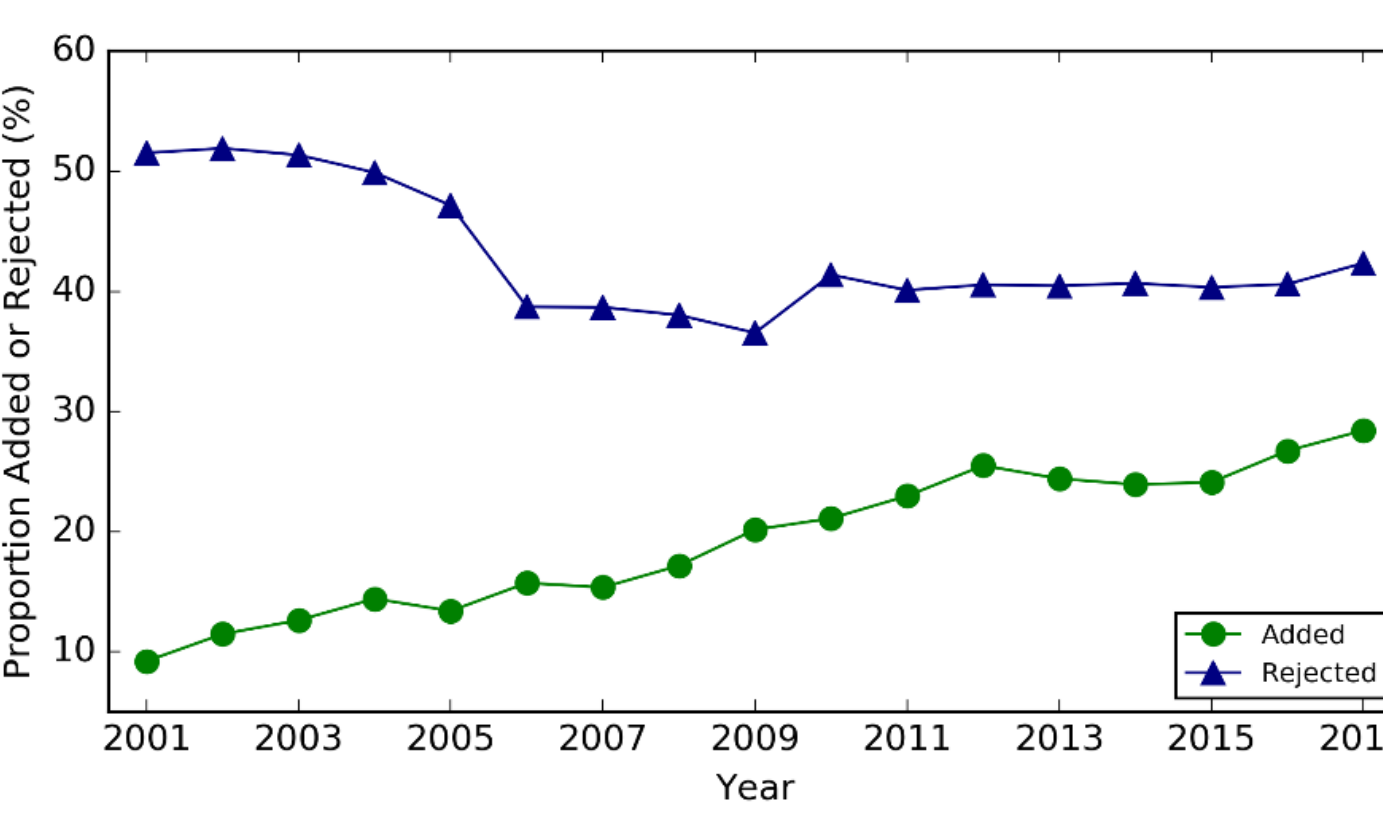
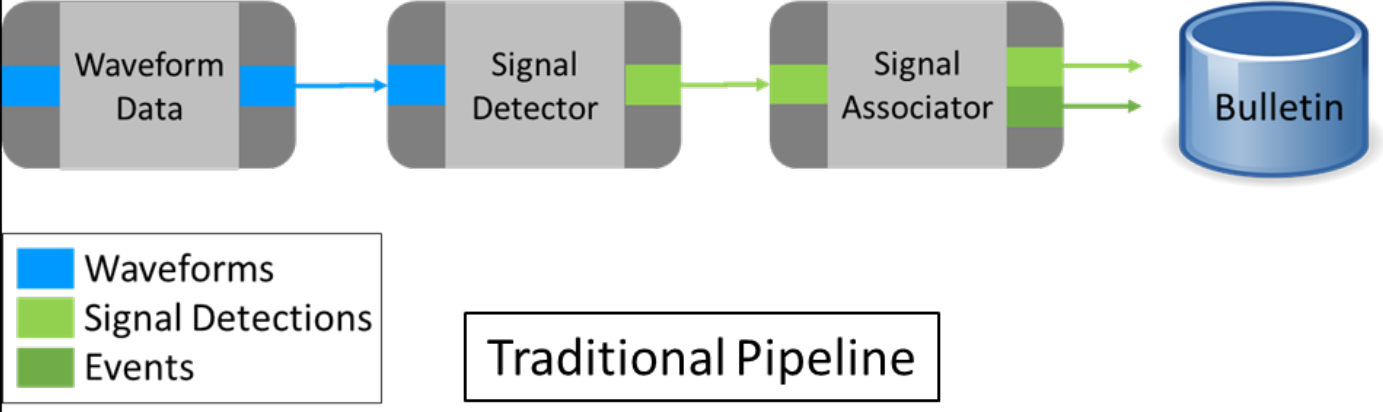




OVERVIEW

In a traditional data processing pipeline, waveforms are acquired, a detector makes the signal detections (i.e., arrival times, slownesses, and azimuths) and passes them to an associator. The associator then links the detections to the fitting event hypotheses to generate an event bulletin. Most of the time, this traditional pipeline requires substantial 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

