelihood of those features given an hypothesized event parameterization

$$\underbrace{p(\theta \mid Y)}_{\text{Posterior}} = \frac{\overbrace{p(Y \mid \theta)}^{\text{Likelihood}}\overbrace{p(\theta)}^{\text{Prior}}}{\underbrace{p(Y)}_{\text{Evidence}}} = \underbrace{\left(\int \overbrace{p(Y \mid F)}^{\text{Signal Likelihood}}\overbrace{p(F \mid \theta)}^{\text{Feature Likelihood}} dF\right)}_{\text{Marginalize over features}} \frac{p(\theta)}{p(Y)}$$

Feature Based Inference for Seismic Monitoring

- Waveform Features
  - P and S arrival time
  - Waveform feature within window e.g. total signal power
- P and S arrival times and uncertainty can be found using models like AK135
- We can build a statistical model for the signal power using simulations and background models



Waveform Likelihood based on arrivals and signal power feature

Bayesian Inference:

$$\underbrace{p(\theta \mid Y)}_{\text{Posterior}} \propto \underbrace{p(\theta)}_{\text{Prior}} \int p(Y \mid \underbrace{t_p, t_s, P_{1:5}}_{\text{Features: P and S Arrivals and Window Powers}})\, p(t_p, t_s, P_{1:5} \mid \theta)\, dt_p dt_s dP_{1:5}$$

$$= p(\theta) \int \underbrace{p(Y_{1:5} \mid t_p, t_s, P_{1:5})}_{\text{Uniform}}\, \underbrace{p(P_{1:5} \mid t_p, t_s, \theta)}_{\text{Simulations and Background process}}\, \underbrace{p(t_p, t_s \mid \theta)}_{\text{Travel time model}}\, dt_p dt_s dP_{1:5}$$

Assuming conditional independence:

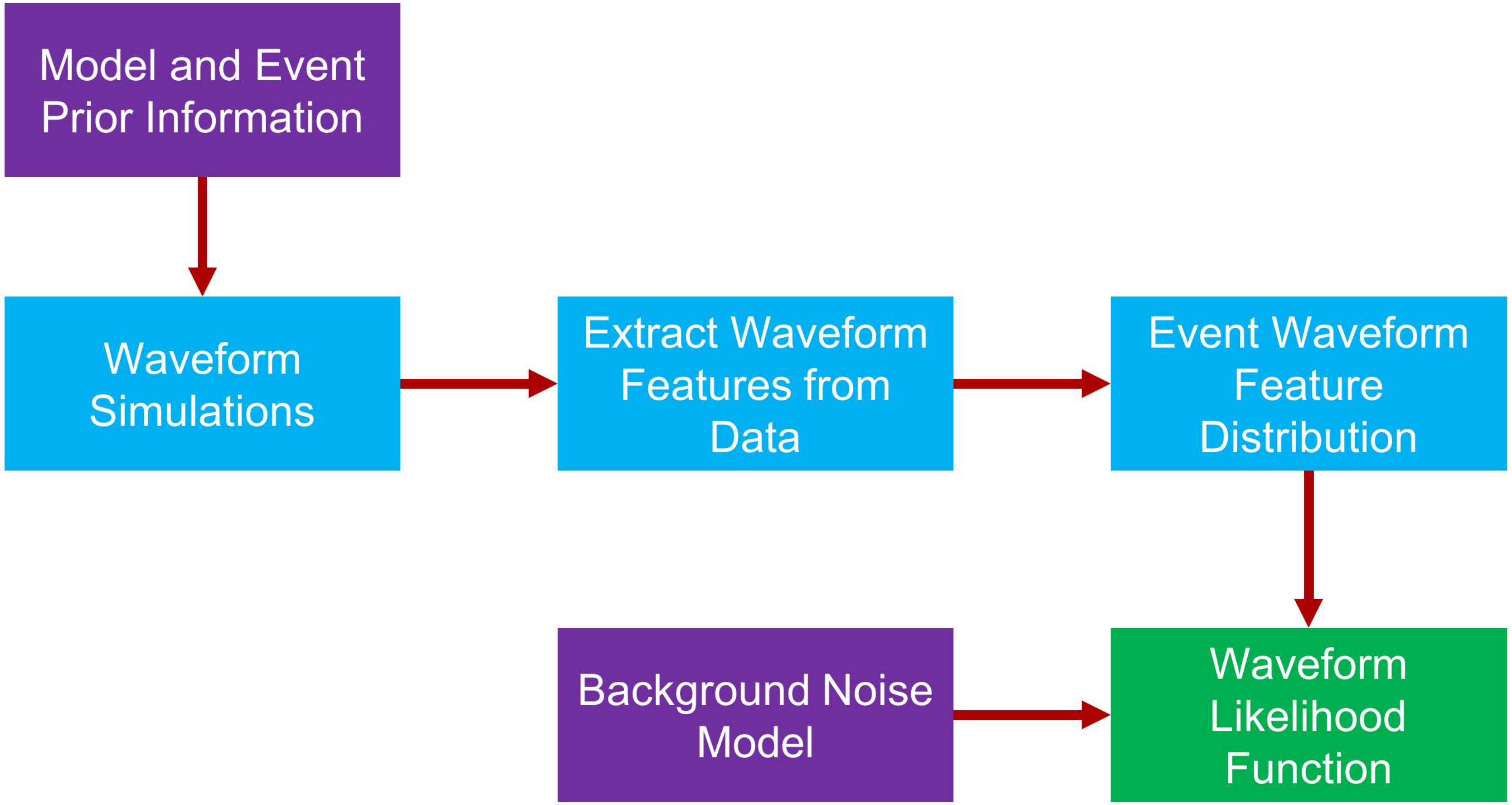
$$p(Y_{1:5} \mid t_p, t_s, P_{1:5}) = \prod_{i=1}^{5} p(Y_i \mid t_p, t_s, P_i)$$

