Application of empirical signal detectors in future global pipelines

CASIS 2016

Doug Dodge
Dave Harris

May 18, 2016

LLNL-PRES-685865
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

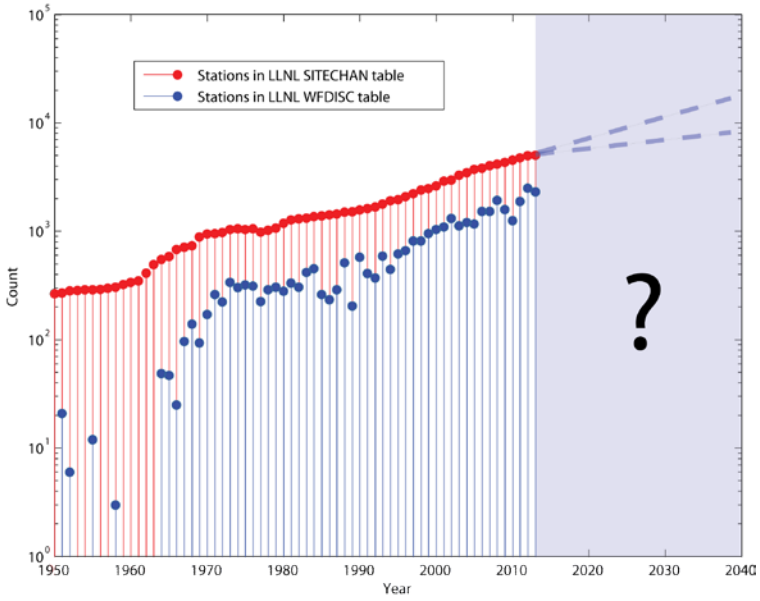
Lawrence Livermore National Laboratory

Lowering monitoring thresholds and increasing channel counts requires a paradigm change.

Current practice is predicated on hundreds of events / day observed on hundreds of channels.

Requires event review by many highly trained human analysts

But sensor counts are growing at a super-linear rate.



The increase in channels coupled with decreased thresholds means perhaps 1000X more signals to analyze.

It won't be possible to have "eyes on" every signal, so automated processing that utilizes the full signal bandwidth to reliably classify signals will be required.

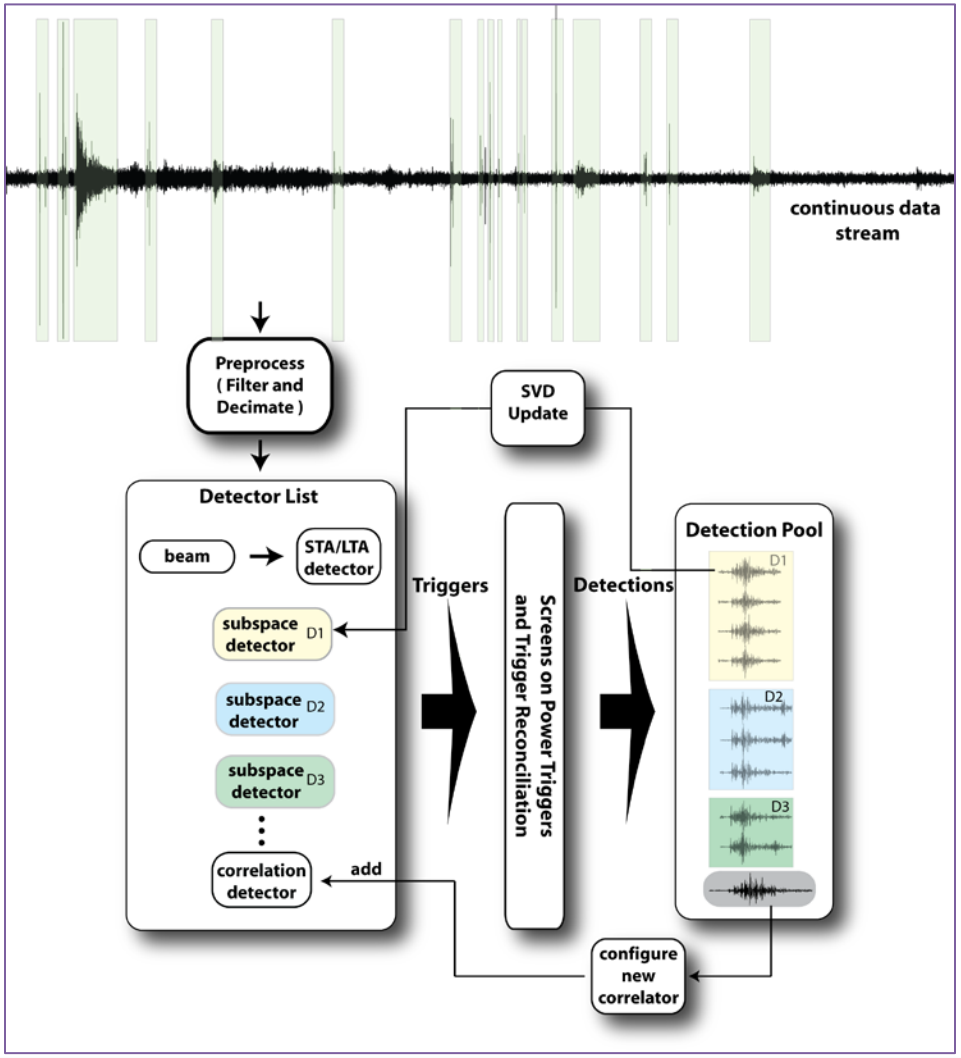
We think large-scale deployments of empirical signal detectors will be an important component of such processing.

