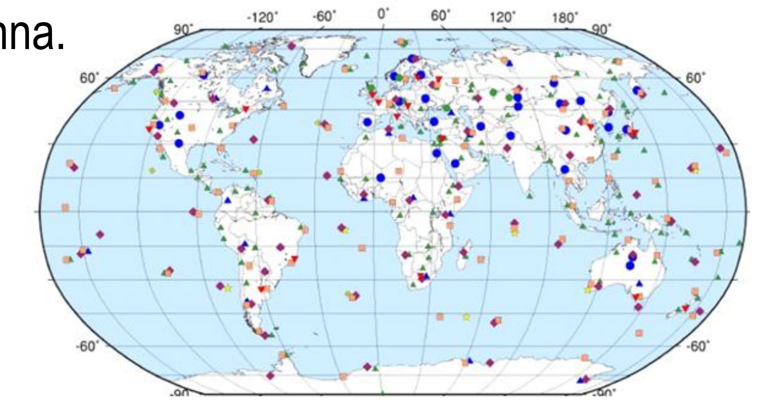
Scalable Probabilistic Inference for Global Seismic Monitoring (S43B-2238)

Nimar S. Arora, Tony Dear, and Stuart Russell
nimar@cs.berkeley.edu, tonydear@berkeley.edu, and russell@cs.berkeley.edu

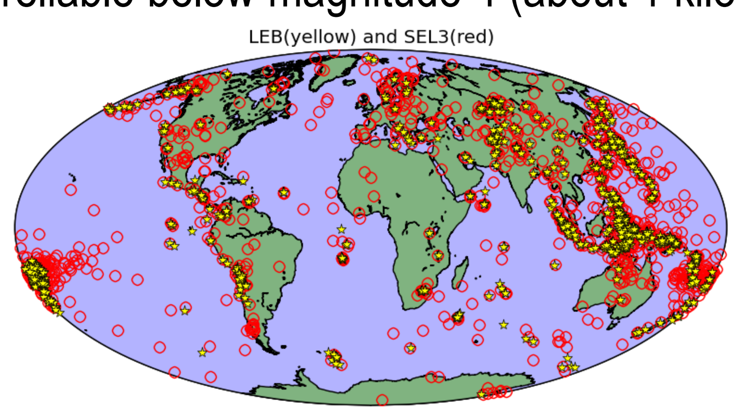
Introduction

- The **Comprehensive Nuclear-Test-Ban Treaty (CTBT)** bans all nuclear explosions on Earth, regardless of purpose.
- Global seismic monitoring aims to recover all seismic events in the magnitude range of interest, given a set of detections.
- Data from the International Monitoring System (IMS) are processed in real time at the International Data Centre (IDC) in Vienna.



Blue dots and triangles are primary seismic stations.