Uniform waveform distribution conditioned on features:

$$p(Y \mid t_p, t_s, P) = \frac{1}{\underbrace{\int \delta(Y^T Y - P)\, p(Y)\, dY}_{\text{Quantifies the size of the signal space with given features}}} \propto \frac{\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{n}{2}} P^{\frac{n}{2}-1}}$$

BUILDING FEATURE-BASED WORKFLOW

Data driven workflow



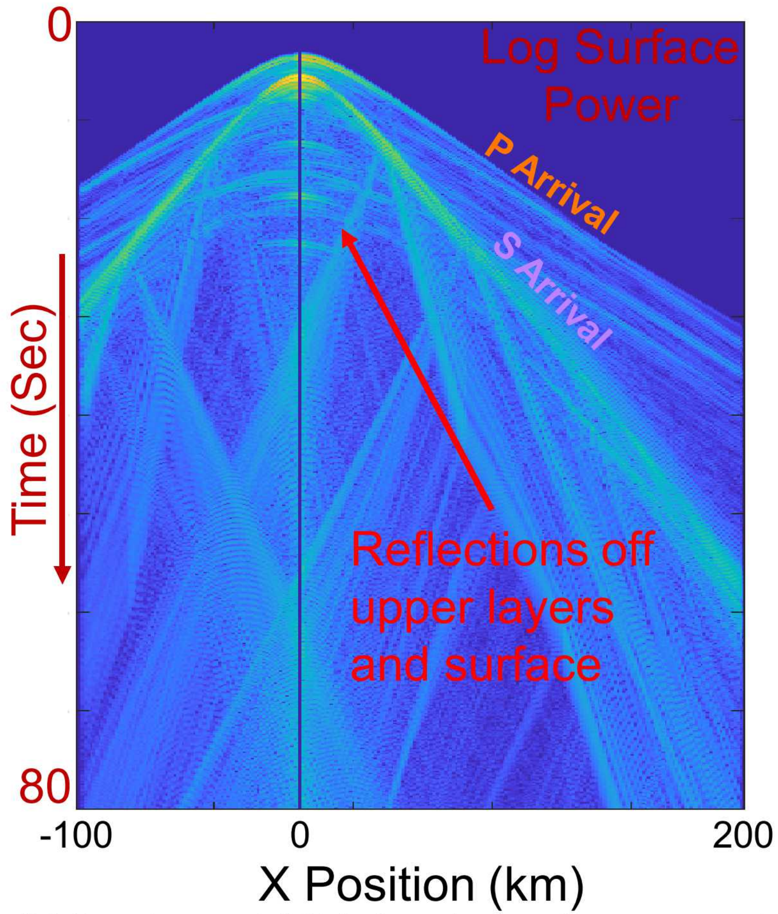
Bayesian Inference Problem

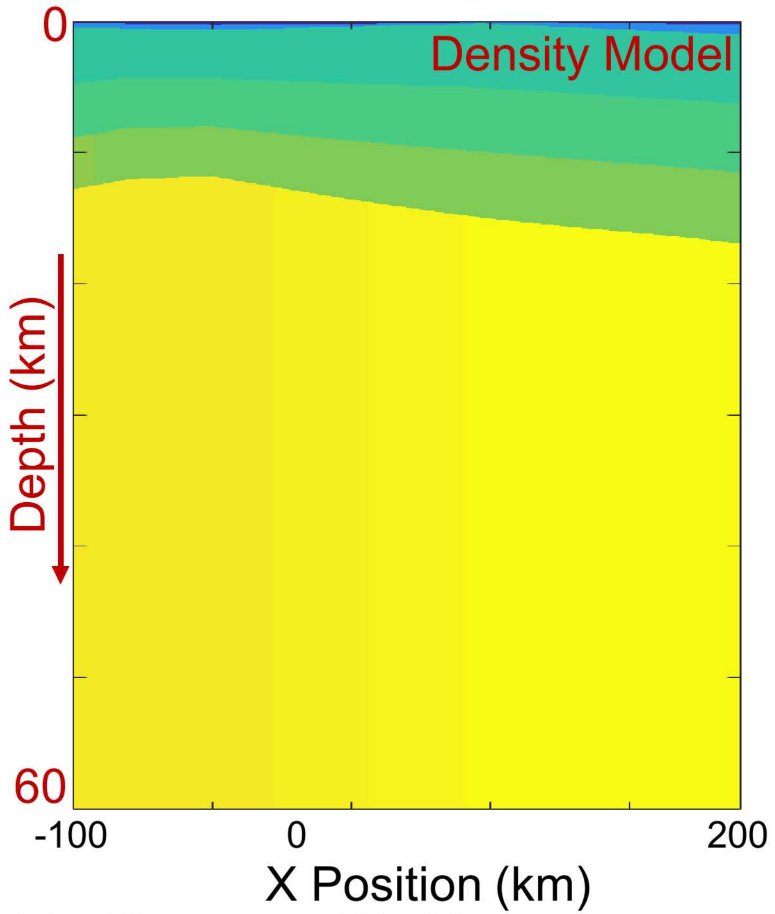
- Parameters
  - Event Parameters: Latitude, Longitude, Depth, Magnitude, Origin Time
  - Uncertainty Parameters: Travel time uncertainty
- Feature Model
  - AK135 for mean travel time and approximate travel time uncertainty
  - Waveform Simulations build signal power distribution as a function of distance from the source and marginalize over sources of uncertainty like focal mechanism.
- Background Noise Process
  - Assume a process modeled as a stationary Gaussian process within each window with known covariance
  - Independent of the event signal