Dynamic Correlation Processor (DCP)

- Initial implementation was for automated aftershock processing.
- Discovers its own templates including many signals not seen elsewhere
- Ideal for characterizing the signal space observed by a single seismic stream

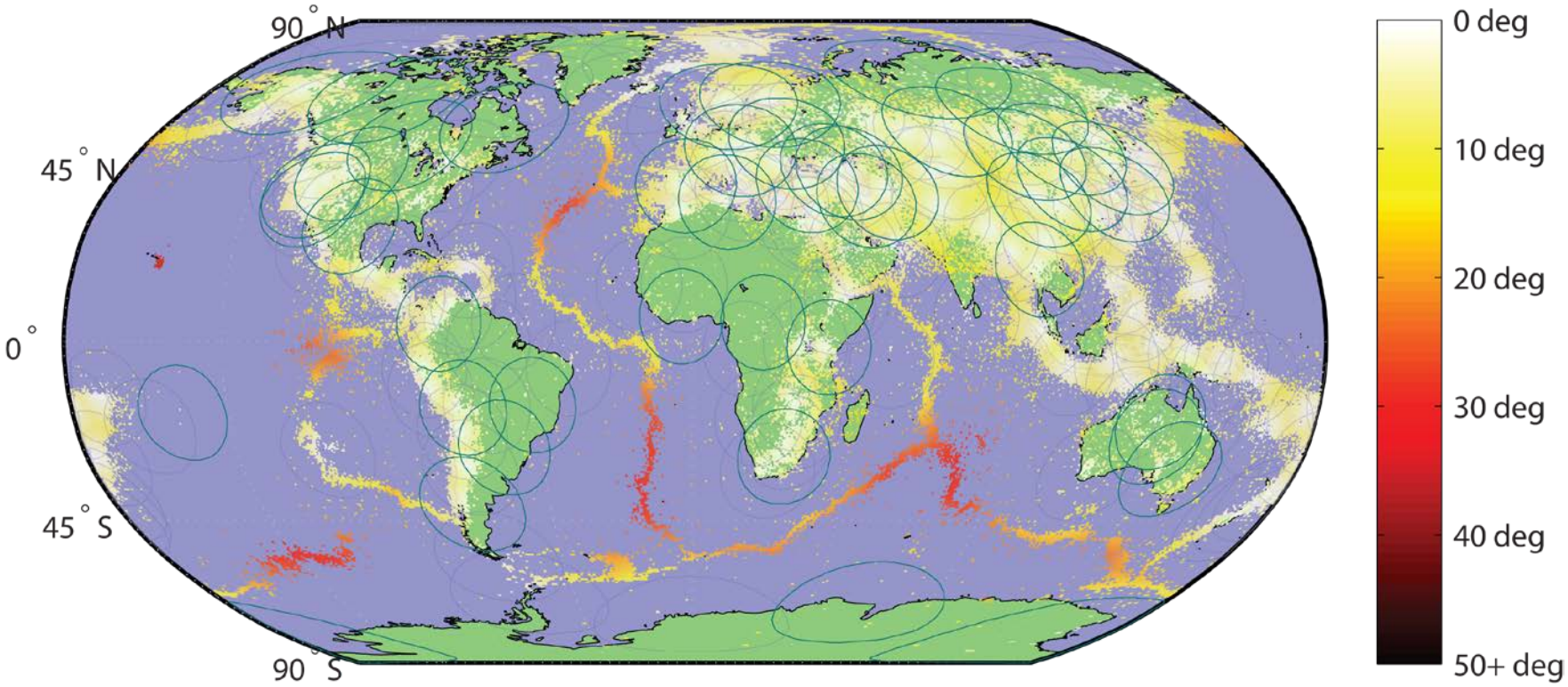
Dynamic Correlation Processor (DCP)