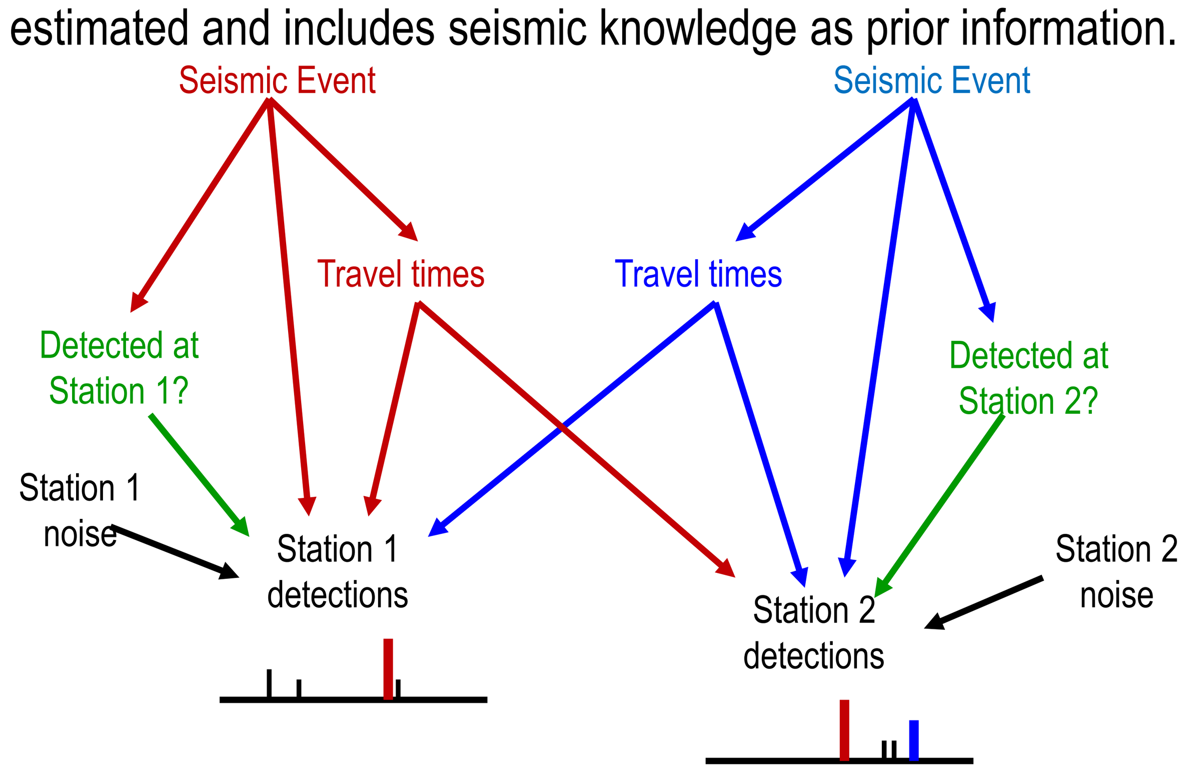
- The current automated system (**SEL3**) detects 69% of real events and creates twice as many spurious events.
- 16 human analysts find more events, correct existing ones, throw out spurious events, and generate **LEB** ("ground truth").
- Unreliable below magnitude 4 (about 1 kiloton).



Yellow stars – LEB, Red circles – SEL3. Results for 1 week.

The NET-VISA Model

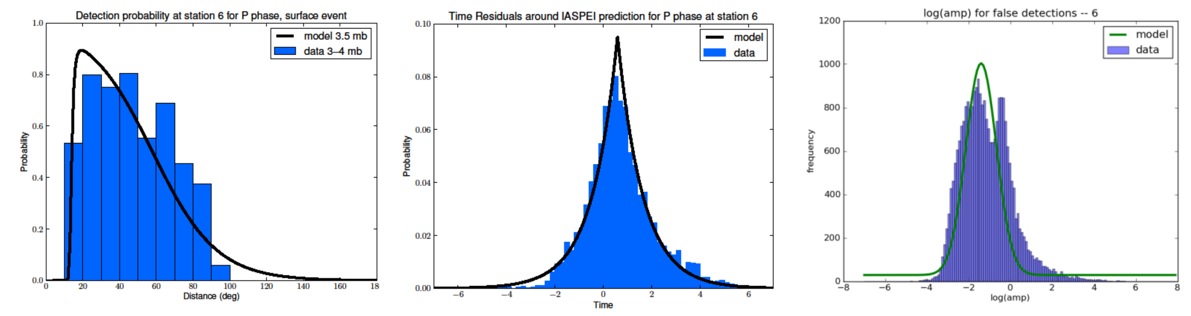
- Unlike SEL3, which processes the data in stages, we propose a single vertically integrated probability model that is empirically estimated and includes seismic knowledge as prior information.



