The Iterative Processing Framework: a new paradigm for automatic event building

Rigobert Tibi[1], Andre Encarnacao[1], Sanford Ballard[1], Christopher Young[1], and Ronald Brogan[2] (*Email: rtibi@sandia.gov*)

[1]) *Sandia National Laboratories;* [2]) *ENSCO, Inc., Falls Church, VA*

Sandia National Laboratories

SL18-Signal Analysis-NDD2Ad

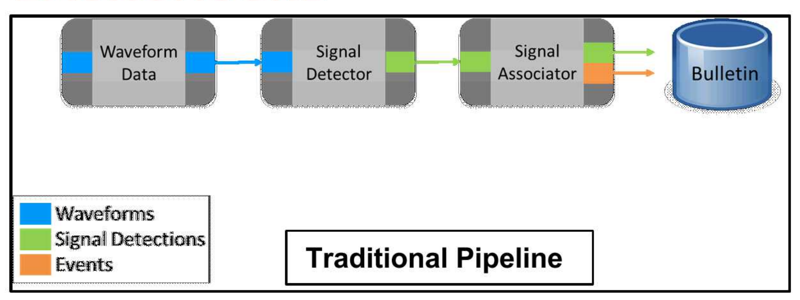
National Nuclear Security Administration

SAND2019-2387C

OVERVIEW

In a traditional data processing pipeline, signal detections (i.e., arrival times, slownesses, and azimuths) are passed to a signal associator to form events. The associator then links the detections to the fitting event hypotheses to generate an event bulletin. Most of the time, this traditional pipeline requires heavy human analyst involvement to improve the quality of the resulting event bulletin. For example, at the International Data Center (IDC) human analysts spend a significant amount of time and effort to correct the automated bulletin. We propose an Iterative Processing Framework (IPF) that includes a new data processing module that incorporates automated analyst behaviors (Auto Analyst) into the event building pipeline. In the proposed framework, through an iterative process, Auto Analyst takes over many of the tasks traditionally performed by human analysts.

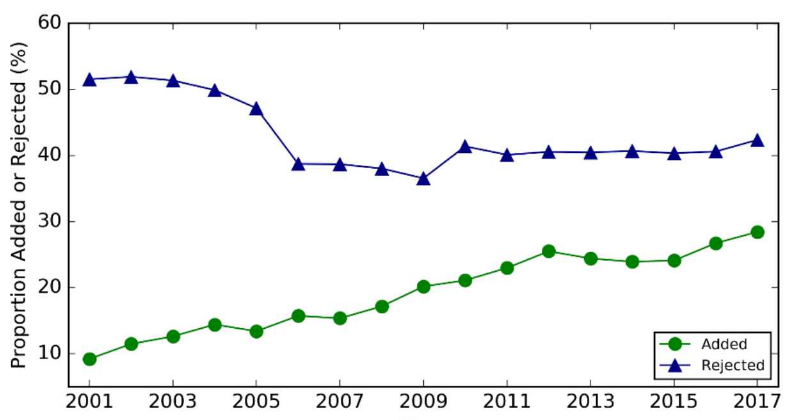
BACKGROUND



In a traditional pipeline of event building,

1) Waveform data are processed by a signal detector,
2) The signal detections are passed to a signal associator, and
3) The associator links the detections to the fitting event hypotheses to generate an event bulletin.

The traditional pipeline requires heavy human analyst involvement to improve the quality of the resulting event bulletin.

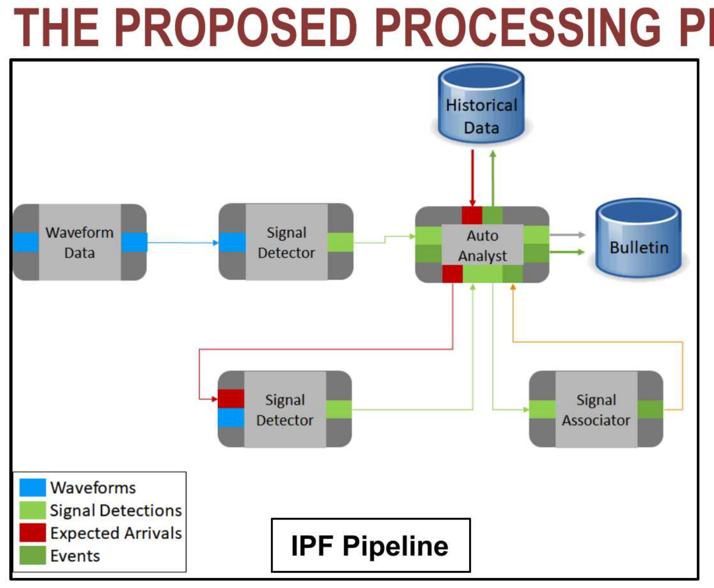


At the International Data Center (IDC):