- Processes a stream of 1-N channels in a single pass band.
- Maintains a pool of detectors whose templates span every signal ever seen.
- Growth of the pool is limited only by available memory.



Large scale test of dynamic correlation processors

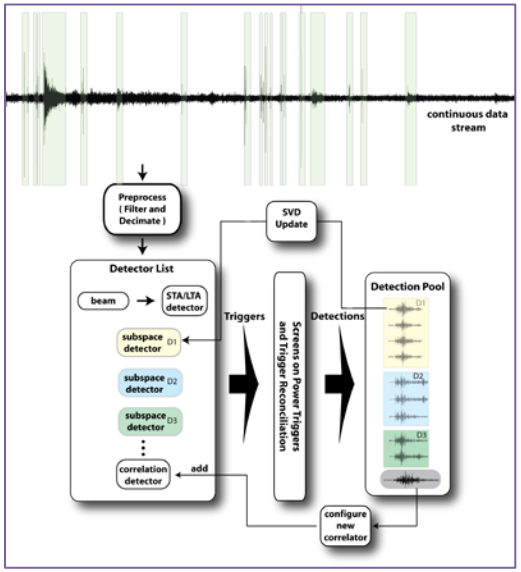
Nearly all continental seismicity is within ~10 of at least one IMS station.
How many of these events could be detected by correlation?



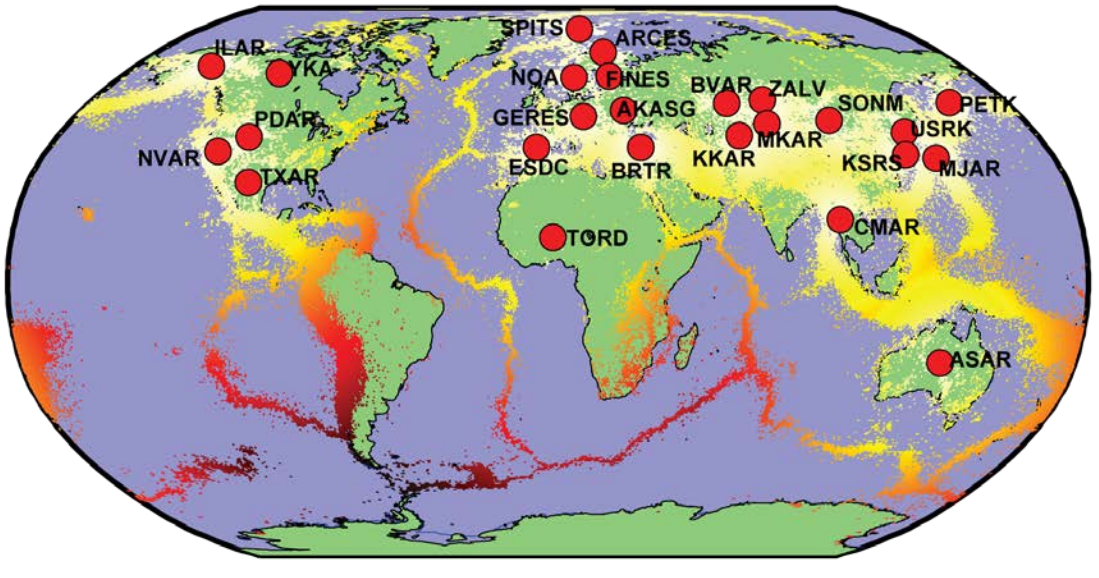
Dodge and Harris, 2016 (BSSA)

Experiment configuration

- 11 terabytes of array data (25 arrays 4,400 channel-years).
- Average duration was ~16 years.
- Data were processed on a 9-node, large-memory SMP cluster.
- Each DCP used 12 beam-forming power detectors as spawners.
- Triggers were screened by SNR, duration, and apparent velocity.