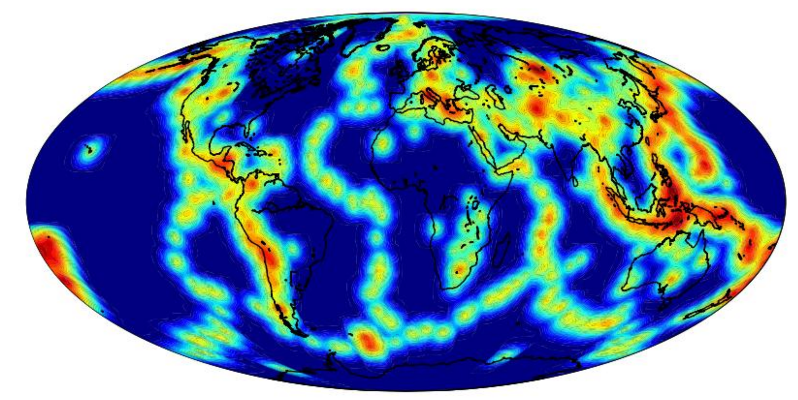
- Event magnitudes are distributed as per the Guttenberg Richter distribution (exponential distribution with rate log(10)).
- Event detection probabilities depend on the station, the seismic wave type (phase), event magnitude, and distance from the event to the station.

The NET-VISA Model (Continued)

- Event parameters – arrival time, azimuth, amplitude, etc. – have station-specific distributions.



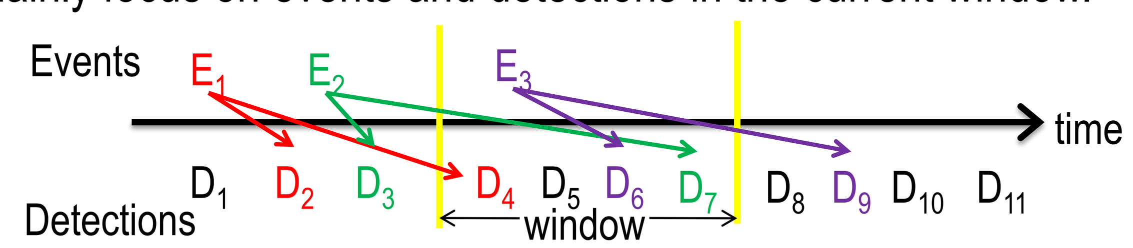
- Events are generated by a time-homogenous Poisson process.
- Earthquakes are located according to a kernel density estimate, while explosions have a uniform prior.



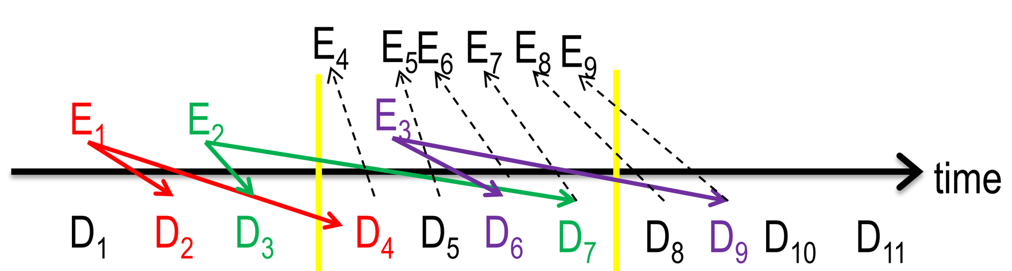
- The residual distribution for the travel time, azimuth, and slowness are mostly modeled as Laplacian distributions.
- Noise detections are generated by a station-specific time-homogenous Poisson process.
- All parameters are estimated from historical training data.

Performing Inference

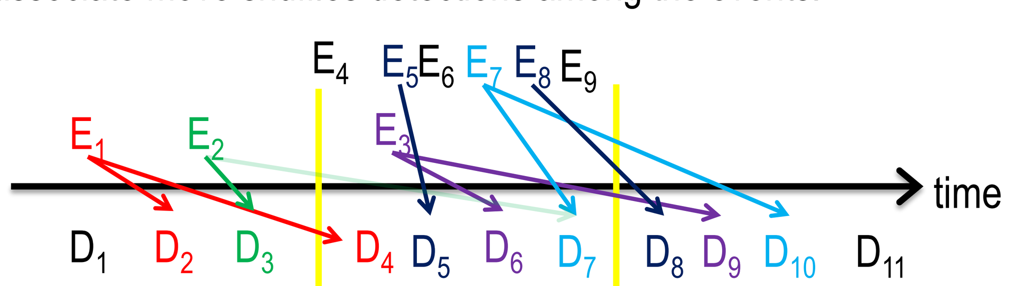
- Given the set of detections at all the stations, we need to infer **the most probable explanation (MPE)** – a sequence of events and the association of events to detections.
- Our algorithm goes through a sequence of moves which mainly focus on events and detections in the current window.



