The Bayesian Approach to Association

...

(AGU-2017: S23D-07)

Nimar S. Arora (Bayesian Logic, Inc.)

The Association Problem

Hypothesize an event, find strong evidence to support it

Goodness of Fit Approach

Hypothesize ***any*** event, find **strong *positive* evidence** to support it.

Evidence can be:

- An STA/LTA-based detection with measured arrival parameters close to predicted parameters (phase, time, azimuth, slowness)
- A high correlation with a prior waveform from an event in a nearby location

Strength:

- Rules written down by experts (e.g. 3 detections with P phase arrival time within 2 seconds of predicted)
- Goodness of fit formula

Bayesian Approach

Write down a generative model that explains ***all*** of the data (that could be used as evidence).

Include as much of the seismological knowledge into the model as appropriate.

Calibrate model components on historical data.

Hypothesize an event that

- explains all the data
- is better than the alternate explanations for the data (such as noise, other events)

Bayesian Formula (for a fixed alternate hypothesis)

H -> Hypothesis, A -> Alternate Hypothesis, D -> Data

$P(H \mid D) = P(H) P(D \mid H) / \{ P(H) P(D \mid H) + P(A) P(D \mid A) \}$

Proportional to : $P(H) \{ P(D \mid H) / P(D \mid A) \}$

Bayesian Formula (for a fixed alternate hypothesis)

H -> Hypothesis, A -> Alternate Hypothesis, D -> Data

$$P(H \mid D) = P(H)\, P(D \mid H) \,/\, \{ P(H)\, P(D \mid H) + P(A)\, P(D \mid A) \}$$

Proportional to : $P(H)\ \{ P(D \mid H) / P(D \mid A) \}$

Strength of evidence -- calibrated by historical data

Prior belief in hypothesis = everything we know of seismology

- Travel time tables
- Earth's natural seismicity
- Amplitude decay curves
- etc.

Strength of the evidence

P (D | H) / P (D | A)

P (D | A) : odds of random noise explaining the observed detections

- Some stations are very noisy
- Similarly some waveform templates correlate with almost everything!

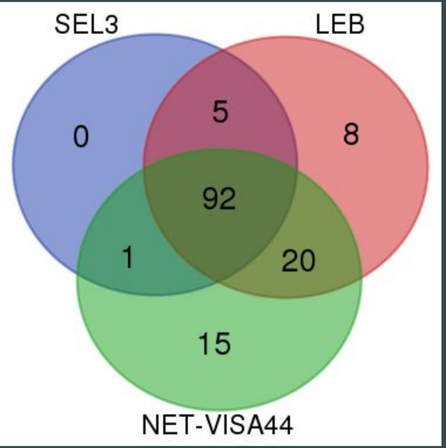
P (D | A) : also considers the odds of other events explaining these detections

- The coda of large events often contain multiple detections

P (D | H) : also includes the mis-detection probability, plus phase misclassification

Results of NET-VISA at IDC

- Running continuously from 2011 to present
- Detected all DPRK nukes
- Detects 88% of events in LEB bulletin with 45-50% “false” events
  - GA detects 70% of events in LEB with 50% false events
- “False” events contain real events verifiable through ISC bulletin