Distance to nearest array



Results

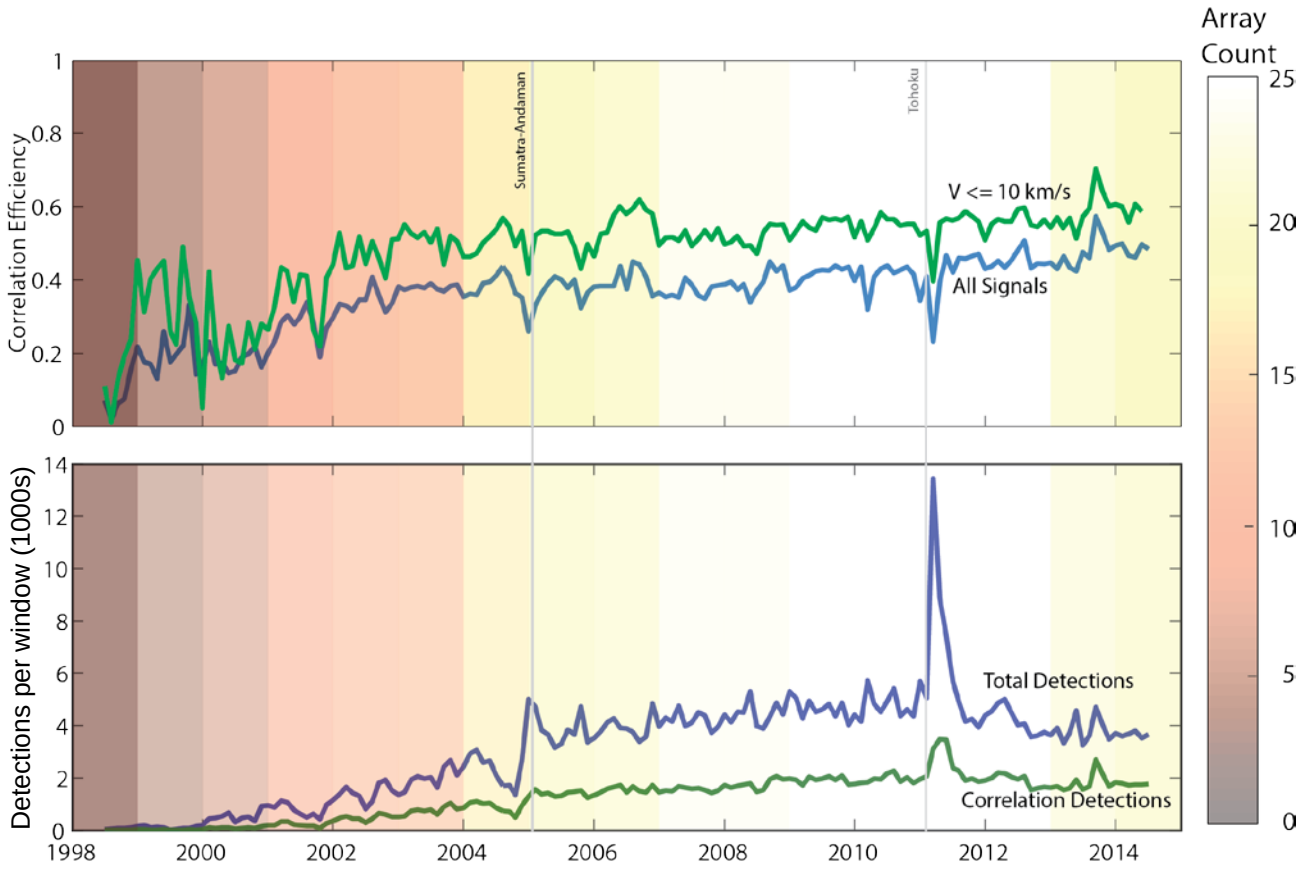
- Processing required ~2 months.
- Over 360,000 detectors were created that collectively triggered on about 3.2 million unique signals.
- Detector counts ranged from 2390 at AKASG to 41,726 at MKAR.
- After classification, ~290,000 detectors with ~460,000 detections were retained.
- ~46,000 of those detectors produced more than 1 detection (~181,000 correlation detections).
- Over ~16 years of operation nearly 40% of detections are from correlators.