- The birth move adds new events by probabilistically "inverting" detections.

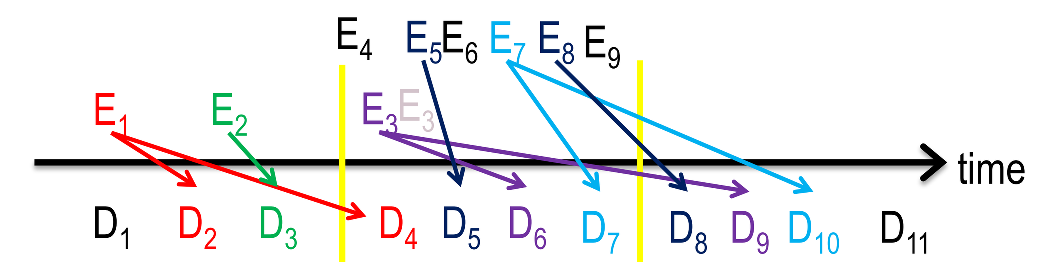


- The re-associate move shuffles detections among the events.

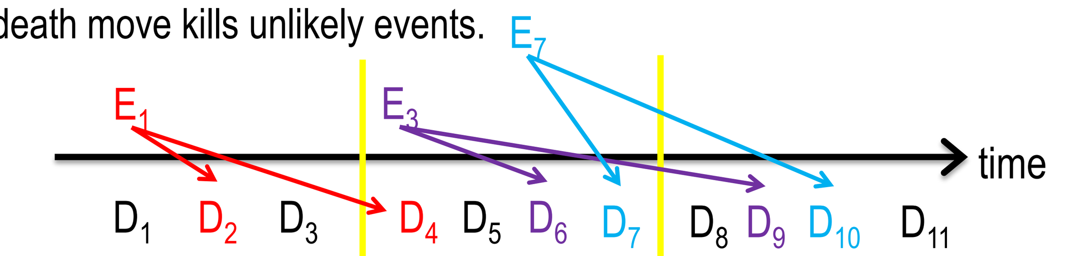


- The relocate move changes event locations.

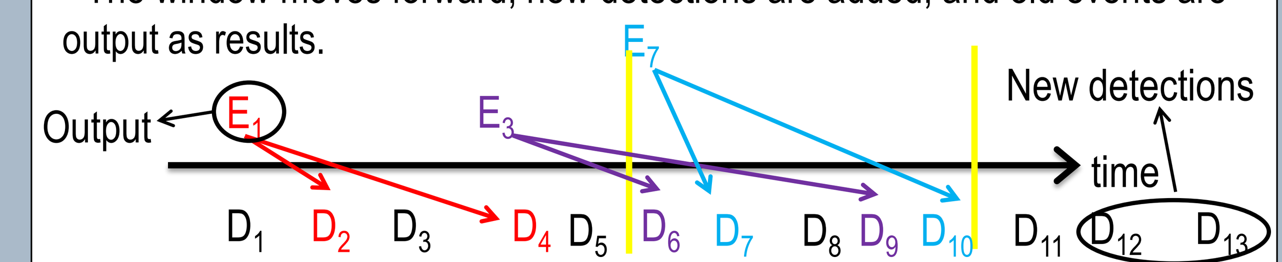
Performing Inference (Continued)



- The death move kills unlikely events.



- The window moves forward, new detections are added, and old events are output as results.



Parallel and Distributed Computing

- In actual implementation of the algorithm, we massively split up the computation using both multiple cores and multiple machines.
- We first split all detections among multiple threads and propose events independently of each other.
- During event relocation, each thread randomly perturbs an event independently of each other, each for a fraction of the original number of iterations. At the end, we simply keep the best one.
- We further parallelize the proposal phase by splitting detections among multiple machines in a computer cluster on Amazon's EC2.
  - We use MIT's StarCluster software to manage such networks.