- About 43% of the yearly events included in the automatic bulletin (SEL3) are rejected by human analysts as false.
- The proportion of legitimate events added by human analysts to the Late Event Bulletin (LEB) has steadily increased from ~10% in 2001 to ~30% in 2017 (see figure on the left), requiring increasingly significant human effort.
- The workload on human analysts will continue to increase as additional IMS stations come online.

To reduce the human-analyst workload during the seismic event building process, we are proposing an Iterative Processing Framework (IPF) that includes a new data processing module, which incorporates automatic analyst behaviors into the event building pipeline.

THE PROPOSED PROCESSING PIPELINE



The proposed pipeline, IPF, includes:

- All the components of a traditional pipeline, and
- A new module that incorporates automatic analyst behaviors (Auto Analyst, AA) and provides the possibility to query historical data for empirical information.

AA accomplishes the following tasks, grouped into **2 processes**:

1) Evaluate small (low *NDEF*) events to improve their formation

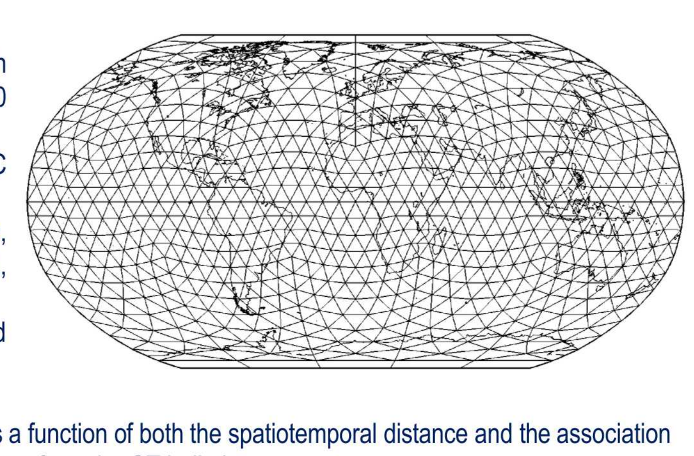
- Refine arrivals from arrays for low *NDEF* events by performing *f-k* analyses with varying parameters
- Rerun associator if any arrivals are modified in the process

2) Scan unassociated arrivals to build potential missed events

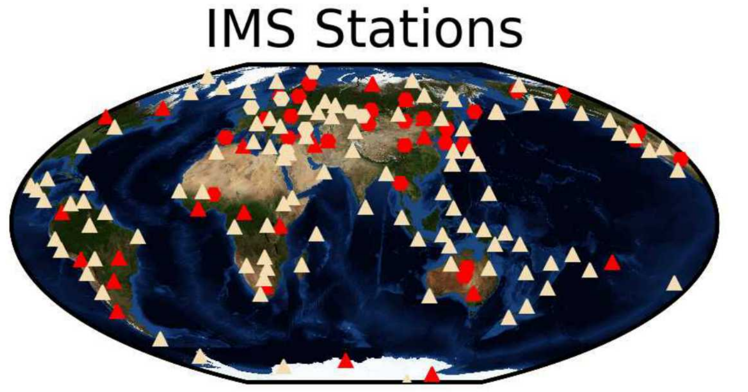
- Identify unassociated arrivals from array stations
- Compute single station locations
- Querry historical data for missing expected arrivals
- Reprocess waveforms searching for expected arrivals
- Rerun the associator with any new detected arrivals