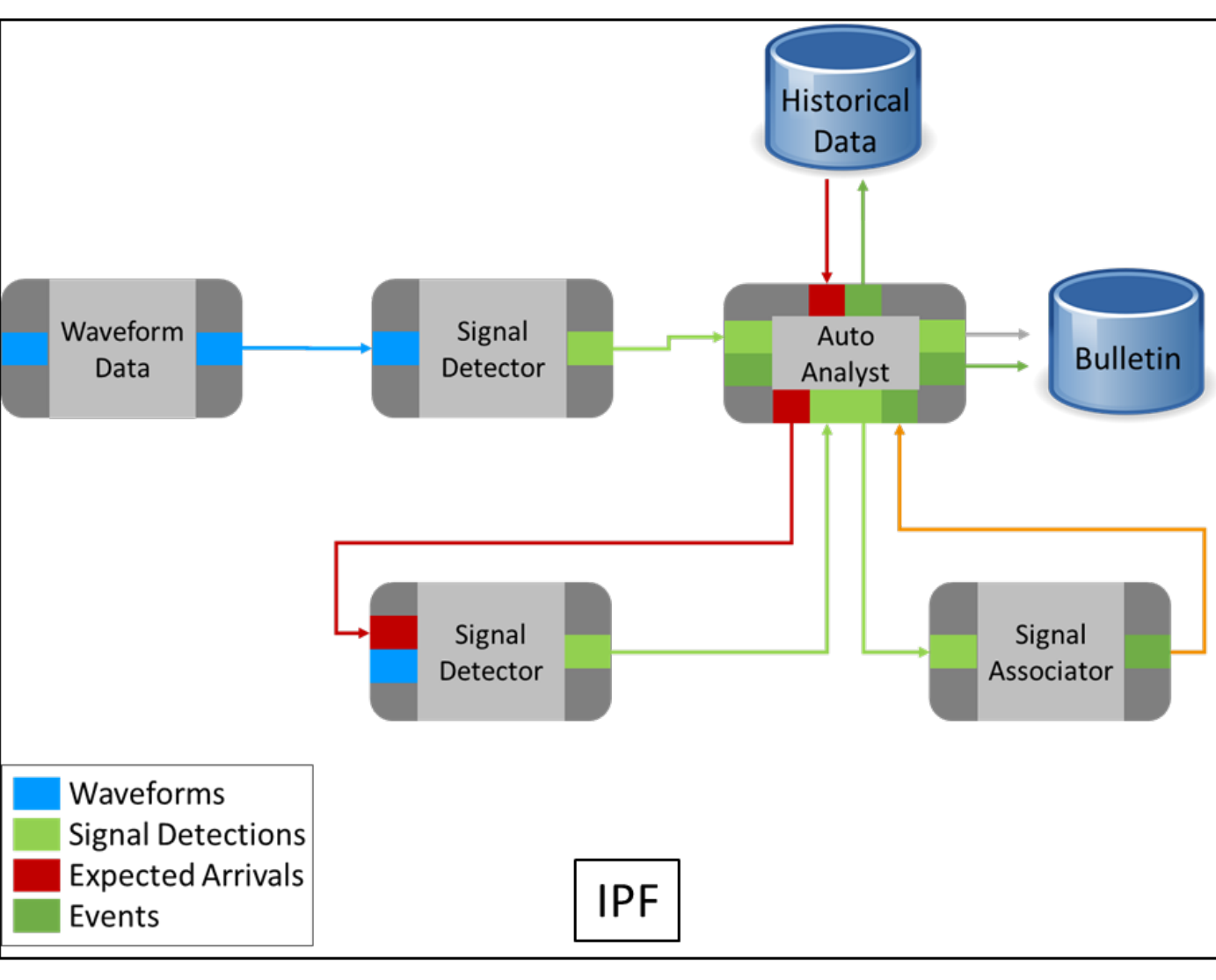


In a traditional pipeline of event building,
 1. Waveform data are processed by a signal detector,
 2. The signal detections are passed to a signal associator, and
 3. The associator links the detections to the fitting event hypotheses to generate an event bulletin.

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 signals will continue to increase as additional IMS stations come online.
To reduce the human-analyst workload during the seismic event building process, rather than a traditional linear sequence of data processing steps, we are proposing an Iterative Processing Framework (IPF) that includes a new data processing module, which incorporates automatic analyst behaviors.