Correlation efficiency improved over time and was higher for local-regional signals.

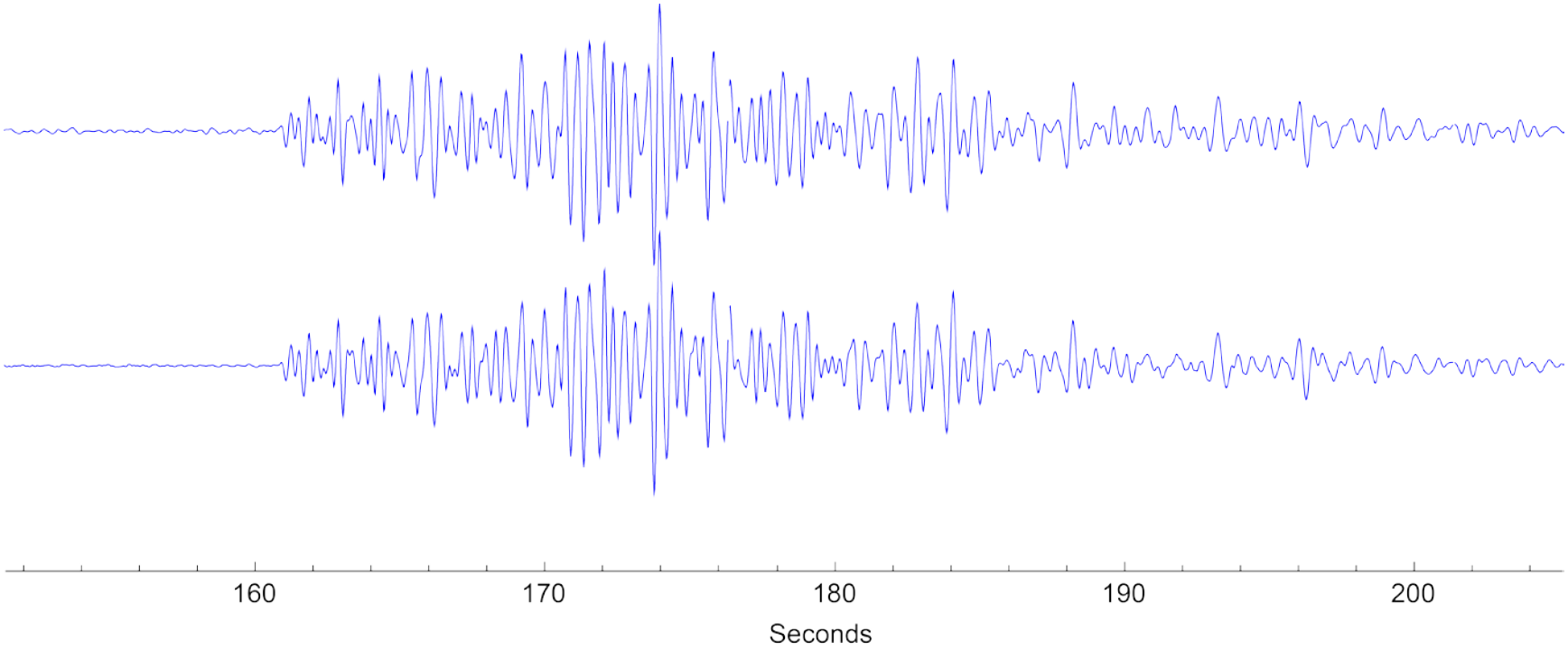
The correlation efficiency has a linear trend of ~0.01 per year
(fit to data from 2004–2014).

In that interval efficiency increases from ~.35 to ~.46

For local-regional signals the efficiency is higher by about 0.1.

