SAND2019-4285J

The Iterative Processing Framework: A New Paradigm for Automatic Event Building

Rigobert Tibi[1], Andre Encarnacao[1], Sanford Ballard[1], Christopher J. Young[1], Ronald Brogan[2], and Amy Sundermier[1]

[1]Sandia National Laboratories,

P.O. Box 5800,

Albuquerque, NM 87185-0404

Email: rtibi@sandia.gov

Phone: 505-844-9843

[2]ENSCO, Inc.

5400 Port Royal Road

Springfield, VA 22151

To be submitted to *Bulletin of the Seismological Society of America*

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.

Abstract

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 the year 2017 for example, International Data Center (IDC) analysts rejected about 40% of the events in the automatic bulletin and manually built 30% of legitimate events. We propose an Iterative Processing Framework (IPF) that includes a new data processing module that incorporates automatic analyst behaviors (Auto Analyst, AA) into the event building pipeline. In the proposed framework, through an iterative process, the AA takes over many of the tasks traditionally performed by human analysts. These tasks can be grouped into two major processes: (1) evaluating small events with low number of location-defining arrival phases (*NDEF*) to improve their formation; and (2) scanning for and exploiting unassociated arrivals to form potential events missed by previous association runs. To test the proposed framework, we processed a two-week period (1–14 May 2010) of the signal detections dataset from the IDC. Comparison with an expert analyst-reviewed ground truth bulletin for the same time period suggests that IPF performs better than the traditional pipelines (IDC and baseline pipelines). Most of the additional events built by the AA are low-magnitude events that were missed by these traditional pipelines. The AA also adds additional signal detections to existing events, which saves analyst time, even if the event locations are not significantly affected.

Introduction

To be able to assess whether a possible treaty violation has occurred anywhere in the world, the Comprehensive Test Ban Treaty Organization's (CTBTO) monitoring system at the IDC processes data from the International Monitoring System (IMS) seismic sensor network to detect, locate, and characterize seismic events that occur daily around the globe. In a traditional data processing pipeline, signal detections—consisting of arrival times, slownesses, and azimuths—are passed to a signal associator to form events (Le Bras *et al.*, 1994; Arora *et al.*, 2013; Draelos *et al.*, 2015). Based on the predictions from an assumed Earth model, the associator 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 (e.g., Gibbons *et al.*, 2016). At the IDC, for example, on average about 43% of the yearly events included in the automatic bulletin are rejected by human analysts as false; and 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 (Fig. 1). 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, rather than a traditional linear sequence of data processing steps (i.e., a "pipeline") we propose an Iterative Processing Framework (IPF) that includes a new data processing module, which automates analyst behaviors (Auto Analyst, AA). In the proposed framework, through an iterative process, the AA automatically performs many of the tasks traditionally performed by human analysts to improve the quality of the produced bulletin before the analysts review it.

Based on their respective end objectives, the AA behaviors involved are grouped into two major processes. One process consists of evaluating small events (with low numbers of defining signal detection, i.e., *NDEF*) by validating or refining their signal detections to improve their formation. This is achieved by performing multiple frequency-wavenumber (*f*-*k*) analyses of the signal detections with varying parameters until the optimum values for those parameters are found. The other process involves scanning for and exploiting unassociated arrivals to form potential events missed by previous association runs. Gibbons *et al.* (2016) made the case for a similar iterative strategy that could be applied to aftershock sequences to relieve human effort. Using their approach that involves a semiautomatic procedure, these authors were able to screen out half of the "nuisance" detections associated with the 2015 Gorkha, Nepal, aftershock sequence, thereby reducing the likelihood of false associations and spurious events hypotheses made with the remaining detections.

To test our framework, we processed a two-week period (1–14 May 2010) of the signal detections dataset from the IDC and compare the results with an expert analyst-reviewed ground truth bulletin for the same time period (Linville *et al.*, 2019). Our main findings are: (1) IPF performs better than the traditional pipelines; (2) most of the additional events built by the AA are low-magnitude events that evaded the traditional processing pipelines; (3) the AA adds additional signal detections to existing events, which saves analyst time, even if for most of these events the locations are not significantly affected; and (4) the performance of the proposed pipeline improves with increasing number of event hypotheses. About 90% of the new event hypotheses added by IPF are valid events.

The Proposed Processing Framework

In a traditional data processing pipeline of event building, waveform data are processed by a signal detector (e.g., energy burst detector such as an STA/LTA algorithm) to find detections which are then passed on to another process to determine useful features, e.g., *f-k* analysis for arrays and polarization analysis for three-component stations to measure azimuth and slowness. The characterized detections are then passed to a signal associator, which – based on model predictions of the signal travel times, azimuths and slownesses – links the detections to the fitting event hypotheses to generate an event bulletin (Fig. 2a). Importantly, in this type of processing, the processing steps always occur in the same linear sequence, e.g., the signal detection step only occurs once. IPF, the iterative framework we are proposing, includes all the components of a traditional pipeline, as described above, plus a new controller module that allows our system to rerun any of the other processing steps as deemed appropriate, thereby more closely modeling the way analysts actually work data. Throughout this paper, we refer to the new controller module simply as Auto Analyst or AA. The non-linear flow chart for IPF is depicted in Figure 2b. IPF also provides the possibility to query historical data for empirical information, which in some instances, as discussed below, guides the decision making about processing strategy. As mentioned above, we grouped the AA behaviors into two major processes, depending on their respective end objective:

(1) *Evaluating small (low NDEF) events to improve their formation*. For array stations, the AA validates or refines detections associated with small (low *NDEF*) events. This is achieved by performing *f-k* analyses of the detections with varying parameters until optimum values for those parameters are found. For any successful refinement (i.e., if

any arrival is modified in the process), the original detection is replaced by its modified version. The AA also searches for additional detections according to the procedure described in the next paragraph. Then the associations are reassessed by rerunning the associator with the updated set of signal detections.

(2) *Scanning for and exploiting unassociated arrivals to form potential events missed by previous association runs*. AA identifies any unassociated arrival from an array station, which has slowness consistent with a first *P* phase (i.e., *Pg*, *Pn*, *P*, *PKP*, *PKPab*, or *PKPbc*) and high signal-to-noise ratio (SNR). Using that arrival, AA computes a single-station location of the causal event. That location is used to query historical information to discover other stations that have a high probability of detecting a first *P* signal from the event (Fig. 3). AA then searches for the missing expected signals by reprocessing the waveforms using a short-term average/long-term average (STA/LTA) detector at lower thresholds, and in optimized detection frequency bands (i.e., station-specific) that are informed by the results of the query. The optimized detector bands and thresholds were both compiled by an expert-analyst with over 30 years of experience and familiarity with the noise characteristics of the IMS stations. The associator is then run again with any new detected arrivals. This capability of creating single-station events based on unassociated detections leads to a significant increase in the number of event hypotheses, which, as discussed below, results in dramatic improvement in IPF performance.

The event building process in our framework is iterative in the sense that the cycle of operations involved is repeated several times until no additional events are built.

Test Settings and Data

This work builds on the success achieved with PEDAL (Probabilistic Event Detection, Association and Location) by Draelos *et al.* (2015). Hence, PEDAL is the association algorithm we use in the proposed pipeline. For the test, we divided the Earth into a triangular tessellation of grid nodes, with 2° edge lengths (Ballard *et. al.*, 2016). The nodes are uniformly spaced at the surface, with depth nodes of 15-km radial spacing only in regions where, based on historical data, deep seismicity is expected to occur (e.g., subduction zones). This resulted in about 13,000 nodes. Since PEDAL assigns event locations to the grid nodes, the location accuracy for the built events is a function of the grid size. The selected grid resolution was established as optimal in the PEDAL study both in terms of the quality of the results and computational performance (Draelos *et al.*, 2015).

The input used to test the IPF pipeline consists of a two-week period (1–14 May 2010) of IMS network (Fig. 4, left) signal detections automatically generated by IDC data processing, i.e., the same input that the IDC associator used to create the Standard Event List 3 (SEL3) event catalog for that interval. The SEL3 catalog is then examined and corrected by IDC analysts to create the Late Event Bulletin (LEB). The LEB is an obvious choice to use to validate our IPF processing results for this 2-week interval, but the PEDAL study established that though the LEB is a high-quality bulletin, it is far from complete. To support multiple data processing research studies at Sandia National Laboratories, a seismic analyst with over 30 years of experience and high familiarity with IMS network manually compiled a more complete, high-resolution bulletin, which we refer to as the Unconstrained Event Bulletin (UEB, Fig. 4, right) (Linville *et al.*, 2019). This is the event catalog we use to evaluate our IPF processing results as well as those from the

IDC. The UEB, which consists of a total of 11,378 events and includes all the events from the LEB and many legitimate non-LEB events, is characterized as “unconstrained” because it did not require that events satisfy the stringent IDC criteria of at least three detecting stations. In fact, some events in the UEB were built using *P* and *S* signals from a single station, provided the directional information (i.e., azimuth) was also obtained. Locations for many UEB events should not really be considered ground truth in the classic sense because a one-station 3-component location using only *P*–*S* times and the azimuth garnered from polarization is subject to a large error in location. The goal here was to note that an event occurred, mark and associate the signals, and provide the best guess for the event location using the information available, even if only one station was used.

In this paragraph, for reason of simplicity, we refer to any bulletin to be evaluated as B_E. Our method for bulletin comparison uses the event commonality metric described in Draelos *et al.* (2015). The event commonality score (ECS) between an event from the B_E and a reference event from the UEB takes into account both spatio-temporal differences between hypocenters as well as differences in sets of associated phases and is defined as:

$$ECS = STDF + FAF + MAF, \quad (1)$$

where *STDF* is the spatio-temporal difference factor, *FAF* the false association factor, and *MAF* the missed association factor.