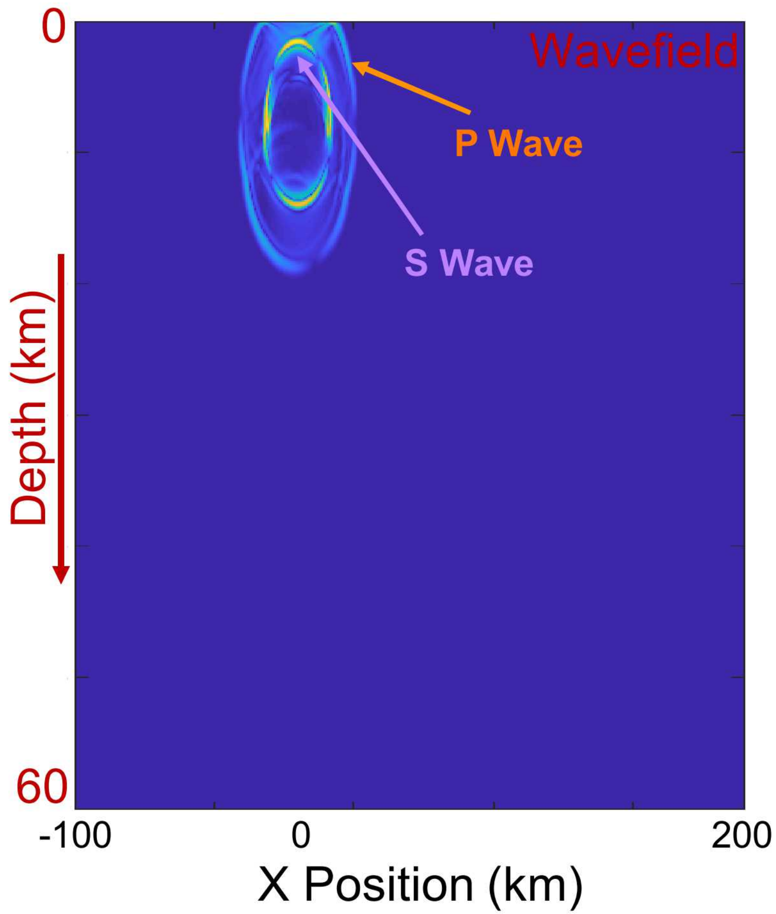
Building Feature Model

Simulation Environment

- 2D waveform simulations[4] on 300 km x 60 km domain from Crust 1.0 cross-sections of Utah
- Simulated 1k events at 10 sensors with uniformly distributed event and focal mechanism parameters