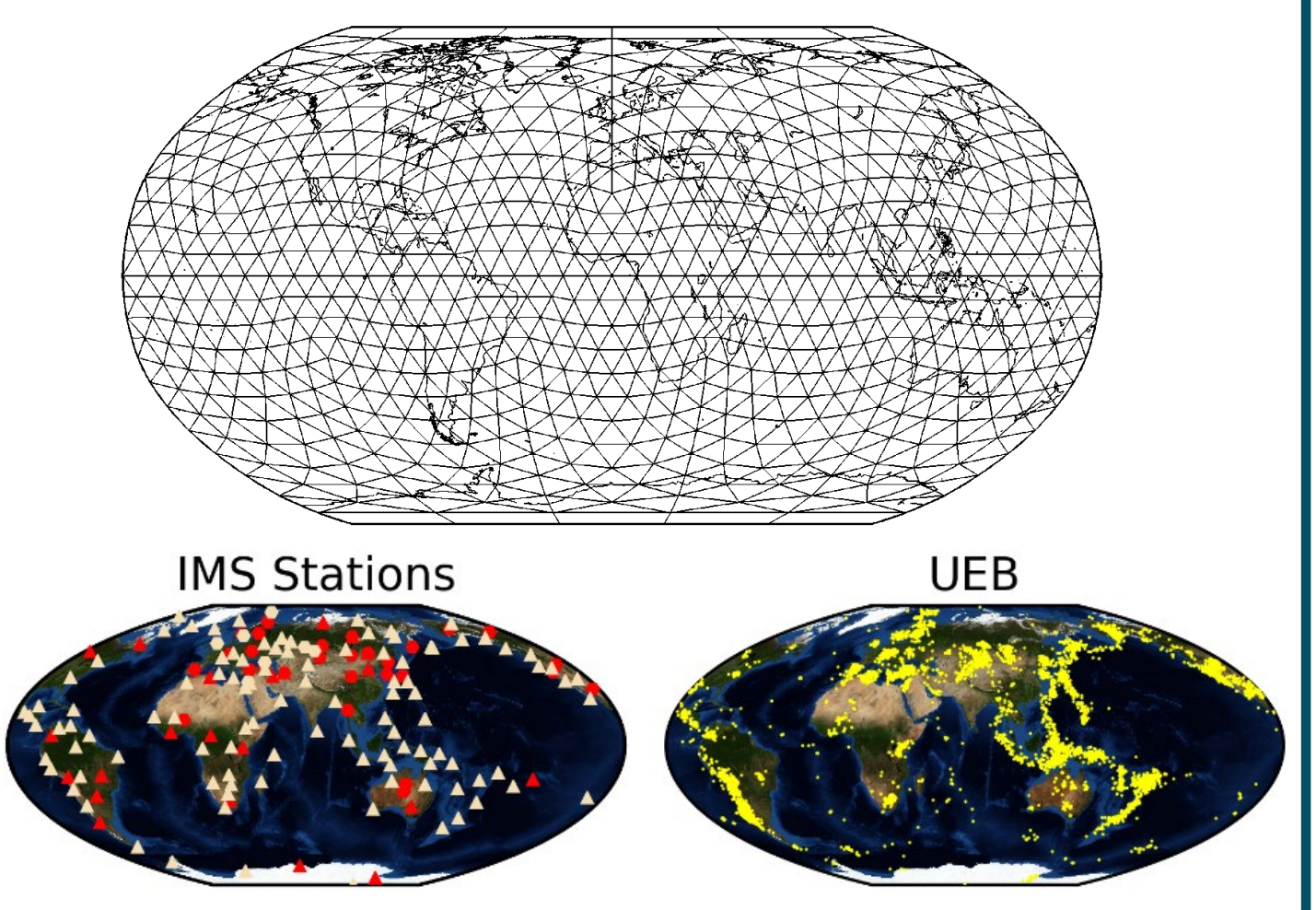
THE PROPOSED PROCESSING FRAMEWORK



The proposed framework, IPF, includes:
 • All the components of a traditional pipeline, plus
 • A new controller module (Auto Analyst, AA) that allows our system to more closely model the way analysts actually work data, including querying 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
 • Search for additional detections
 • Rerun associator if any arrivals are modified or new arrivals are detected in the process
2. Scan unassociated arrivals to build potential missed events
 • Identify unassociated arrivals from array stations
 • Compute single station locations
 • Query historical data for missing expected arrivals
 • Reprocess waveforms searching for expected arrivals
 • Rerun the associator with any new detected arrivals

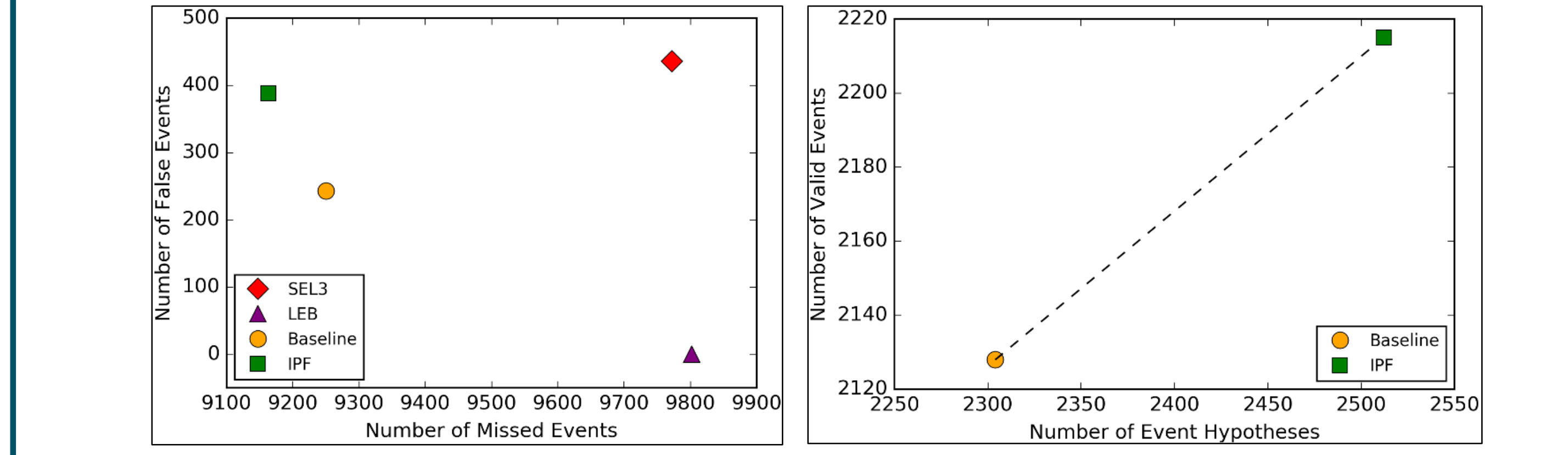
TEST SETTINGS AND DATA

- The association algorithm used in IPF is PEDAL (Draeos *et al.*, 2015).
- The Earth is divided into 2° triangular 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 (15–28 May 2010) of signal detections automatically generated by IDC.
- 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 far right figure).
- UEB is the ground truth (GT) used to assess the quality of the bulletins (baseline, and IPF bulletins, as well as SEL3, and LEB).
- 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.