TEST SETTINGS AND DATA

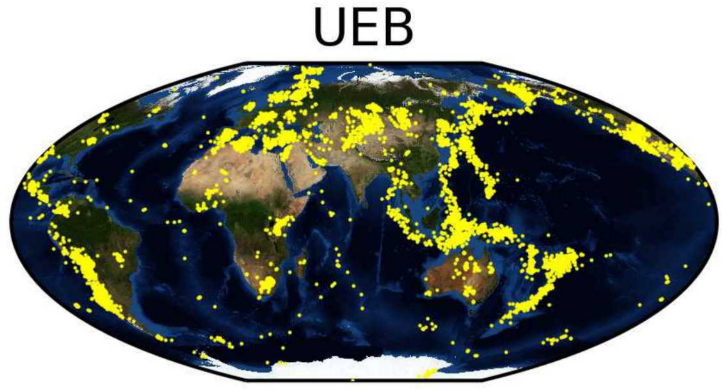
- The association algorithm used in IPF is PEDAL (Draelos *et al.*, 2015).
- The Earth is divided into 2°-grids uniformly spaced, with depth nodes in region where deep seismicity is expected to occur, resulting in ~13,000 nodes (see figure on the right).
- Test dataset consists of a 2-week period (1–14 May 2010) of the IDC signal detections.
- For the same time period, an expert-analyst compiled a more complete, high-resolution bulletin (Unconstrained Event Bulletin, UEB (Linville *et al.*, 2019)).
- UEB contains 11,378 events and includes all the events from the LEB and many legitimate non-LEB events (see figure below, right).
- UEB is the ground truth (GT) used to assess the quality of IPF bulletins.
- For bulletin comparison, we used an event commonality score (ECS) that is a function of both the spatiotemporal distance and the association difference between the events from the bulletin to be evaluated and the events from the GT bulletin.



IMS Stations



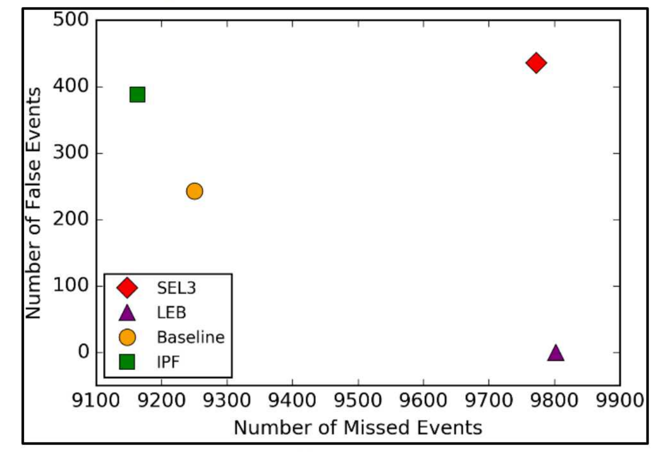
UEB

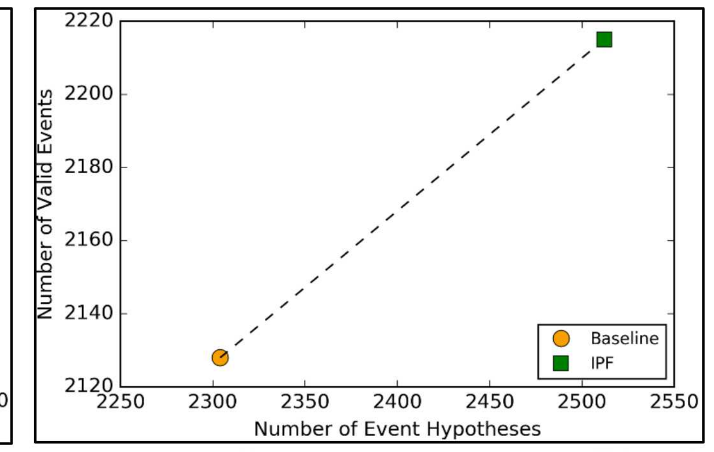


The authors acknowledge the support of the National Nuclear Security Administration Office of Defense Nuclear Nonproliferation Research and Development for funding this work. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions 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.