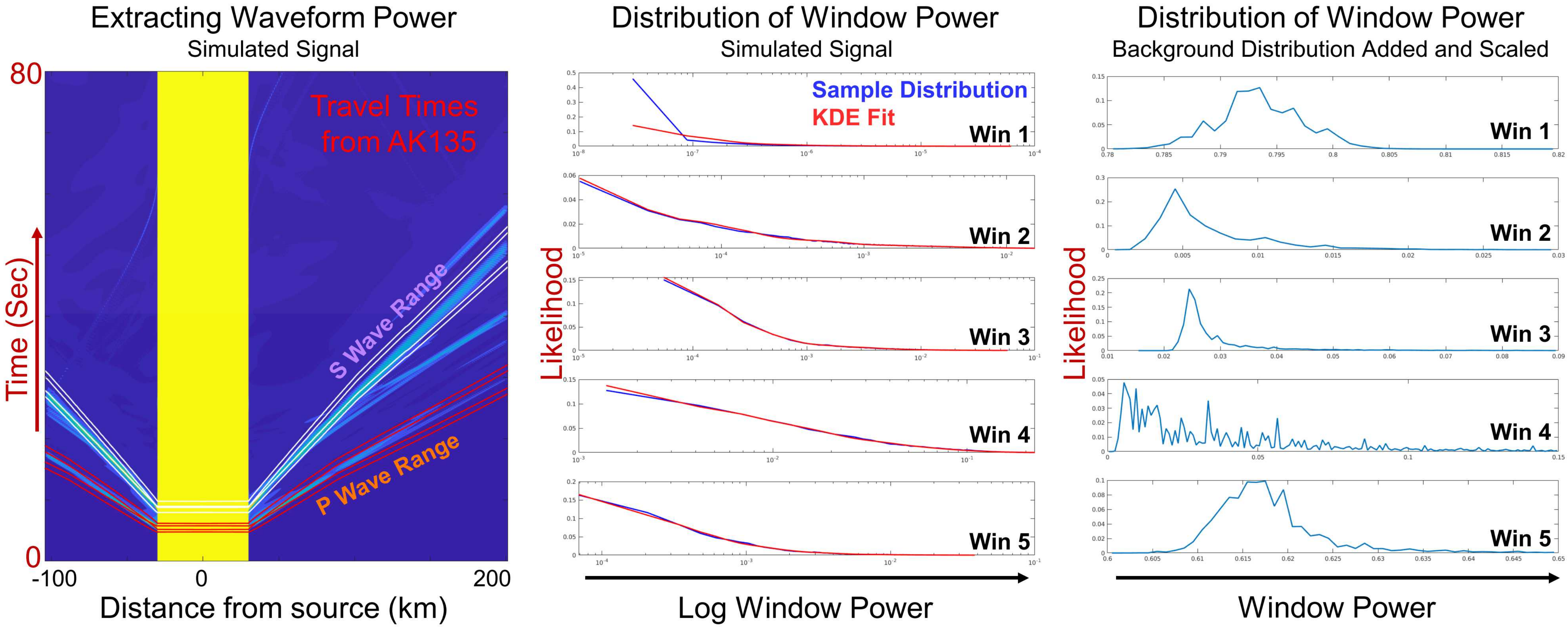




[4]Li, Dunzhu, et al. "Global synthetic seismograms using a 2-D finite-difference method." (2014)

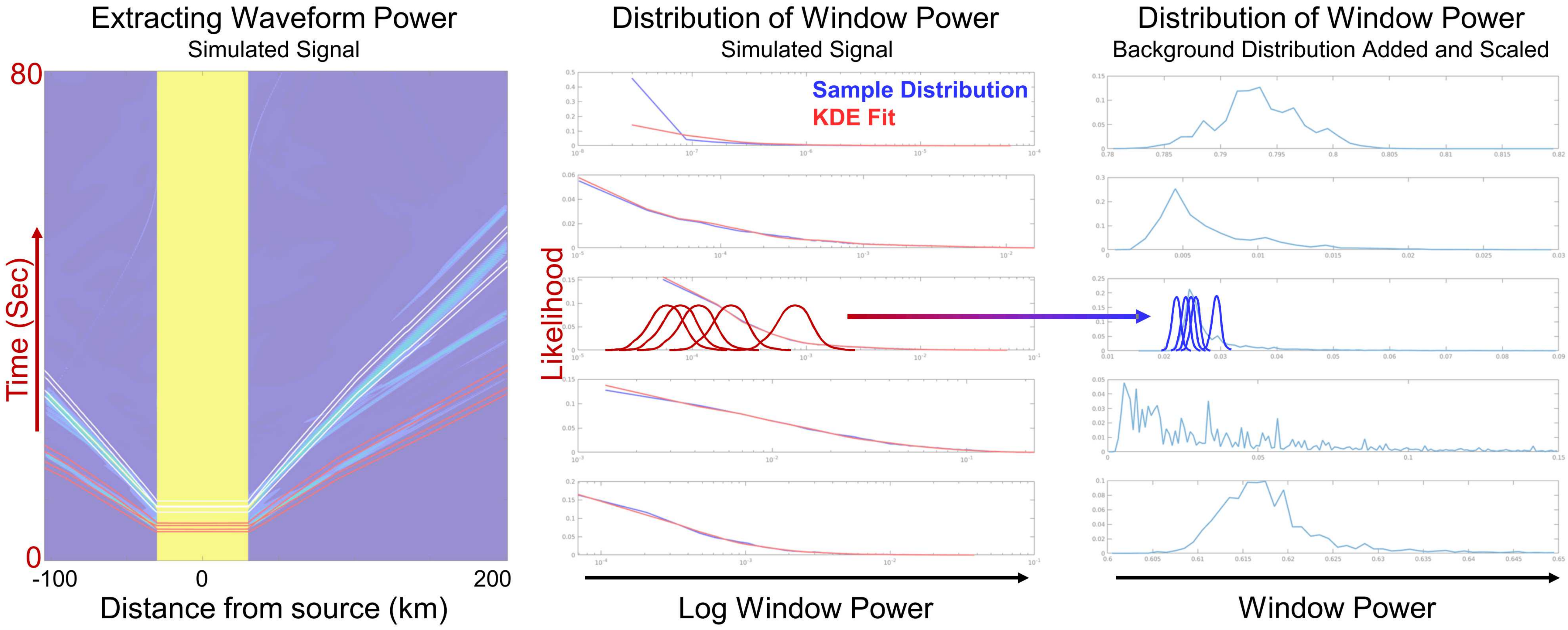
Building Features: Extracting Waveform Features and KDE Model

- P and S travel times, uncertainty, and assumed duration define possible window arrangements
- The window arrangement which contains the maximum event power is used to build the KDE



Building Features: Extracting Waveform Features and KDE Model

- The event magnitude and background process change the distribution of window power. Assuming a Gaussian background process, this can be modeled as shifting and scaling the KDE kernels.