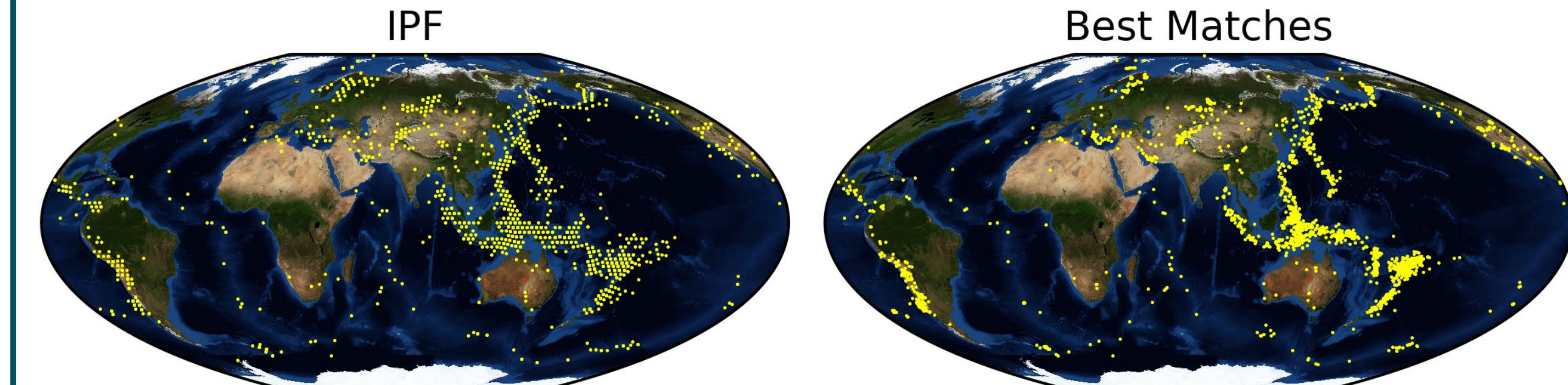


RESULTS

Bulletin Comparison

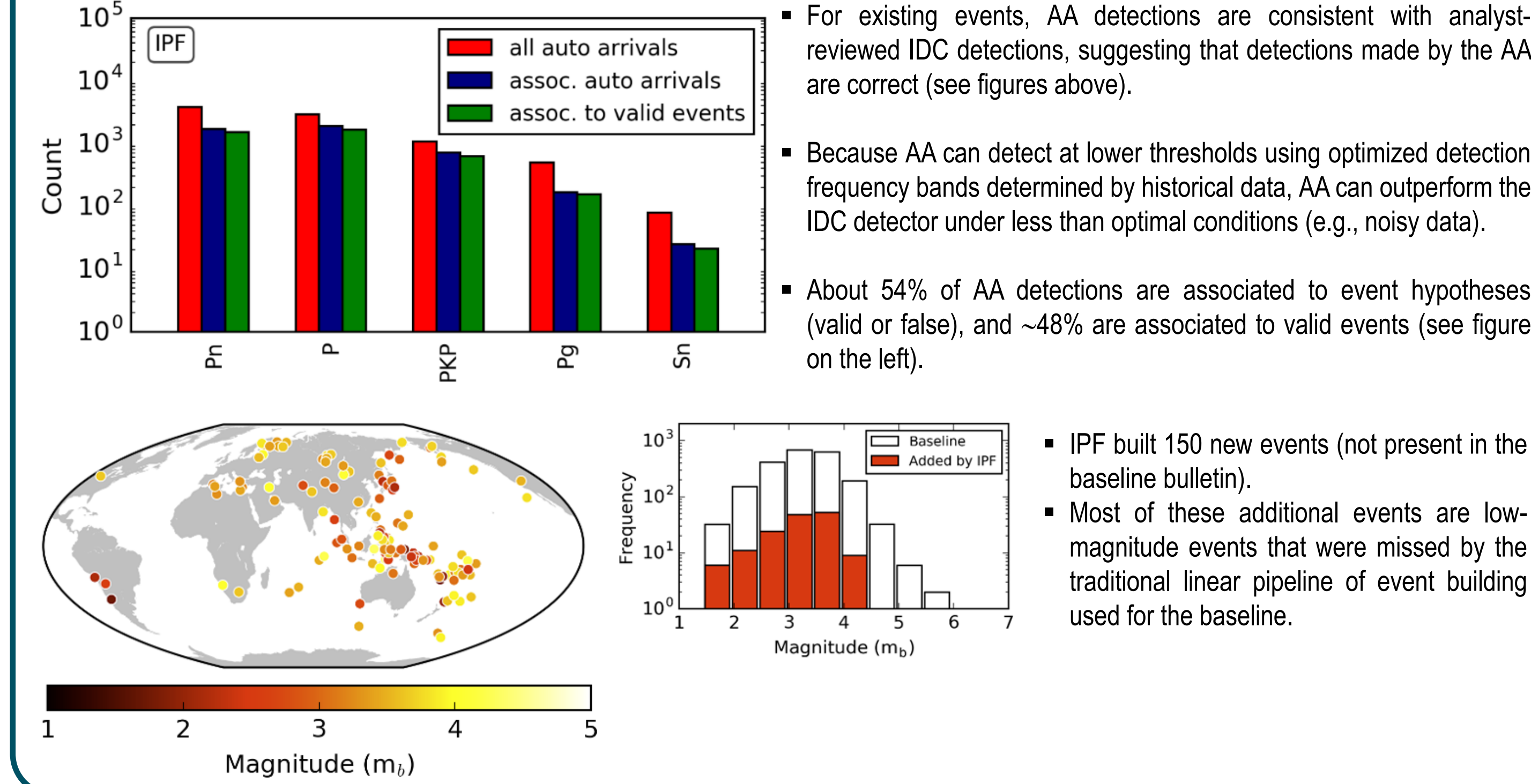
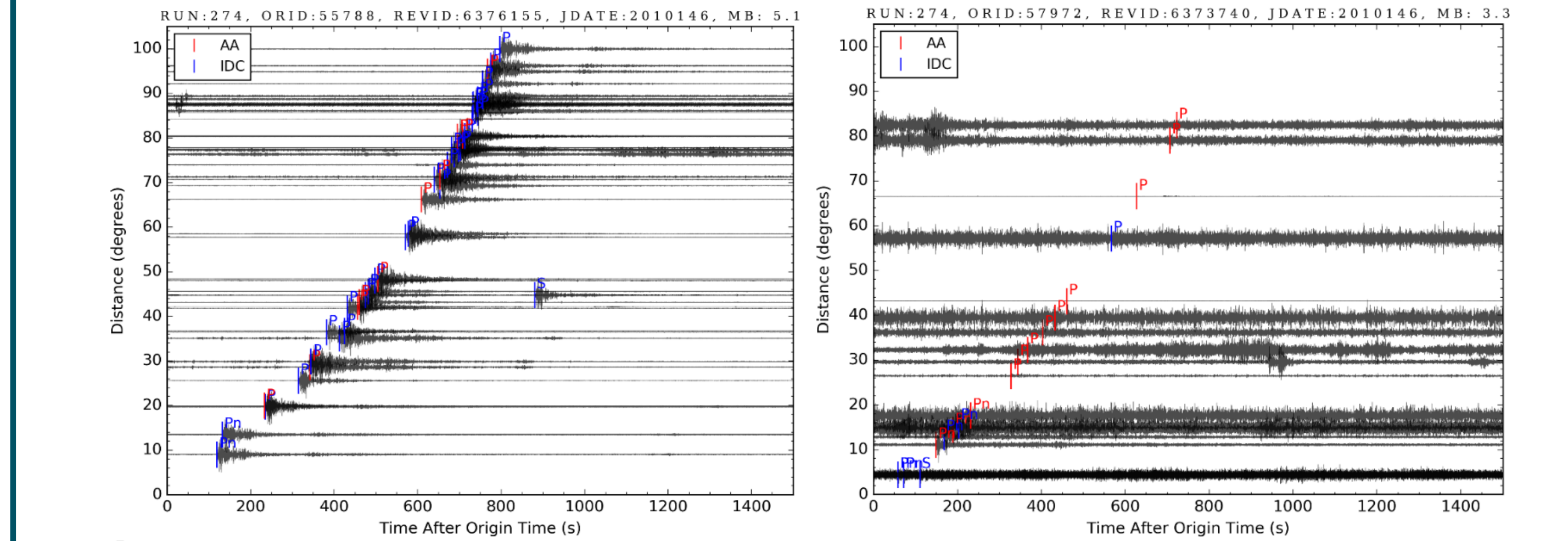