RESULTS

Bulletin Comparison

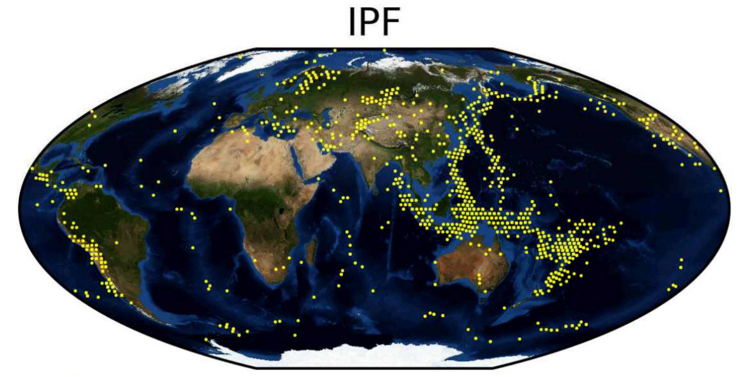




- SEL3 contains high numbers of false events and missed events, explaining why it is reviewed and augmented manually to produce the LEB (top left figure).
- LEB contains no false, but has the largest number of misses (9,802)
- Baseline (PEDAL) misses 522 fewer legitimate events and builds 193 fewer false events compared to SEL3 (GA, Global Associator (Le Bras *et al.*, 1994), the associator currently used at the IDC).

- IPF generates ~6% more legitimate events and ~11% fewer false compared to SEL3.
- IPF bulletin contains ~6% more legitimate events than LEB.
- IPF forms 87 more events than the baseline.
- Thanks to its single-station location capability, IPF creates many more event hypotheses, which leads to a significant improvement of the pipeline performance.
- On average, ~90% of event hypotheses result each in a valid event (top right figure)