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

where $d = D_{epi} + (\Delta T_0 \times V_P)$, D_{epi} is the distance between the two epicenters, ΔT_0 the difference in the adjusted event origin times (each origin time is adjusted to account for the event depth), V_P assumed mantle *P*-wave velocity of 10 km/s; and A is the distance scaling factor of 1500 km.

$$FAF = \frac{a_s}{a_{AUTO}}, \tag{3}$$

where a_s is the number of shared associations between the events, and a_{AUTO} the total number of associations for the automated event in B_E.

$$MAF = \frac{ta_s}{ta_{REF}}, \tag{4}$$

where ta_s is the number of shared time-defining associations between the events, and ta_{REF} the number of time-defining associations for the reference event in the UEB.

The higher the ECS, the better the match between the two events. The event from B_E that best matches an event in the UEB and has an ECS of 0.3 or larger is characterized as *valid*. Events in the B_E having best match in the UEB with ECS < 0.3 are *false*. *Split* events are events in the B_E that share arrivals with the same reference event in the UEB. A *split* event is *valid* if ECS $\geq$ 0.3, and false if ECS < 0.3. B_E events having ECS $\geq$ 0.3 with multiple events in the UEB are characterized as *merged* and treated as false. A *missed* event is any event in the UEB with either no match in the B_E or with a best match that has an ECS below 0.3.

For the test, we investigated the following two cases:

- The baseline scenario, which is equivalent to a traditional (linear) event building pipeline with PEDAL as the associator, and does not involve the AA.
- The full IPF, which involves the implemented iterative AA behaviors.

Results and Discussion

Bulletin Comparison

For each of the bulletins produced using the test dataset (the baseline, and IPF bulletins, as well as SEL3, and LEB), we compare against the UEB to establish the number of false events and the number of missed events. The results are shown in Figure 5, left. A perfect match with the UEB would mean zero false events and zero missed events, and hence would plot at the coordinate origin. SEL3 has 436 false and 9772 missed events (Table 1), explaining why this bulletin must be reviewed and significantly augmented manually by analysts to produce the LEB. The LEB contains no false events but has the largest number of misses (9802) of any bulletin, including SEL3. The LEB has more missed events than SEL3 perhaps because the analysts follow more stringent criteria for what constitutes an event for the LEB than the automatic pipeline had to follow for SEL3. Comparing our baseline bulletin with SEL3 contrasts the performance of PEDAL (the association algorithm used in the pipeline) with GA (Global Associator), the associator currently used at the IDC (Le Bras et al., 1994). PEDAL misses 522 (~5%) fewer legitimate events and builds 193 fewer false events compared to GA. IPF generates a bulletin with 609 (~6%) fewer misses and 48 (~11%) fewer false compared to the automatic pipeline at the IDC. The proportion of reduction in missed events remains at about 6% for the IPF bulletin when compared with the LEB. IPF forms 150 additional valid events not in the baseline bulletin, but also doesn’t form 63 valid baseline events. The net result is that the IPF bulletin containins 87 events more than the baseline (Fig. 5, right). The elimination of some valid baseline events during IPF processing is because the associator starts from scratch (no events) and completely reprocesses the interval every time detections are modified or added, hence there is no guarantee that events from a prior iteration will be rebuilt. We investigated the eliminated valid baseline events by tracking down arrival associations. We found that most of the arrivals previously linked to those events are associated to IPF events with higher ECS values than the original events in the baseline bulletin. This suggests that AA uses those arrivals in the IPF pipeline to build events that better match with events in the ground truth bulletin. In summary, IPF performs better than the traditional pipelines, and is fully automated, which saves analyst time and effort.

Figure 6, left shows the grid-node locations of all the valid events obtained using the IPF. Although the event locations from the framework are constrained by the 2-degree grid location, it is clear that these locations correspond to known seismic zones, including convergence margins, and mid-oceanic ridges. The locations of the corresponding best matches from the ground truth bulletin are shown in Figure 6, right. As expected, for the most part the location patterns are consistent between the two bulletins. For cases in which an IPF event has a geographic location significantly different from the corresponding best match in the ground truth bulletin, the high commonality score (ECS ≥ 0.3) is due to the large number of shared associated arrivals between the two events (see Equation 1). Our results suggest that about 90% of the new, IPF-generated event hypotheses turned out to be valid events (Fig. 7, and Table 1).