- SEL3 contains high numbers of false events and missed events, explaining why it is reviewed and augmented manually to produce the LEB (top left figure).
- LEB contains no false, but has the largest number of misses (9,802) of any bulletin, including SEL3.
- Baseline (PEDAL) misses 522 fewer legitimate 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 framework performance.
- About 90% of the new, IPF-generated event hypotheses turned out to be valid events (top right figure).



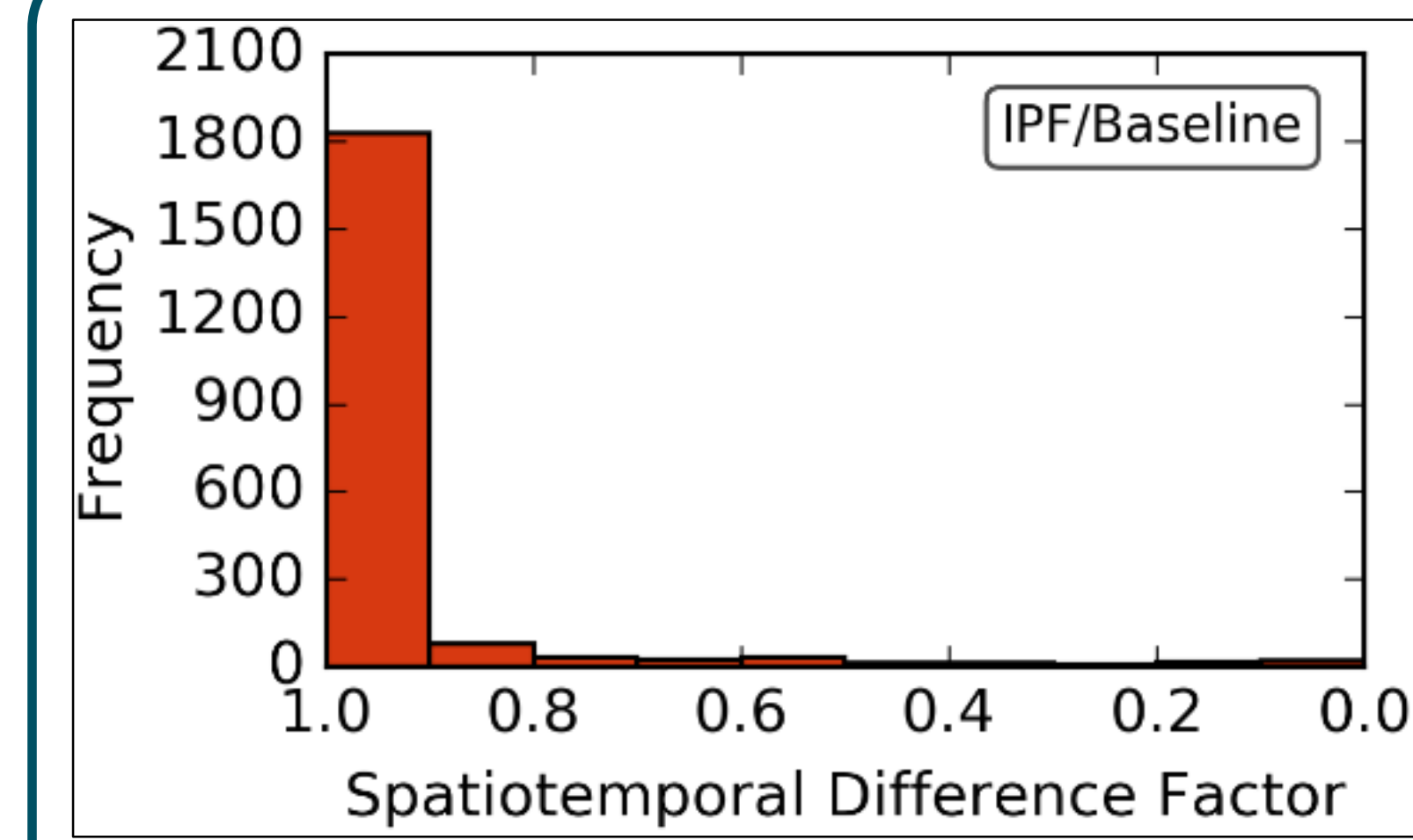
- The location of the events from the IPF framework are constrained by the 2-degree grid size.
- Nevertheless, these locations are good enough to outline known seismic zones.
- For the most part, location patterns are consistent between IPF bulletin and the best matches from the GT.