Results

- We train on 2.5 months of data and evaluate on one week.
- The predicted events are matched to a ground truth event from the LEB within 5 degrees and 50 seconds.
- Recall: % of ground truth events that matched a predicted event.
- Error: Average distance between predicted matched events.

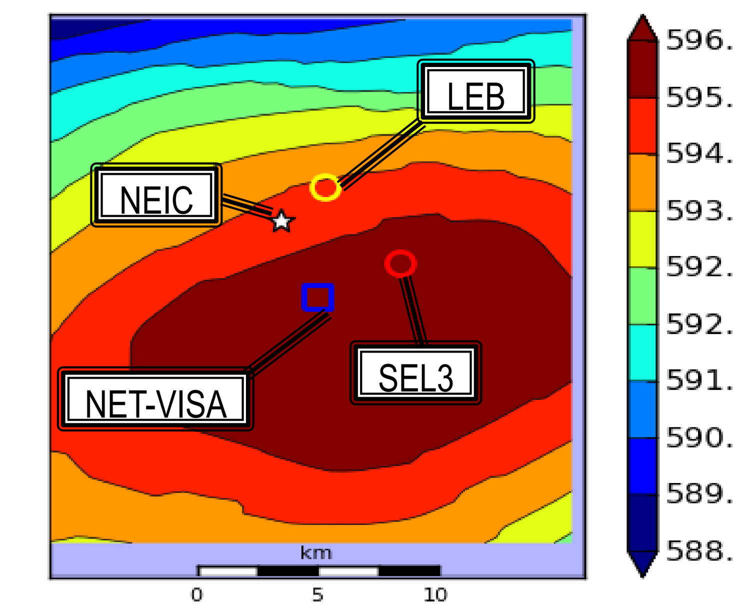
m_b range	# of Events	SEL3		NET-VISA			
		Rec (%)	Err (km)	8 cores CPU-hours: 180 hrs Elapsed: 22.5 hrs		192 cores CPU-hours: 214 hrs Elapsed: 3.65 hrs	
				Rec (%)	Err (km)	Rec (%)	Err (km)
0-2	74	64.9	101	86.5	79	86.5	90
2-3	36	50.0	186	80.6	153	83.3	155
3-4	558	66.5	104	87.1	120	86.7	118
>4	164	86.6	70	92.7	74	92.7	69
Avg	832	69.7	99	88.0	108	87.7	107

- NET-VISA significantly outperforms SEL3, reducing missed events from 30.3% down to about 12%.
- Parallel computing leads to speedup with no loss of performance.

Results (Continued)

Evaluation on DPRK Nuclear Explosion (25 May 2009)

- NET-VISA also has no difficulty locating nuclear events.
- Our prediction agrees well with NEIC and LEB (roughly 5 km).
- NET-VISA associated the event with detections from 53 stations, versus 39 stations by SEL3.



Dealing with Large Aftershock Sequences

- Large events, such as the Dec. 2004 Sumatra earthquake and the March 2011 Tohoku earthquake, can overwhelm processors.
- IDC bulletins are often delayed by weeks or months.
- We take full advantage of parallel processing for speedup.
- Results are shown for experiments on 72 hours of data from the Sumatra event and 48 hours from the Tohoku event.

Event	Cores	CPU-hours	Elapsed (hrs)	SEL3		NET-VISA	
				Recall (%)	Error (km)	Recall (%)	Error (km)
Sumatra	1	645	645	70.8	124	81.9	119
	8	686	85.7			81.9	113
	192	859	23.4			81.3	114
Tohoku	8	2010	251	71.7	119	78.9	108
	192	2470	38			78.4	101

- As with the validation results, using multiple cores has resulted in much more reasonable computation time.

Conclusions and Further Work

- NET-VISA reduces detection failures by more than a factor of 2 relative to SEL3.
- NET-VISA is currently being tested in the CTBTO environment for possible deployment in operations.
- We are now developing the SIG-VISA prototype with an extension of the generative model down to the waveform level, including the step of signal detection in the model.

Acknowledgements

- This project has been funded by CTBTO and Amazon EC2.