Analysis of Auto Analyst Impact in the Pipeline

In this section, we investigate in some detail the impact of the AA—the key module in the IPF pipeline—on the event building process. For that purpose, we examined/contrasted common events between the IPF and the baseline bulletins in terms of signal detections, associations, locations, and ECS scores; moreover, we quantitatively assessed how the IPF pipeline changed these parameters. Figure 8a shows the waveform section for a valid magnitude m_b 5.1 event built by the IPF, an event that was also in our baseline. The original detections from the IDC and AA detections are indicated with blue and red vertical bars, respectively. It is intriguing that IPF provided new detections for this event since it seems unlikely that IDC processing would have missed such high SNR signals. After investigations, however, we were able to establish that IPF had *f-k* analyses redone because the azimuth and/or slowness values for some of the IDC detections were inconsistent with the event location. As expected, across the range of distance shown in the waveform section, AA and IDC *P* picks are all consistent with the same moveout. The consistency of the moveout between AA and the IDC picks is also valid for the m_b 3.3 event, whose waveforms are displayed in Figure 8b. In this case, however, the SNRs are so low that none of the *P* phases emerge above the background noise level on the single traces. The indicated AA detections are based on the beamed waveforms obtained after *f-k* analyses for the array stations. The observed consistency between AA detections and analyst-reviewed IDC detections implies that detections made by the AA are correct. Because AA can detect at lower thresholds using optimized detection frequency bands determined from historical data, AA can outperform the IDC detector under less than optimal conditions (e.g., noisy data). This is the case for the m_b 3.3 event, for which the AA provides twice as many arrival detections as the IDC detector (Fig. 8b). For the current test, AA detections consist mostly of first *P* phases (*P*, *Pg*, *Pn*, and *PKP*); many of them are IDC detections that have been refined by the AA during the event evaluation processes. About 54%

of AA detections are associated to event hypotheses (valid or false), and the proportion of AA detections associated to valid events is ~48% (Fig. 9).

For the test dataset, the IPF built 150 new events not present in the baseline bulletin. The locations of these events are displayed in Figure 10a, while their magnitude distribution is shown in Figure 10b together with that of the baseline bulletin. Most of the additional events built by the IPF are low-magnitude events that were missed by the traditional linear pipeline of event building used for the baseline. These events have evaded the traditional pipeline probably due to the low quality of their signals that—when inspected on single traces—hardly appear above the background noise levels. IPF can improve and exploit these signals because of iterative processing that is informed by empirical information, which allows it to detect at lower thresholds.

To evaluate the extent to which AA behaviors impact the built events, we investigated the differences in terms of origin times, locations, and arrival associations for events that are in common between the IPF and the baseline bulletins. In the process, two events (one from each bulletin) are declared common only if they shared the same matching event in the ground truth bulletin, according to the ECS formalism described in section “Test Settings and Data”. We used the spatio-temporal difference factor (STDF), defined in Equation 2, to assess the combined differences in locations and origin times between the common event in two bulletins. An STDF value close to 0 indicates that the two events are distant in space and/or time, and a value of 1 for this parameter indicates that the events are spatio-temporally co-located. Figure 11a shows the histogram of the calculated STDF values for the 2065 events common between the IPF and the baseline bulletins. Most of the STDF values lie between 0.9 and 1, suggesting that for the large majority of events in the baseline bulletin that are also in the IPF bulletin, the locations and origin times have not been significantly affected by the IPF processing.

Figure11b is a 3D plot showing the calculated STDF values on the vertical axis, with the number of associated arrivals for the IPF and the baseline on the two horizontal axes, respectively, for the 2065 events that are common between the two bulletins. The reported STDF values suggest that, 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 in the IPF pipeline. With respect to the baseline, the data points of events for which arrivals have been added to (or removed from) the association by the IPF processing are in the farther (or nearer) half of the 3D plot (Fig. 11b); data points on the vertical plane along the left-right diagonal are events whose number of associated arrivals were unchanged after IPF processing. For most events in the baseline bulletin, the AA in the IPF pipeline provides additional arrivals. This is implied from the observation that the majority of the population of data points in Figure 11b lies in the farther half of the 3D plot.

The locations of all the events from the IPF bulletin are shown in the map displayed in Figure 12a. Each circle in the map is shaded according to the ECS between the event it represents and the event's best match in the ground truth bulletin. About 85% of the IPF events have ECS values of 0.6 or higher, implying good matches. The distribution of the differences in ECS values for common events between the IPF and the baseline bulletin is shown in Figure 12b. A positive (negative) difference indicates that the ECS for the IPF event is greater (smaller) than that for the corresponding baseline event. Events with positive changes in ECS values constitute the largest population, demonstrating that for most existing events, the IPF pipeline improves the ECS values. That is, by adding arrivals to the association of existing events, IPF processing generally leads to better comparison with events in the ground truth bulletin, although in most cases the event locations are not significantly affected.

Conclusions and Outlook

Traditional linear event building pipelines require substantial human-analyst involvement to improve the quality of the generated event bulletins. To reduce the heavy reliance on human analysts, which will only get worse if an organization wants to monitor to lower detection thresholds, we propose an iterative processing framework that includes a new automated controlling module (the "Automatic Analyst" or AA) that mimics analyst behaviors. The AA behaviors in the IPF pipeline are grouped into two iterative processes. The first process consists of evaluating small events (with low *NDEF*) to improve their formation by tasking the signal detector to reprocess targeted waveform intervals for selected high probability of detection stations based on queries of archived processing results. The second process involves scanning for and exploiting unassociated arrivals to form potential events missed by previous association runs. Just as with a traditional pipeline, IPF automatically produces an event bulletin, but because the bulletin is of higher quality, the amount of analyst review and editing is reduced. We tested IPF by processing two-weeks of May 2010 IDC signal detections and validating against a high-resolution analyst-reviewed bulletin for the same time period, and found that IPF performs better than a traditional linear pipeline. Most of the additional events built by IPF are low-magnitude events that evaded the traditional processing pipelines. In addition to building events that were missed by the traditional pipeline, IPF also added additional signal detections to existing events, which saves analyst time. We found that 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. For the two-week period of the test dataset, IPF built 87 events more than the baseline pipeline. If we assume that the seismicity rate, the *b*-value, and the background noise level are all the same as for the test dataset, we project that for a one-year period, IPF would have built over 2000 more events. These are events that human analysts would not have to painstakingly build manually.