Analysis of Auto Analyst Impact



- For existing events, AA detections are consistent with analyst-reviewed IDC detections, suggesting that detections made by the AA are correct (see figures above).
- Because AA can detect at lower thresholds using optimized detection frequency bands determined by historical data, AA can outperform the IDC detector under less than 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).

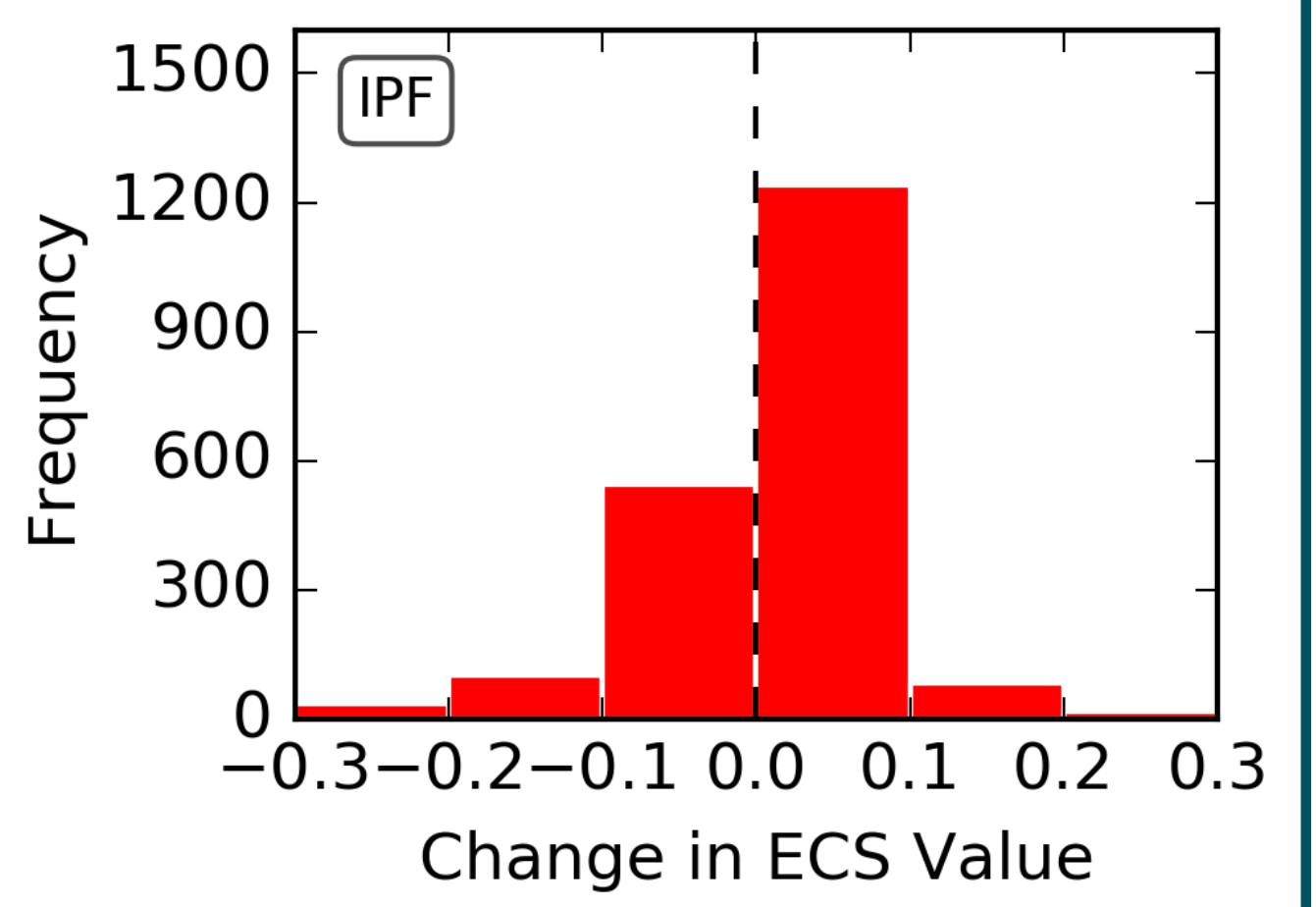
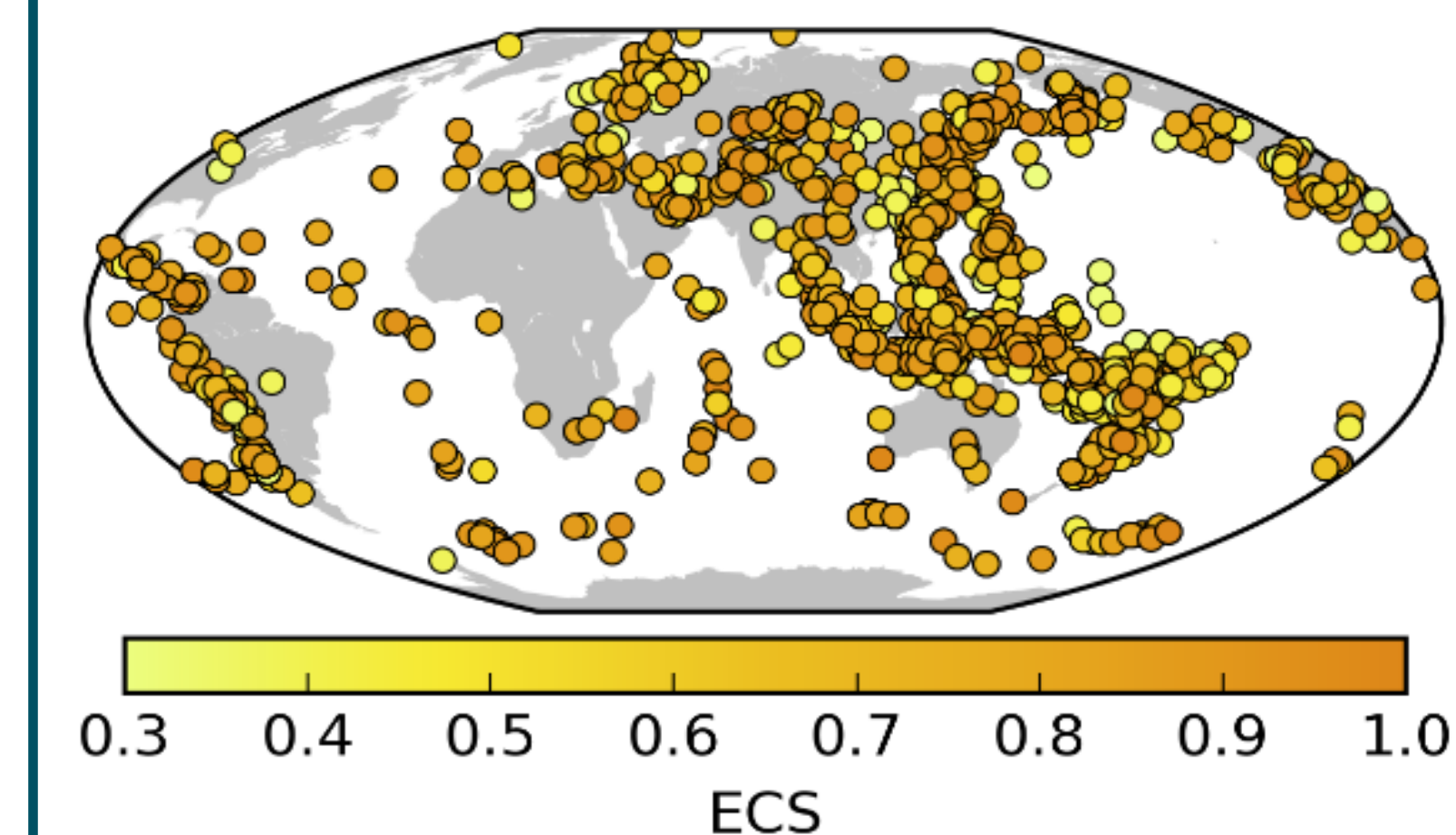
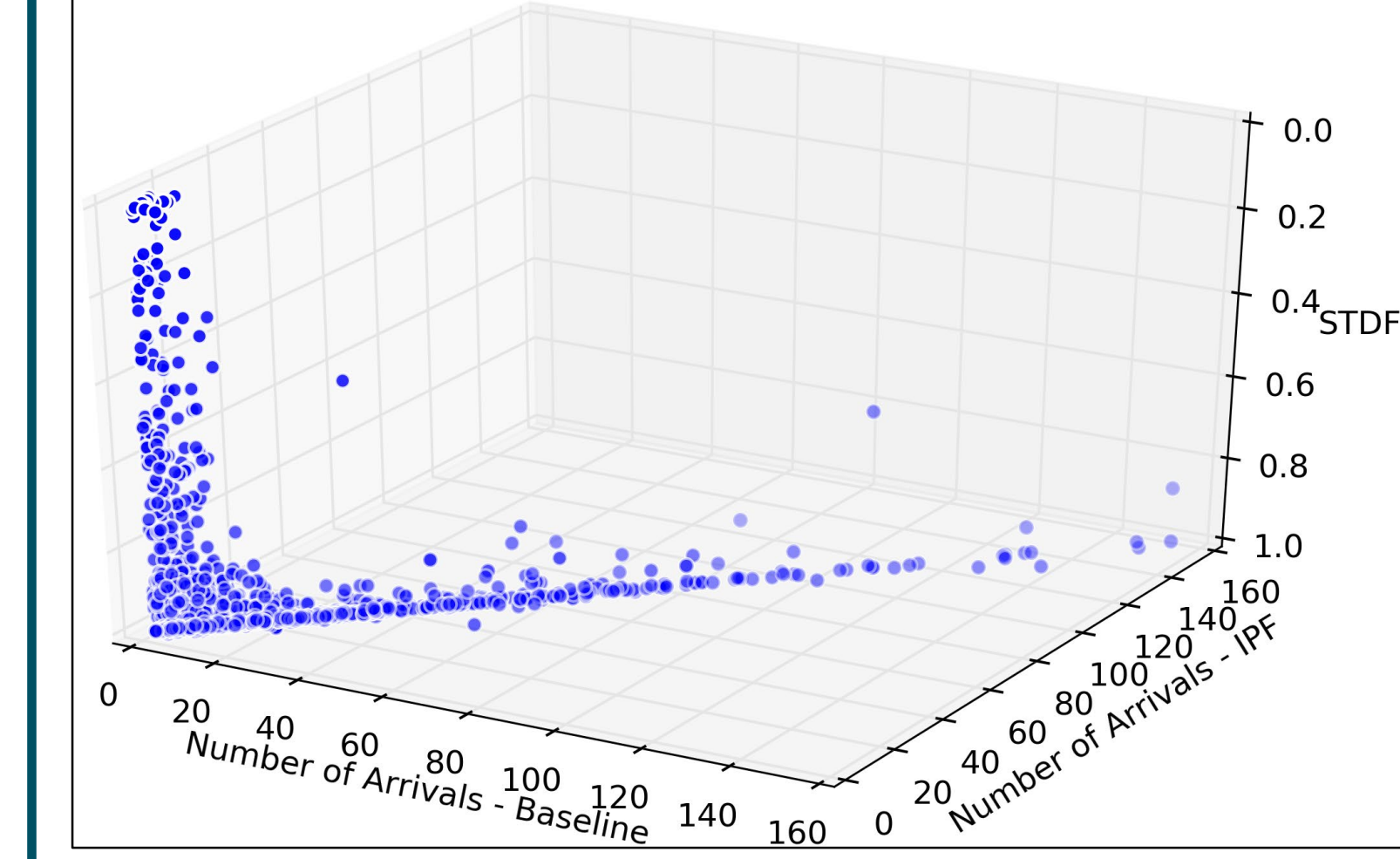
- IPF built 150 new events (not present in the baseline bulletin).
- Most of these additional events are low-magnitude events that were missed by the traditional linear pipeline of event building used for the baseline.