Markov Chain Monte Carlo

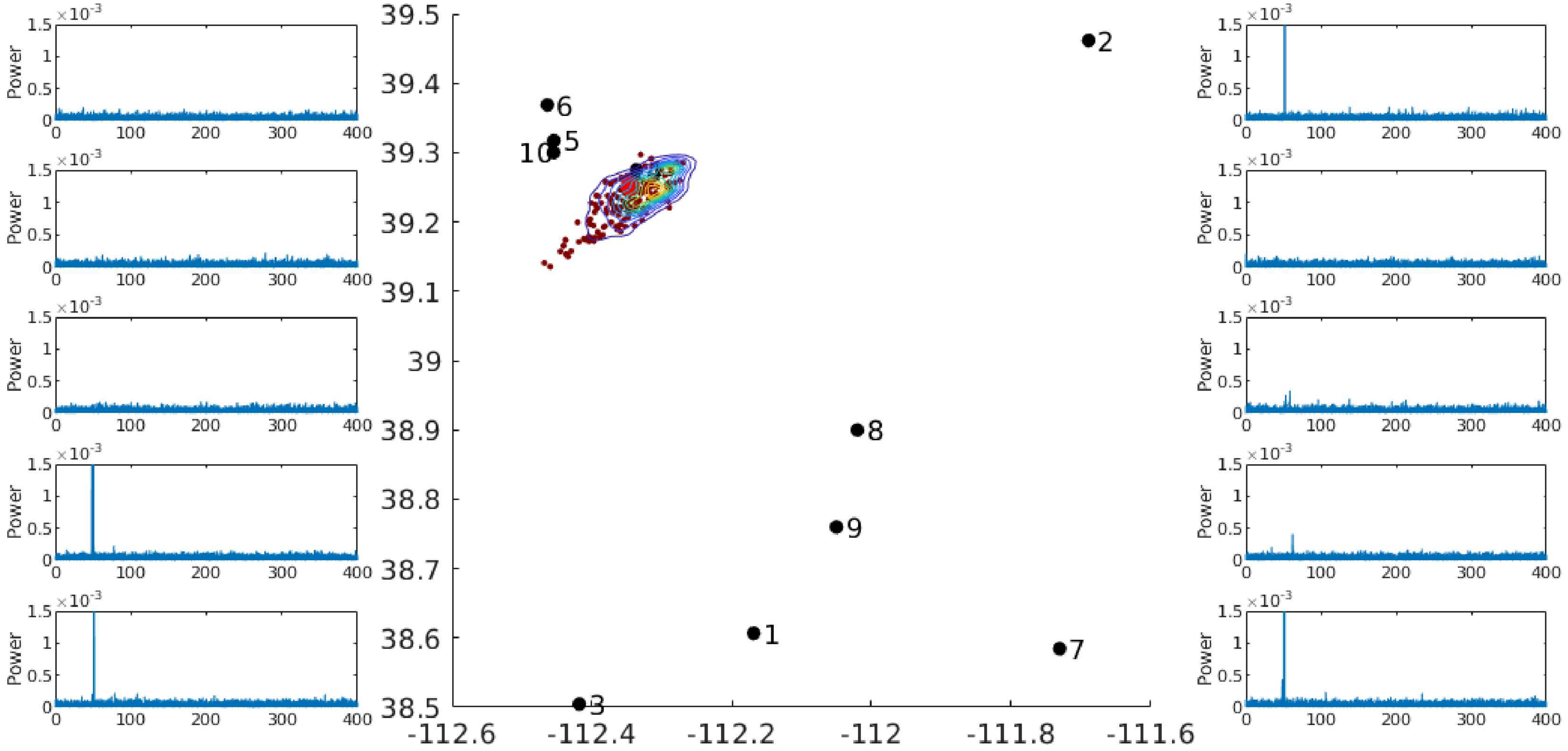
- Sequential Tempered MCMC[5,6]
  - Combines Annealing, Importance Sampling, and MCMC into a single algorithm to efficiently sample the posterior
  - Enables parallel sampling to utilize HPC and model evidence estimates for event detection
  - The annealing schedule can be tuned to avoid poorly identified posterior distribution when only a limited number of sensors influence the likelihood

- Pseudo-Marginal MCMC[7]
  - Enables better uncertainty quantification by using unbiased estimate of the likelihood while still maintaining the posterior distribution.
  - Therefore we can more marginalize over sources of uncertainty e.g. travel time uncertainty.
  - Adaptive methods can be integrated into ST-MCMC to better importance sample the travel time distributions and determine how many trials are needed to get a reasonable likelihood estimate.