Our framework is flexible and could accommodate additional processing modules. From the perspective of global monitoring, repeating events like aftershock sequences or mining explosions constitute nuisance events that any monitoring system must resolve. For repeating events, waveform correlation is known to be superior to traditional detection algorithms because of its ability to template-match signals with extremely low SNRs (e.g., Gibbons and Ringdal, 2006; Slinkard *et al.*, 2016). Integrating a waveform-correlation detector (WCD) into the proposed pipeline could significantly improve its performance by dramatically increasing the number of events hypotheses during the event building process. The impact of a WCD on IPF performance is a research direction we may explore in future work, possibly leveraging our existing approximate nearest neighbor (ANN) approach that rapidly searches for matching waveforms in very large template libraries (Tibi *et al.* 2017).

Data and Resources

To test the performance of the Iterative Processing Framework (IPF) pipeline, we used two-weeks (1–14 May 2010) of the International Data Center (IDC) detections from the Comprehensive Nuclear-Test-ban Treaty Organization (CTBTO). The IPF bulletin is validated against the Unconstrained Event Bulletin (UEB, Linville *et al.* 2019). The UEB is a high-resolution bulletin for 1–14 May 2010, compiled by an analyst with over 30 years of experience and familiarity with the International Monitoring System (IMS) network. We also contrasted the IPF bulletin against the 1–14 May 2010 automatic Standard Event List 3 (SEL3), and the human-reviewed Late Event Bulletin (LEB) from the IDC.

Acknowledgments

This research was funded by the U.S. Department of Energy. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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-NA-0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government. We would like to thank John Merchant for discussions.

References

Arora, N. S., S. Russell, and E. Sudderth (2013). NET-VISA: Network processing vertically integrated seismic analysis, *Bull. Seismol. Soc. Am.*, **103,** no. 2A, 709–729.

Ballard, S., J. Hipp, B. Krauss, A. Encarnacao, and C. Young (2016). GeoTess: A generalized Earth model software utility, *Seismol. Res. Lett.*, **87,** no. 3, 719–725.

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.

Gibbons, S. J., T. Kværna, D. B. Harris, and D. A. Dodge (2016). Iterative strategies for aftershock classification in automatic seismic processing pipelines, *Seismol. Res. Lett.*, **87,** no. 4, 919–929.

Gibbons, S. J., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.* **165,** 149-166.

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*.

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.*

Slinkard, M., S. Heck, D. Schaff, N. Bonal, D. Daily, C. Young, and P. Richards (2016). Detection of the Wenchuan aftershock sequence using waveform correlation with a composite regional network, *Bull. Seismol. Soc. Am.* **106,** no. 4, 1371–1379, doi: 10.1785/0120150333.

Tibi, R., C. Young, A. Gonzales, S. Ballard, and A. Encarnacao (2017). Rapid and robust cross-correlation-based seismic signal identification using an approximate nearest neighbor method, *Bull. Seismol. Soc. Am.*, **107,** no. 4, 1954–1968.

Full mailing addresses

Rigobert Tibi
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0404

Andre Encarnacao
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0401

Sanford Ballard
Sandia National Laboratories
P.O. Box 5800,
Albuquerque, NM 87185-0404

Christopher J. Young
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0401

Ronald Brogan
ENSCO, Inc.
5400 Port Royal Road
Springfield, VA 22151

Amy Sundermier
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0401

Tables

Table 1: Bulletin Comparison

Bulletin	Number of Event Hypotheses	Number of Valid Events	Number of Missed Events	Number of False Events
UEB (ground truth)	-	11378	0	0
SEL3	-	1606	9772	436
LEB	-	1576	9802	0
Baseline	2304	2128	9250	243
IPF	2512	2215	9163	388