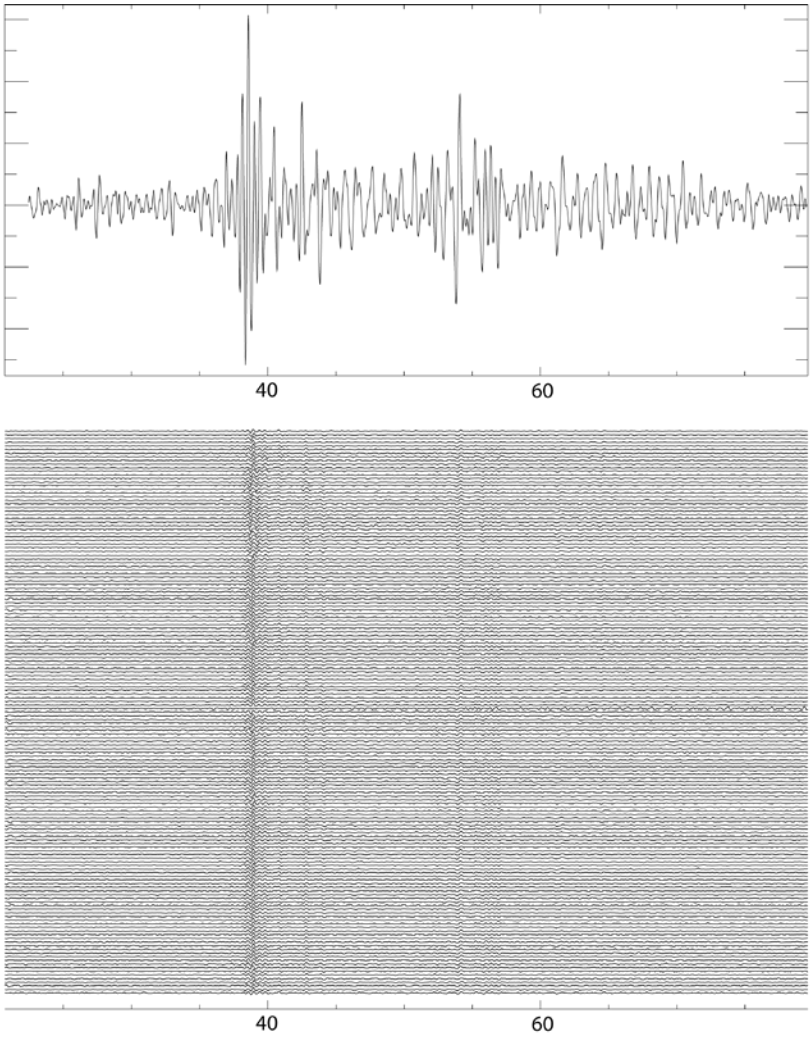


The high efficiency is mostly from the tens of thousands of detectors with only 1 or 2 repeats.



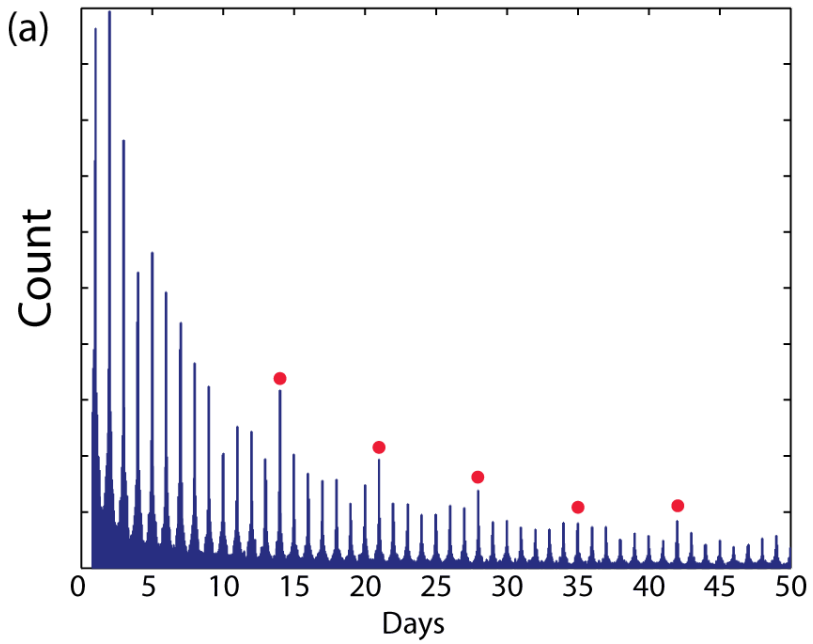
But a small number of patterns are highly repetitive.