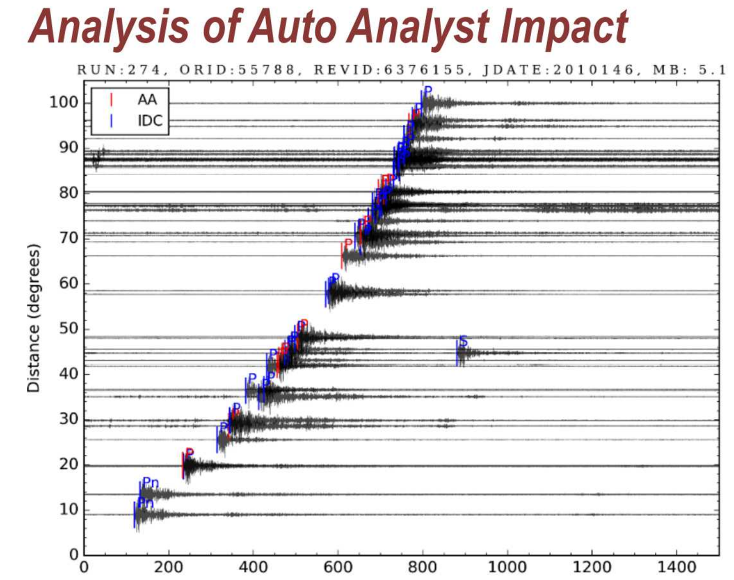
IPF

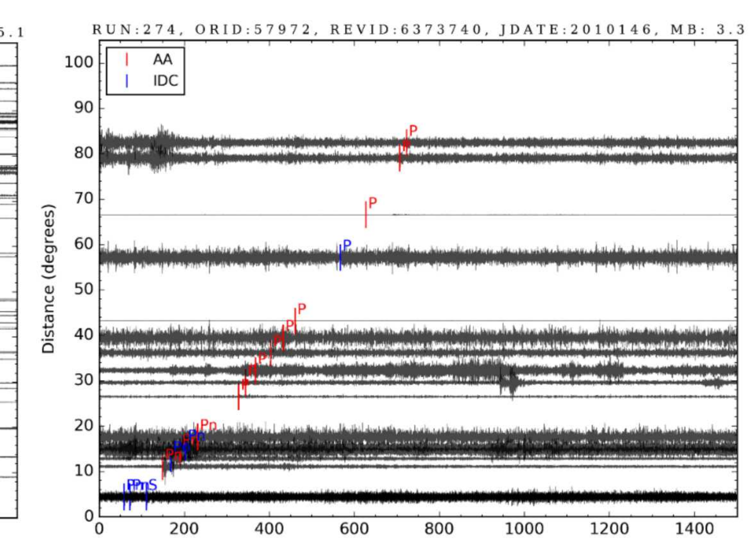


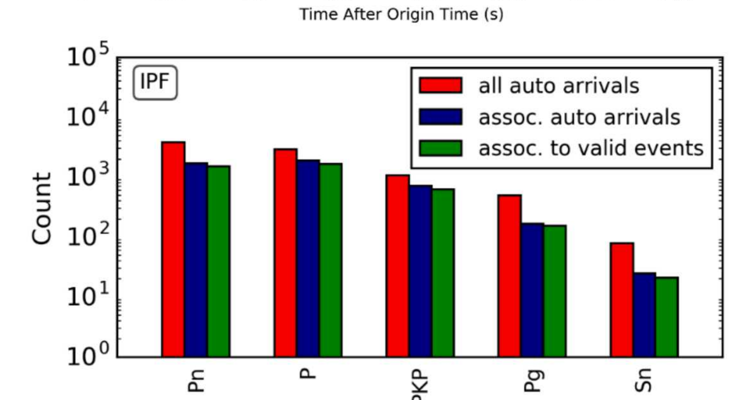
Best Matches

- The exact location of the events from the IPF pipeline are dictated by the grid size.
- Nevertheless, these locations are good enough to outline known seismic zones.
- For the most part, location patterns are consistent between IPF bulletin and the best matches from the GT.