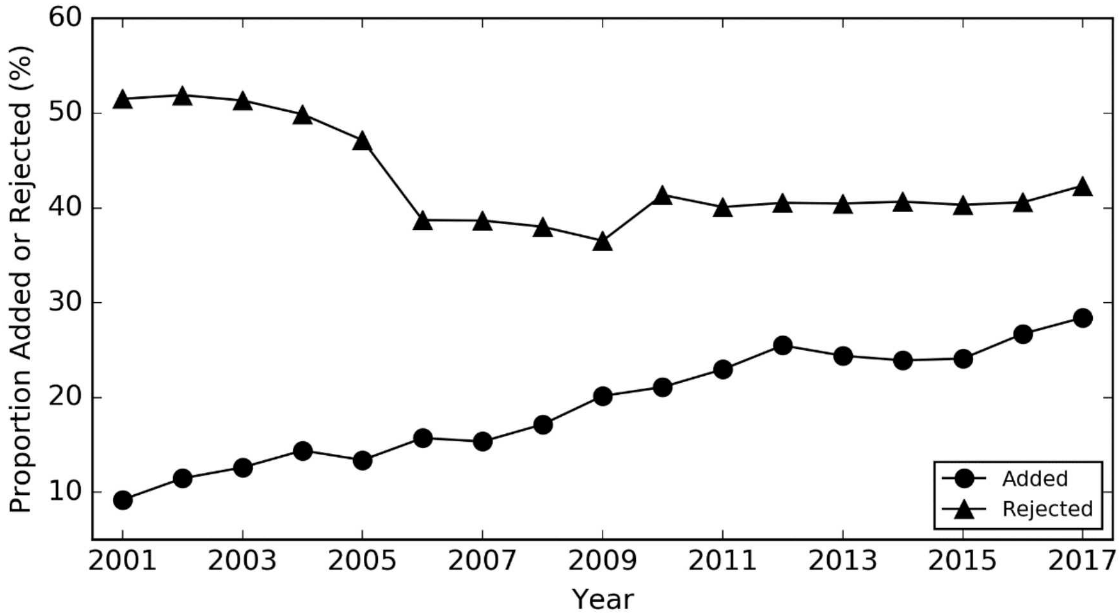

Figure 1. (Triangles) Proportion of yearly events in the IDC automated bulletin (SEL3) that are rejected by human analysts. (Circles) Proportion of yearly events added by human analysts to the IDC reviewed bulletin (LEB).

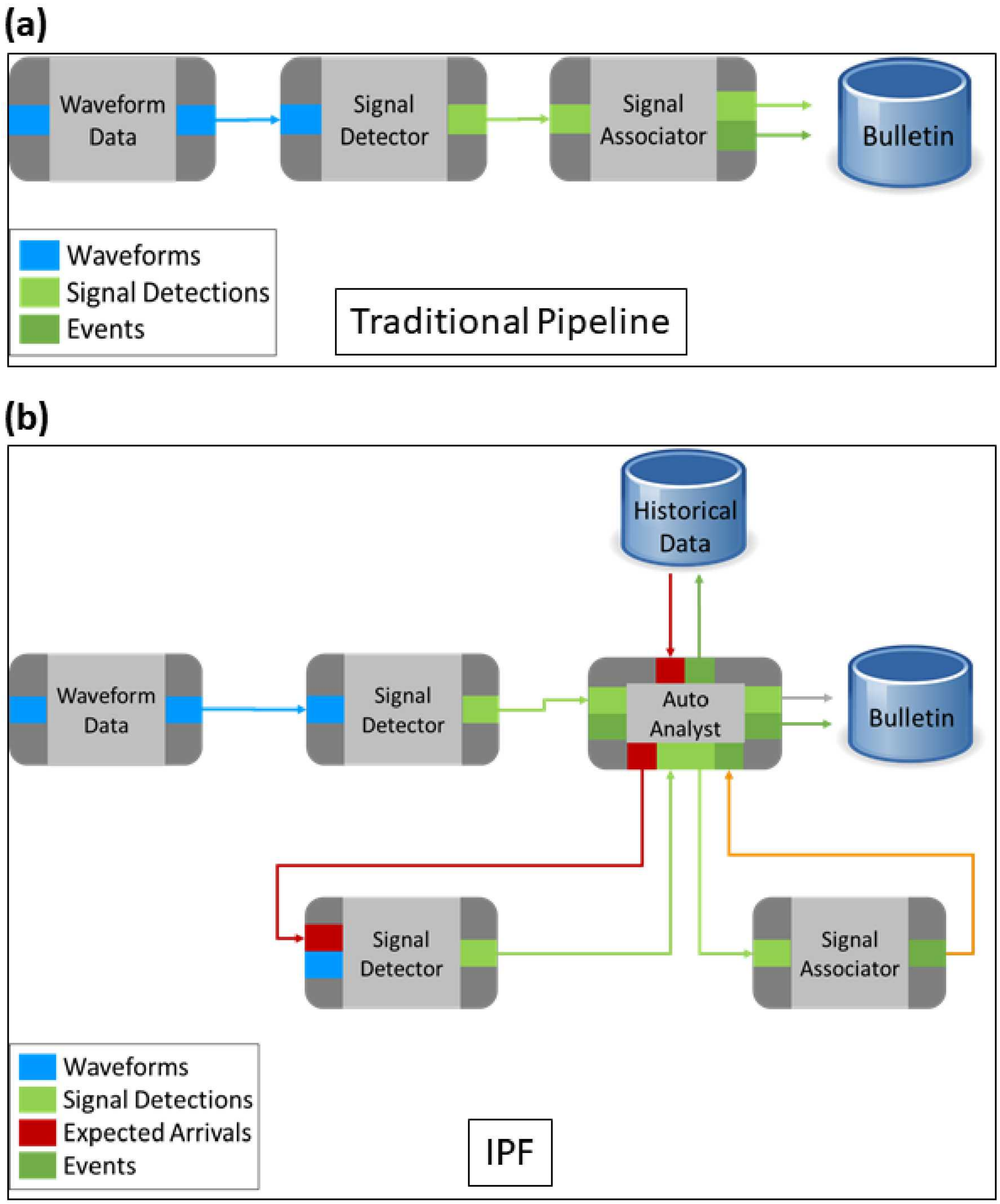

Figure 2: Flowchart for (a) a traditional pipeline of event building, and (b) IPF framework. The IPF framework in (b) includes all the components of a traditional pipeline plus the Auto Analyst, as well as the possibility to query historical data.