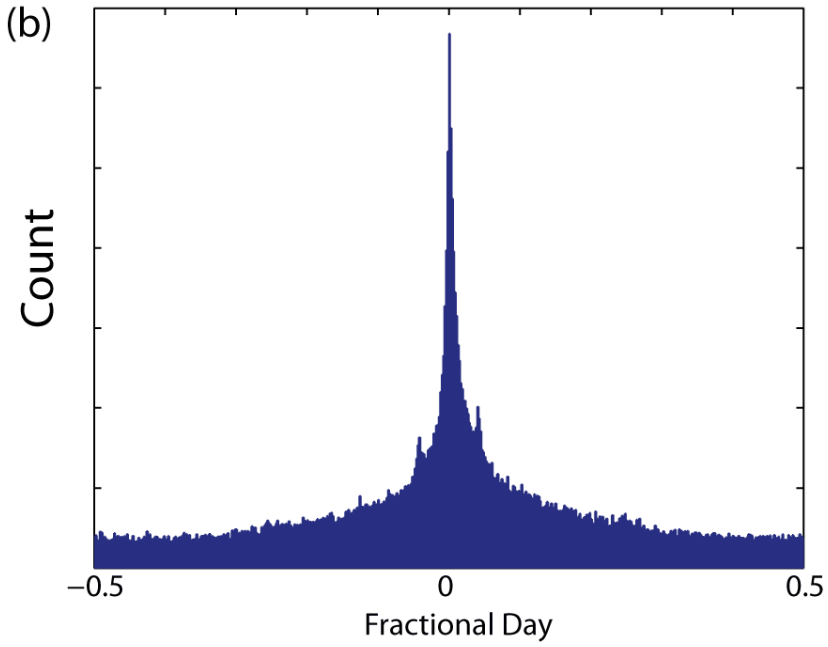
- 225 detectors had over 100 repeats,
- 24 had over 500 repeats,
- 7 had over 1000 repeats,
- And one detector had over 9000 repeats.



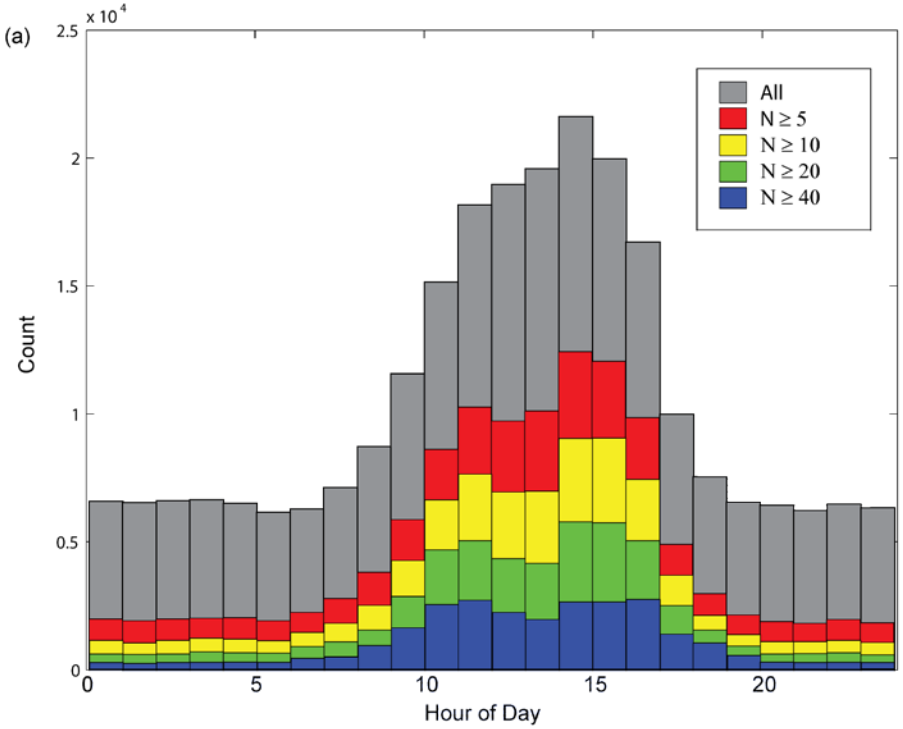
Recurrence intervals suggest many patterns are industrial in origin.