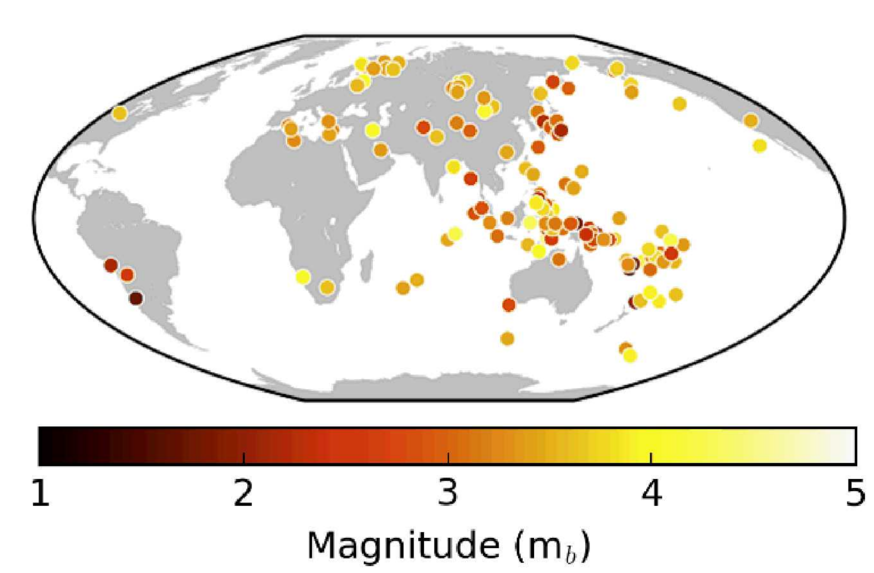
Analysis of Auto Analyst Impact

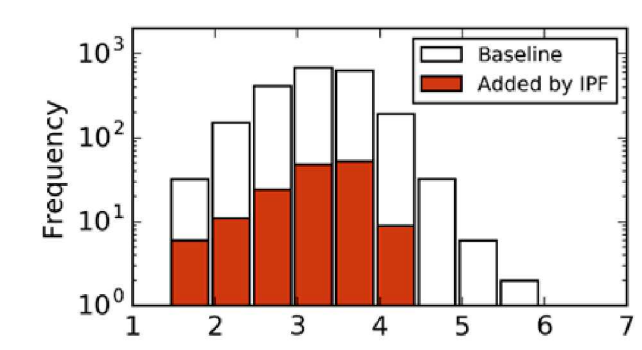




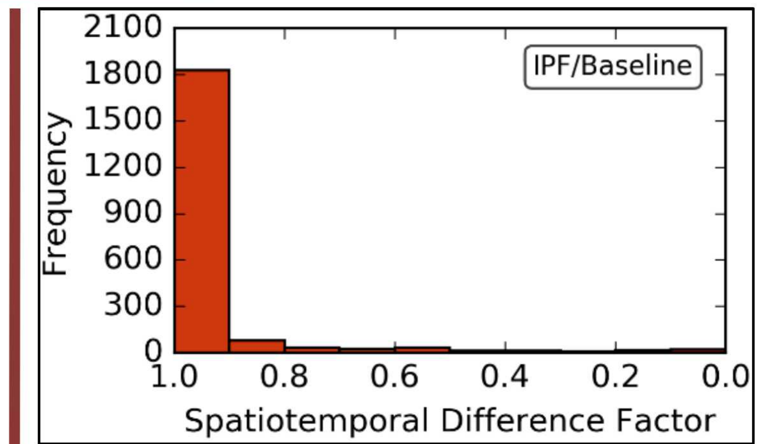


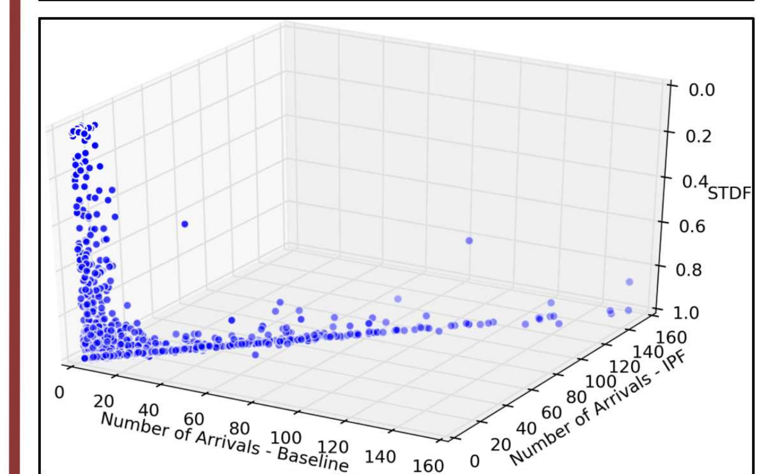
- For existing events, AA detections are consistent with IDC or expert-analyst's detections, suggesting that AA detections are generally correct.
- With its capability to detect at lower thresholds, in optimized detection bands that are informed by historical data, AA outperforms IDC detector in less optimal conditions (e.g., noisy data).
- About 54% of AA detections are associated to event hypotheses (valid or false), and ~48% are associated to valid events (see figure on the left).





- The AA built 150 additional events (not available in the baseline).
- Most of these additional events are low-magnitude events that were missed by the traditional pipeline of event building.



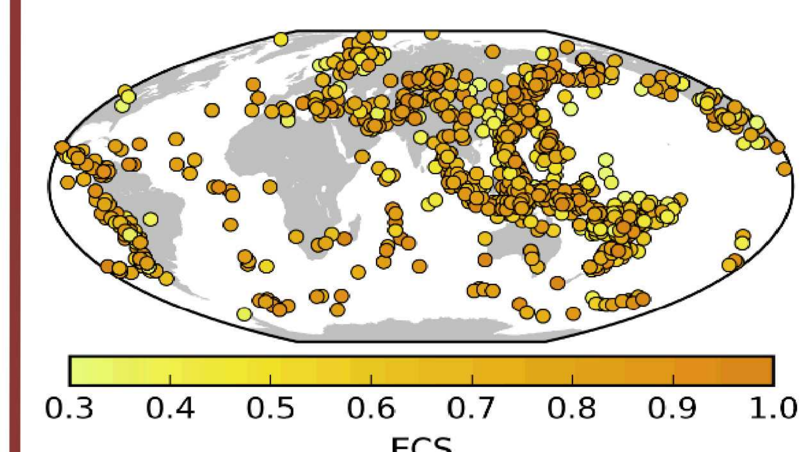


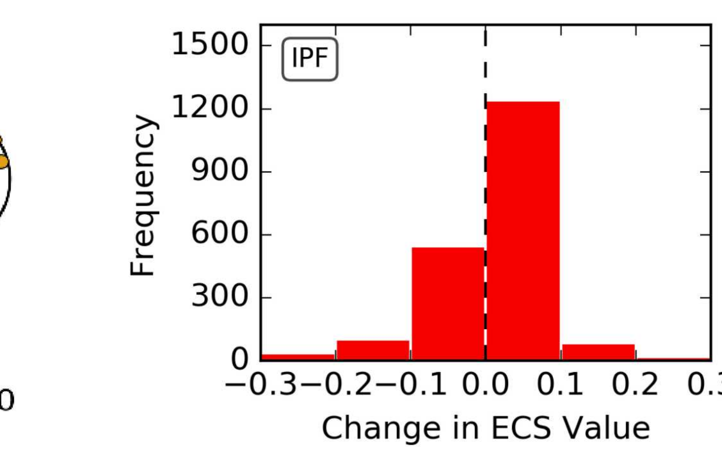
The spatiotemporal difference factor (STDF) between 2 common events from the IPF and baseline bulletins is defined as:

$$STDF = e^{-(d/A)^2},$$

where $d = D_{epi} + \Delta T_0 \times V_P$; D_{epi} is the distance between the epicenters, ΔT_0 the difference in adjusted origin times, V_P the *P*-wave velocity of 10 km/s, and *A* the distance scaling factor of 1500 km.