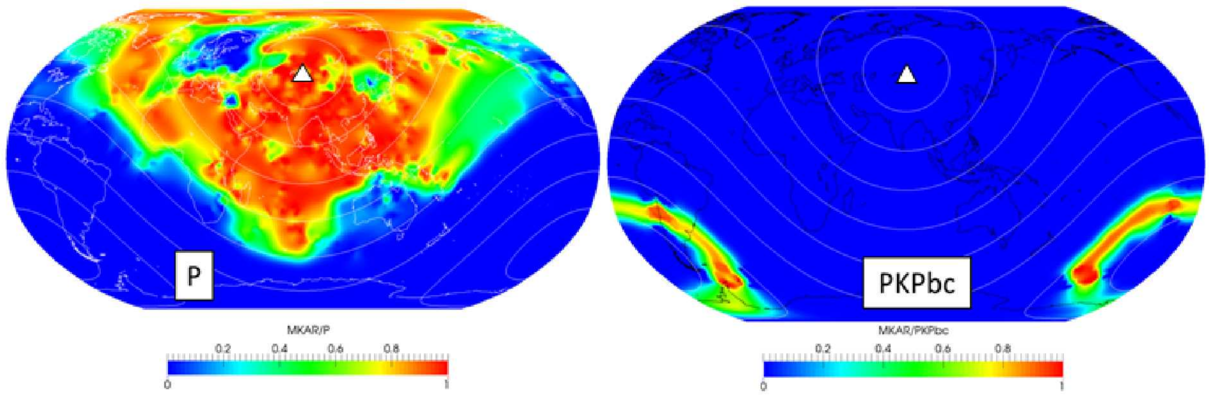

Figure 3. Maps showing the probabilities of detection for the IMS station MKAR for (left) *P* and (right) *PKPbc* arrivals. The probabilities were obtained after processing (including extrapolation and smoothing) of historical data from IDC LEB.

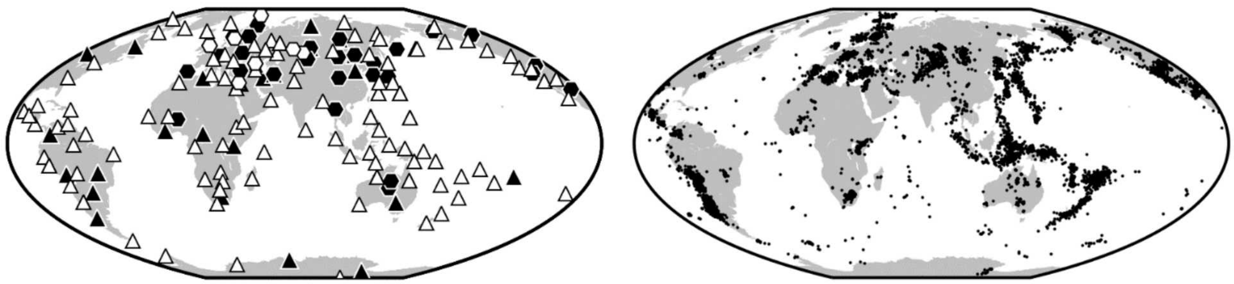

Figure 4. (Left) Map showing locations of the IMS stations. Triangles represent three-component stations; and hexagons are arrays. Filled and open symbols represent primary and auxiliary stations, respectively. (Right) Locations (dots) of the UEB events for the 2 weeks May 2010 interval.