The spatiotemporal difference factor (STDF) between events that are in common between 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 in km between the 2 epicenters, ΔT_0 the difference in adjusted event origin times, V_p the mantle *P*-wave velocity of 10 km/s, and A the distance scaling factor of 1500 km.

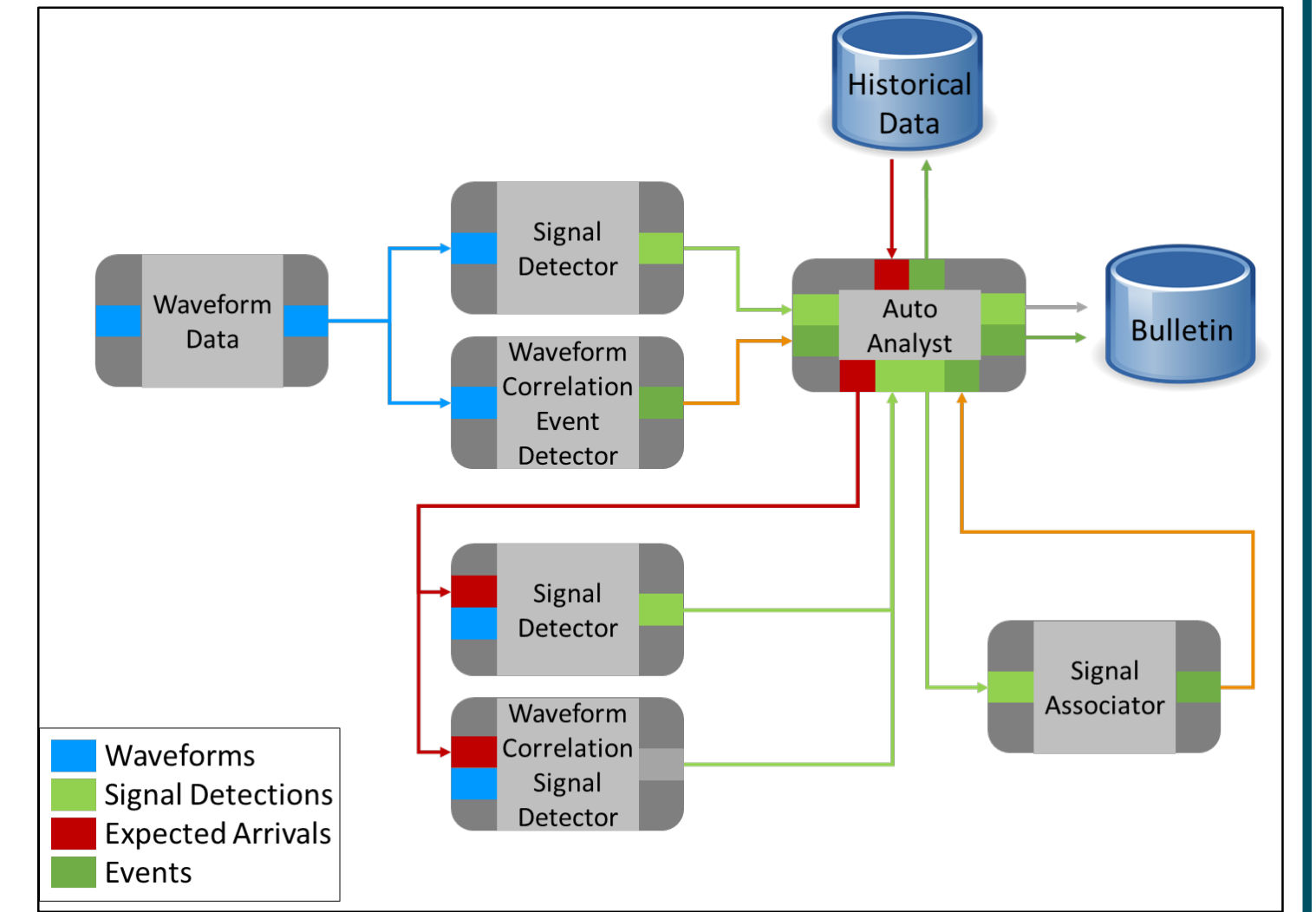
- An STDF value close to 0 indicates that the 2 events are distant in space and/or time.
- For the 2065 events common between the baseline and IPF bulletins, most of the STDF values lie between 0.9 and 1 (see figure on the top left), suggesting that for the large majority of events, the locations and origin times have not been significantly affected by the IPF processing.
- In general, only baseline events with a limited number of associated arrivals (< 10) are prone to significant changes in hypocentral locations and/or origin times when associations are modified by the IPF processing.
- For most events in the baseline bulletin, the AA in the IPF provides additional arrivals. This implied from the observation that the majority of data points in the figure on the left lies in the farther half.



About 85% of the IPF events have ECS values ≥ 0.6 , implying good matches (see figure above on the left). For most events that are in common between the baseline and IPF bulletins, IPF improves the ECS values (see figure above on the right). That is, IPF processing leads to better comparison with events in the GT.

CONCLUSIONS AND OUTLOOK

- IPF performs better than a traditional linear pipeline.
- Most of the additional events built by IPF are low-magnitude events that were missed by the traditional processing pipelines.
- IPF also added additional signal detections to existing events, which saves analyst time, even if the event locations were not significantly affected.
- The performance of the IPF improves with an increasing number of event hypotheses generated by AA.
- About 90% of the new IPF-generated event hypotheses turned out to be valid events.
- Assuming that the seismicity rate, *b*-value, and background noise level are all the same as for the test dataset, we project that for 1-year period, IPF would have built over 2000 more events than the traditional pipeline. These are events that human analysts would not have to painstakingly build manually.
- 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. Integrating a waveform-correlation detector (WCD) into the framework, as shown in the figure on the right, could significantly improve its performance by dramatically increasing the number of event hypotheses.
- The impact of a WCD on IPF performance is a research direction we may explore in future 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. The views expressed here do not necessarily reflect the views of the United States Government, the United States Department of Energy, the National Nuclear Security Administration or the Sandia National Laboratories. SAND2019-5375 C

Disclaimer: The views expressed on this poster are those of the author and do not necessarily reflect the view of the CTBTO