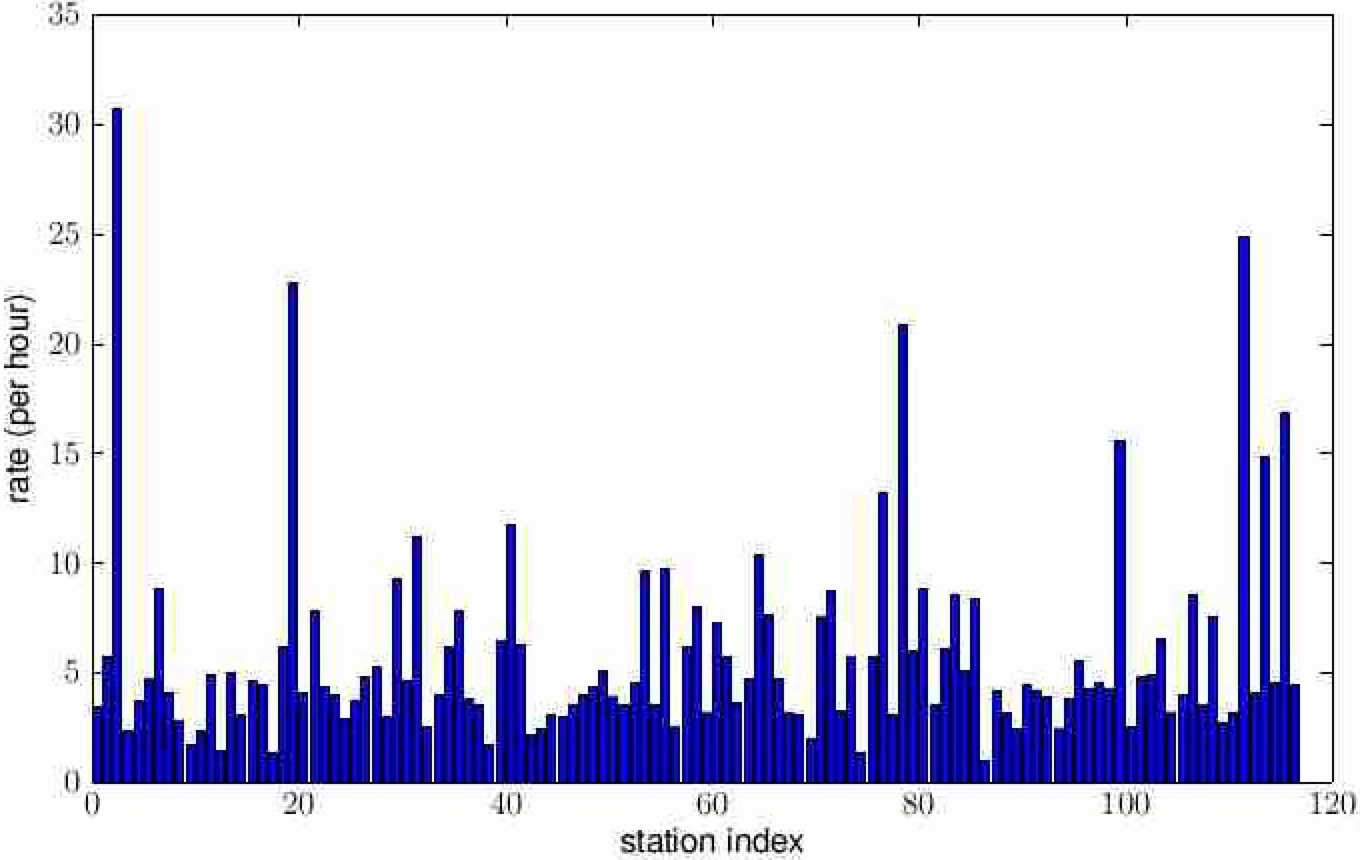


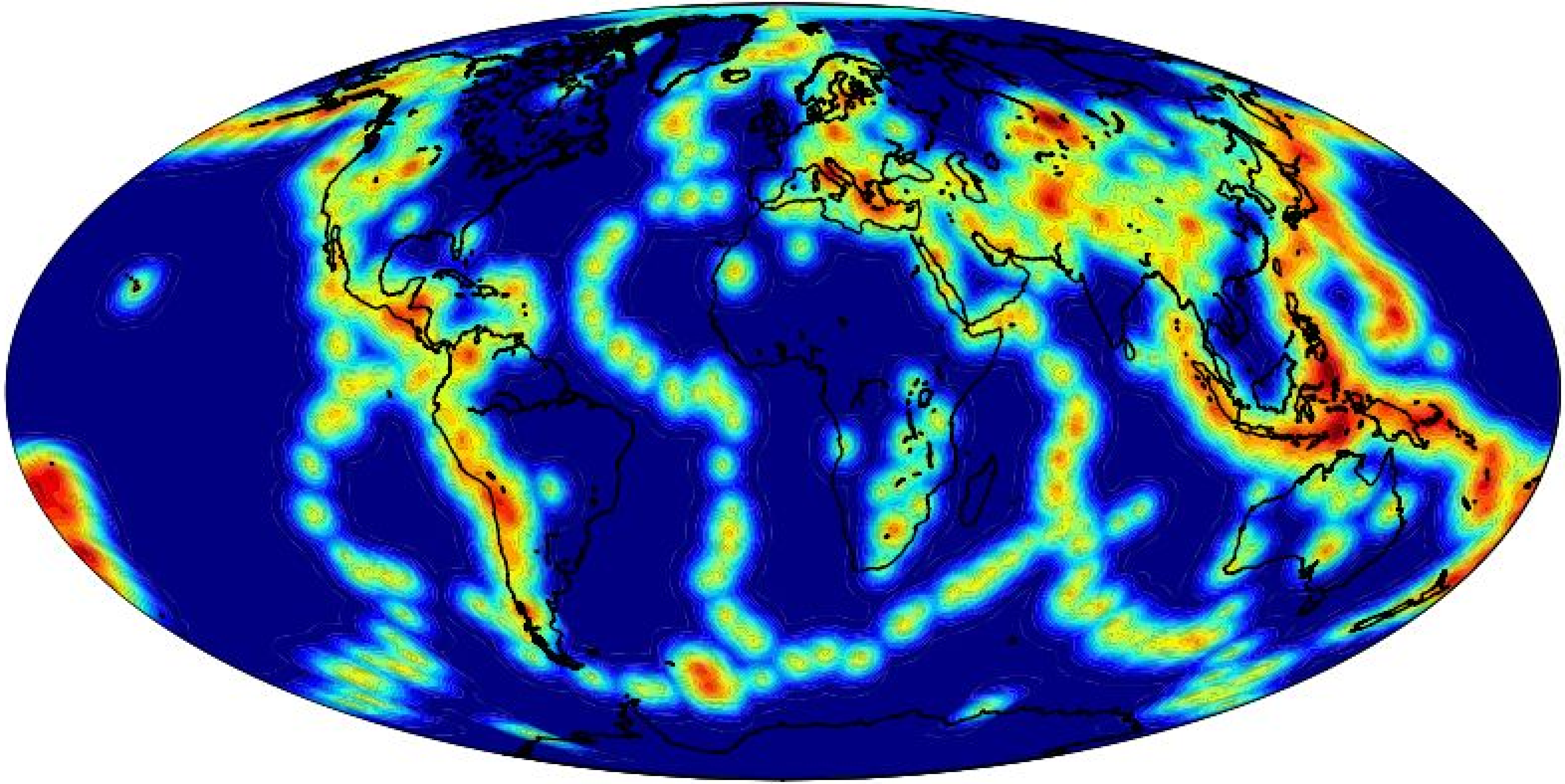
Conclusion - Bayesian Approach

- Quantify rational decision making using all available data and all applicable knowledge
- Other benefits
  - Ability to combine heterogeneous types of data (infrasound and hydroacoustic with seismic)
  - Can handle large number of detections (eliminating false detections is not relevant)
  - Can handle incorrect phase classification
- SIGVISA shows that waveform correlation can also be handled with a generative model very effectively

Appendix

Unassociated Arrivals per hour





Log Prior Density of Events

The effect of the seismic prior

- Allows us to form events with weaker evidence in seismically active regions
- Doesn’t affect the ability to form events in the rest of the globe

What if we didn’t form these low magnitude events in seismically active regions

- Then we have a lot of unexplained detections which lead to more false events being formed
- ***No distinction can be made between weak events in aseismic regions versus weak events in aseismic regions whose detections could have been better explained with a weak event in a seismic region.***