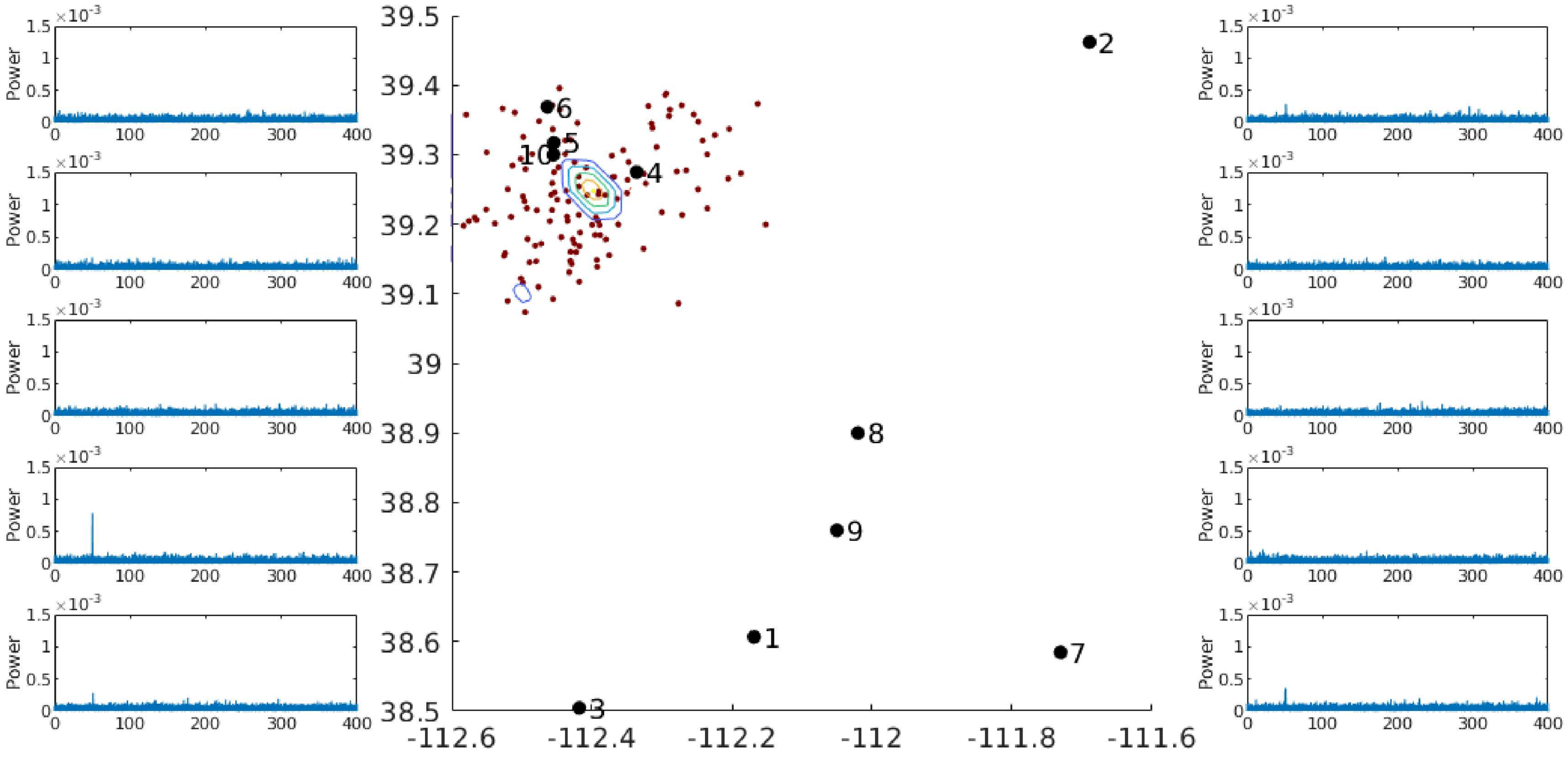
[5]Catanach, T A., and J. L. Beck "Bayesian updating and uncertainty quantification using sequential tempered MCMC with the rank-one modified metropolis algorithm" (2018)
[6]Minson, S. E., M. Simons, and J. L. Beck "Bayesian inversion for finite fault earthquake source models I—Theory and algorithm" (2013)
[7]Andrieu, Christophe, and Gareth O. Roberts "The pseudo-marginal approach for efficient Monte Carlo computations" (2009)

SYNTHETIC EXAMPLE

Example 1: Well identified strong signal



Example 2: Weak signal with more variance



FUTURE WORK AND CONCLUSION

Future Work

- Application and Validation
  - Multiple events and event model selection
  - Compare with Detection-Based and Signal-Based methods
- Better Uncertainty Quantification
  - Integrate complex description of the background noise process
  - Spatial correlation between sensors for more complex arrival time and power uncertainty models
- Richer Features
  - Integrate directional features
  - Preprocess signals to make extracting meaningful features easier such as performing STA/LTA

Conclusion

- Bayesian inference provides a natural way to express and propagate uncertainty for seismic monitoring and decision-making
- Feature-based inference provides a promising approach to signal-based full waveform monitoring that reduces the complexity of the statistical problem
- Advanced MCMC techniques can be employed to reduce the computational burden of the Bayesian inference problem and allow for the explicit integration of uncertainty