- A small STDF value indicates that the 2 events are distant in space and/or time.
- For common events between the baseline and IPF bulletin, most of the STDF values are between 0.9 and 1 (see figure on the top left), suggesting that the event locations in space and time are not significantly affected by the IPF processing.
- In general, only events with limited number of associated arrivals in the baseline bulletin are prone to significant changes in hypocentral location and/or origin time when associations are modified in the IPF pipeline.
- For most events in the baseline bulletin, the AA in the IPF pipeline provides additional arrivals, as implied from the observation that the majority of the data points in the figure on the left lie in the further half.

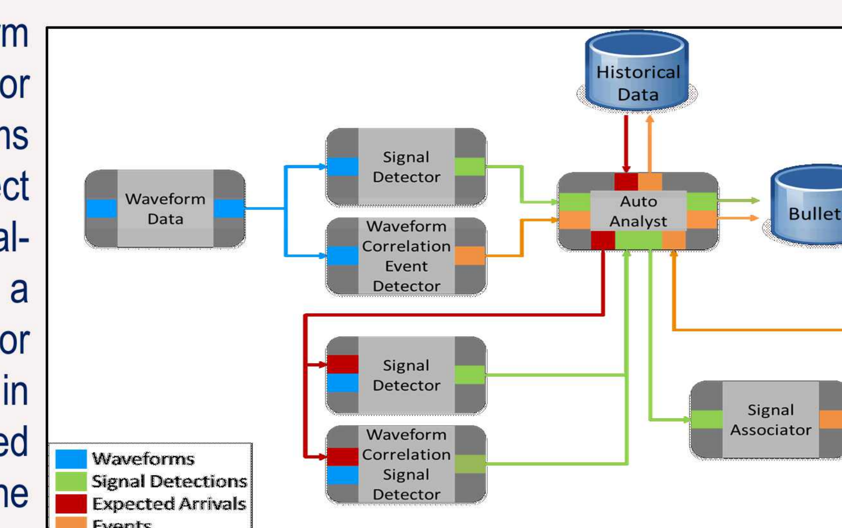




For most events that are common between the baseline and the IPF bulletin, IPF improves the ECS values. That is, IPF processing leads to better matches to events in the GT bulletin.

CONCLUSIONS AND OUTLOOK

- ❑ IPF performs better than traditional pipelines.
- ❑ Most of the additional events built by the AA are low-magnitude events that were missed by the traditional processing pipelines.
- ❑ The AA adds additional signal detections to existing events, which saves analyst time, even if the event locations are not significantly affected.
- ❑ On average, ~90% of event hypotheses result each in a valid event.
- ❑ For repeating events, waveform correlation is known to be superior to traditional detection algorithms because of its ability to detect signals with extremely low signal-to-noise ratios. Adding a waveform-correlation detector (WCD) to the pipeline, as shown in the figure on the right, is expected to dramatically increase the number of event hypotheses during the event building process, which we expect to improve the performance of the pipeline.



- ❑ The impact of a WCD on the pipeline performance will be the focus of a future study.

REFERENCES

- Le Bras, R., H. Swanger, T. Sereno, G. Beall, R. Jenkins, W. Nagy, and A. Henson (1994). Global association final report, *SAIC Technical Report SAIC-94/1155*.
- Draelos, T. J., S. Ballard, C. J. Young, and R. Brogan (2015). A new method for producing automated seismic bulletins: Probabilistic event detection, association, and location, *Bull. Seismol. Soc. Am.*, **105**, no. 5, 2453–2467.
- Linville, L., R. Brogan, C. Young, and K. A. Aur (2019). Global to local high-resolution event catalogs for algorithm testing and source studies. Submitted to *Seismol. Res. Lett.*