- A significant fraction of repeaters have 24-hour and 1-week recurrence intervals.

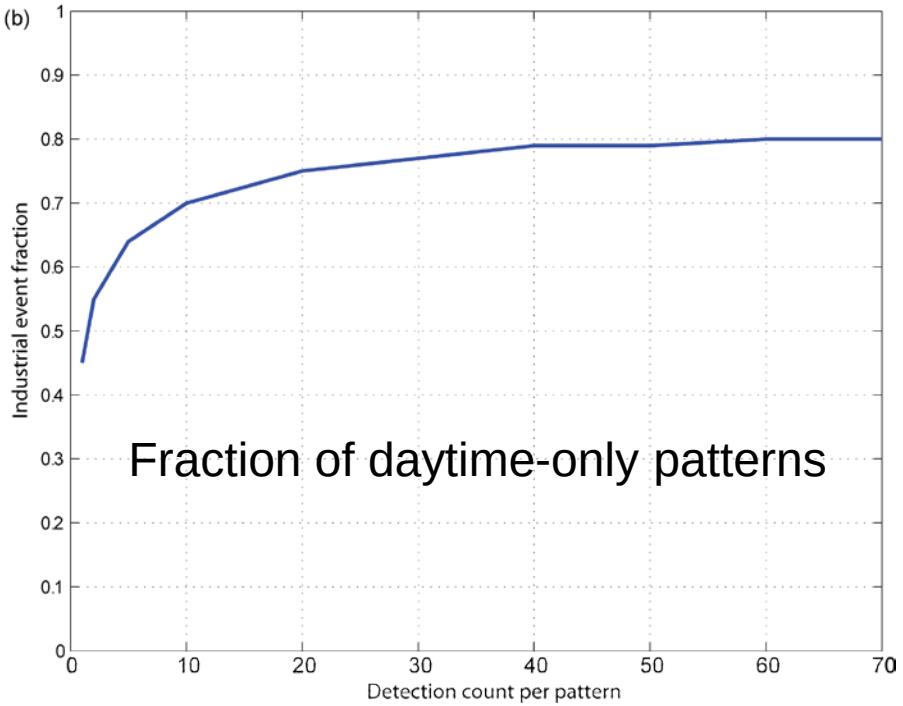




Time of day also suggests an industrial origin.



80% of large multiplets are daytime-only



Applying dynamic correlation processors to a seismic pipeline

From the monitoring perspective most signals are nuisances.

NDCs are tasked with “finding a needle in a haystack” and a disproportionate part of that effort involves manual review of uninteresting signals.

As station counts grow and thresholds are lowered, the size of the “haystack” of uninteresting signals will grow faster than current architectures can be scaled up.

Our results so far suggest that many local-regional signals can be detected by correlation, and if so, that could provide the phenomenological basis for correlation filtering as a component of pipeline processing. **But, a number of technical issues must be addressed to operationalize these capabilities.**

Scaling out

Extending DCP filtering to the full International Monitoring System seismic network would require (currently) 4 terabytes of templates per band with a growth rate of ~0.3 terabytes per year. We are investigating two ways to manage the volume:

- In-memory data grids
- Tiered storage with fast lookup (e.g. Locality-Sensitive Hash lookups or subspace template hierarchies) for flash-hosted templates.

We anticipate that our efforts will be aided by the rapid evolution of data-intensive compute platforms.

Learning the signal space

- Although Dynamic Correlation Processors perform a de facto clustering of the signal space they cannot characterize either the signal source or structure.
- We anticipate that pipeline output will be used to learn classifiers for a subset of signals, but un- or semi-supervised learners will be required for much of the signal characterization work.
- Although we have had success screening signals using classifiers trained on time-domain features, we think robust characterization will be better-achieved using time-frequency representations. We are currently pursuing this approach.

Lawrence Livermore
National Laboratory