BACKUP SLIDES

Sequential Tempered MCMC

- ST-MCMC methods combine:
  1) Annealing: Introduce intermediate distributions
  2) MCMC: Explore the intermediate distributions
  3) Importance Resampling: Discard unlikely chains and multiply likely chains while maintaining the distribution
- Examples: SMC[1], Subset Simulation[2], TMCMC[3], AlTar/Catmip[4], AIMS[5], and AMSSA[6]

[1] Del Moral et al 2006
[2] S.K. Au and J.L. Beck 2001
[3] J. Ching and Y. C. Chen 2007
[4] J.L. Beck and K.M. Zuev 2013
[5] S.E Minson, M. Simons, J.L. Beck 2013
[6] E. Prudecio and S.H. Cheung 2012

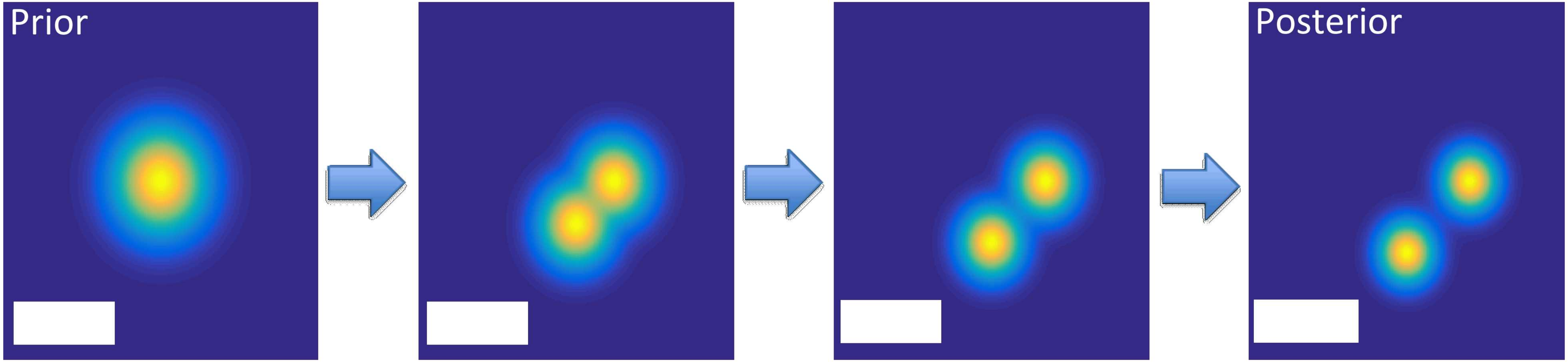
Annealing

β defines how much the data updates the intermediate distribution:

$$\pi_i(\theta) \propto p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_i} \, p(\theta \mid \mathcal{M}) \quad \beta_i \in [0, 1]$$

Intermediate distributions at different β levels

Level 0: $\beta_0 = 0$ Level 1: $\beta_1 = \beta_0 + \Delta\beta_1$ Level 2: $\beta_2 = \beta_1 + \Delta\beta_2$ Level n: $\beta_n = 1$



Annealing: Finding Δβ

Find Δβ such that the coefficient of variation (κ) of the sample weights is 1

Sample weight: $w\left(\theta_j\right) \propto p\left(\mathcal{D} \mid \theta_j, \mathcal{M}\right)^{\Delta\beta_i}$

Coefficient of variation: $\kappa\left(w\right) = \frac{\sigma(w)}{\bar{w}}$

Importance Resampling

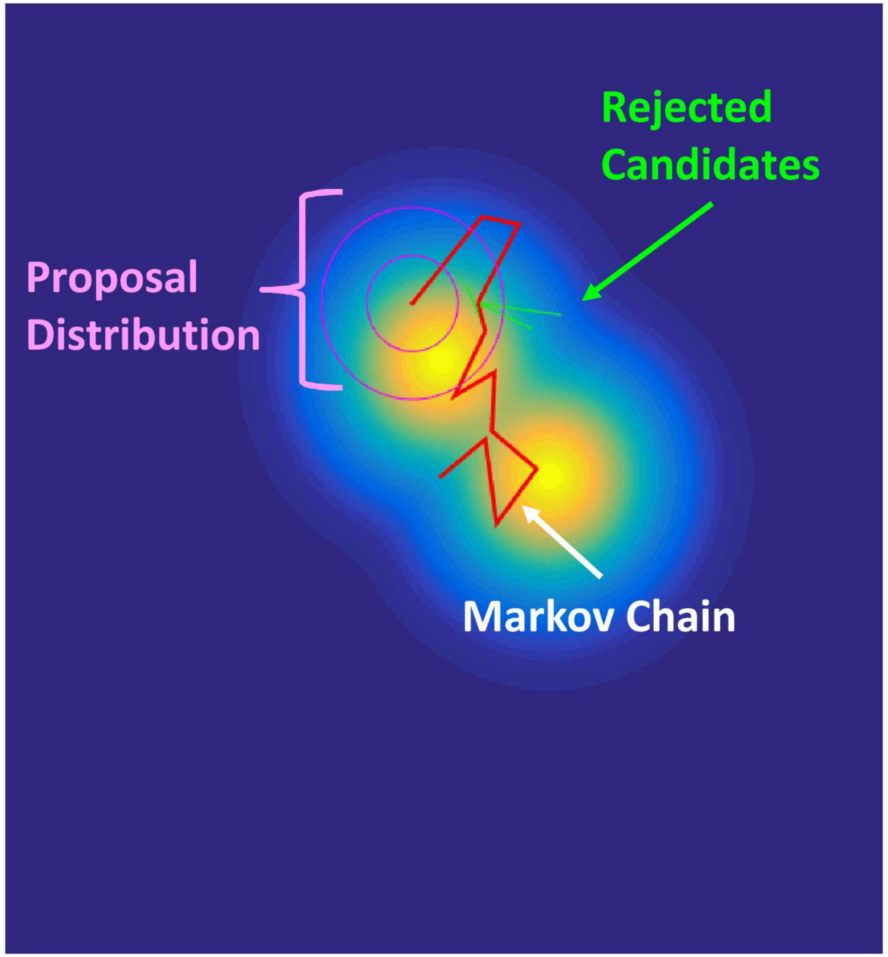
- Resampling the population rebalances the weights as the distribution changes. This discards unlikely samples and replicates likely samples
- Multinomial Resampling from level i-1 to level i:

Probability of selecting sample k: $P\left(\theta_{i,j} = \theta_{i-1,k}\right) = w\left(\theta_{i-1,k}\right)$

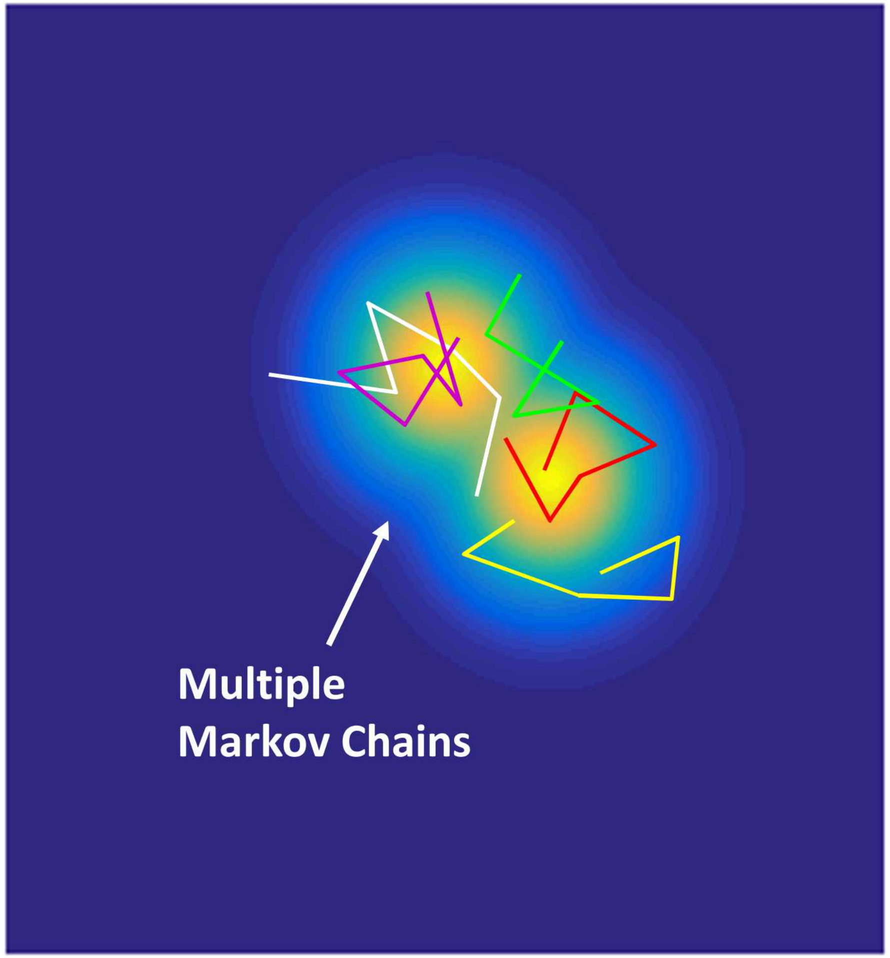
Sample weight: $w\left(\theta_{i-1,j}\right) \propto p\left(\mathcal{D} \mid \theta_{i-1,j}, \mathcal{M}\right)^{\Delta\beta_i}$

Metropolis Hastings MCMC with Parallel Chains

Single MH Markov Chain



Parallel MH Markov Chain



Designing the ST-MCMC Algorithm

- Algorithm Parameters
  - Number of parallel Markov Chains
  - Chain Length or target correlation
  - Annealing/convergence rate i.e. coefficient of variation target
- MCMC Algorithm
  - Freedom to choose the proposal distribution and its properties
  - Design of the Markov Chain kernel
- Resampling scheme for importance sampling