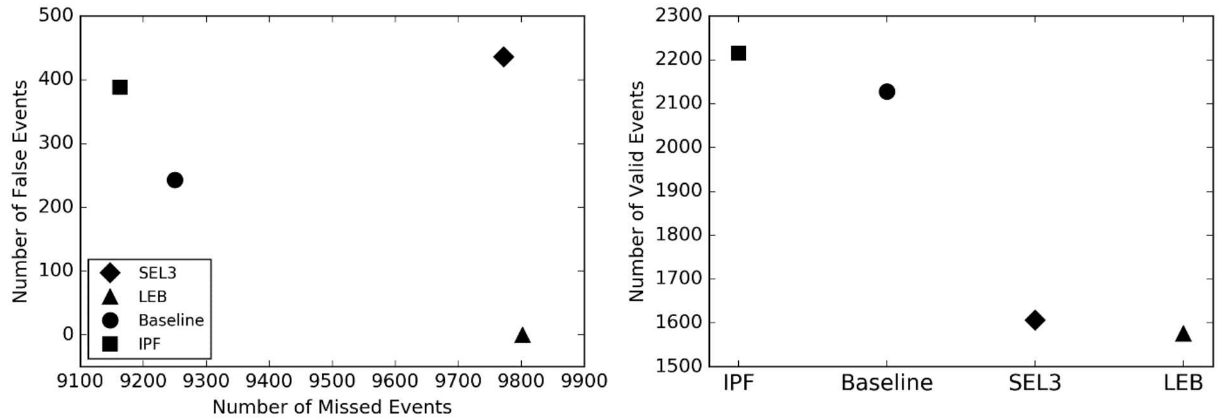

Figure 5. (Left) Number of false events vs. number of missed events for the SEL3 (diamond), LEB (triangle), baseline (circle), and IPF (square) bulletins. (Right) Number of valid events sorted by decreasing order for the different bulletins indicated on the horizontal axis.

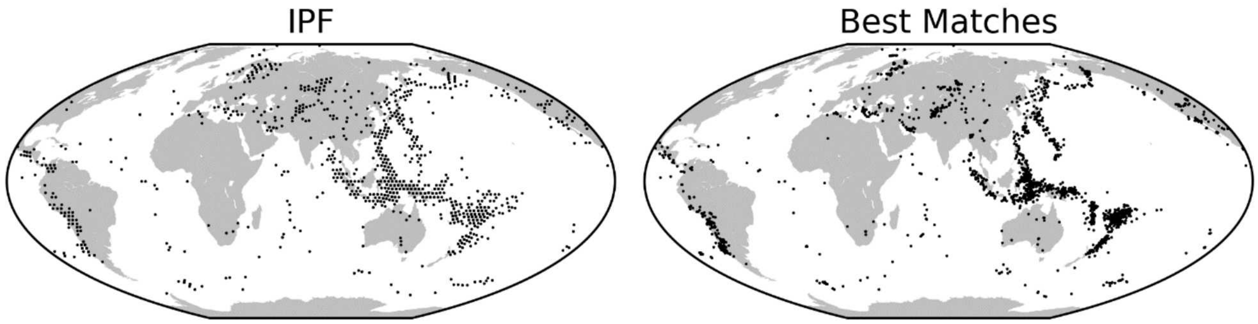

Figure 6. (Left) Map showing the grid locations of the 2215 valid events from the IPF bulletin. (Right) locations of the best matches in the UEB for the IPF events shown in the map on the left.

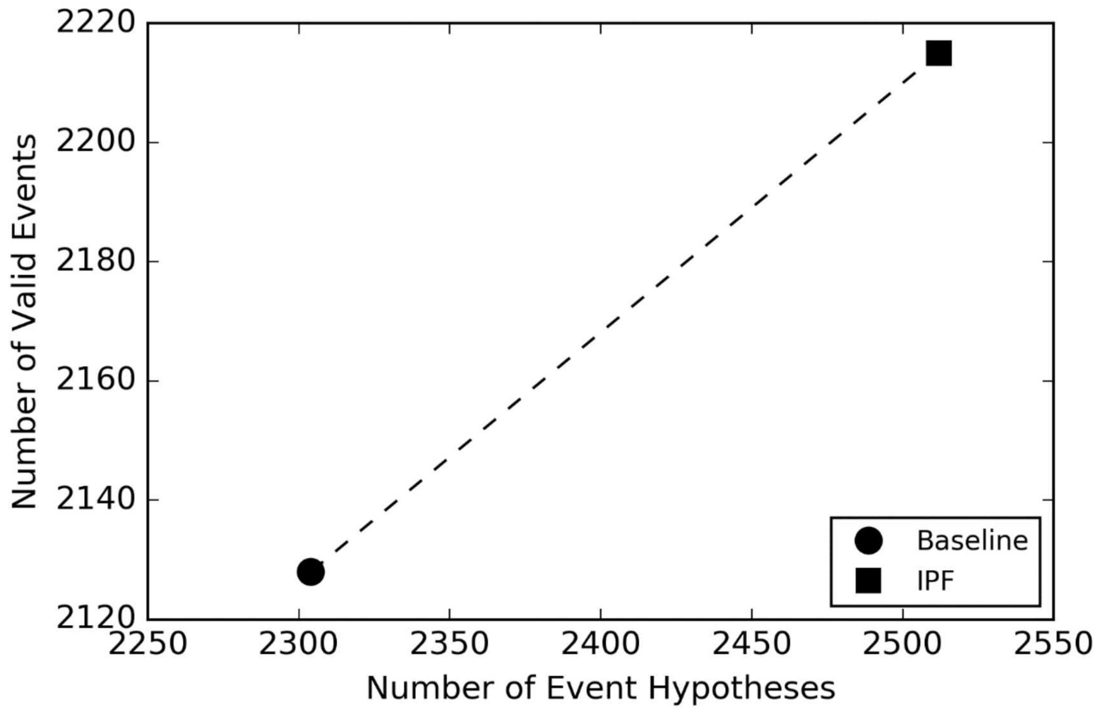

Figure 7. Number of valid events as a function of the number of event hypotheses for the baseline (circle) and IPF bulletins (square).

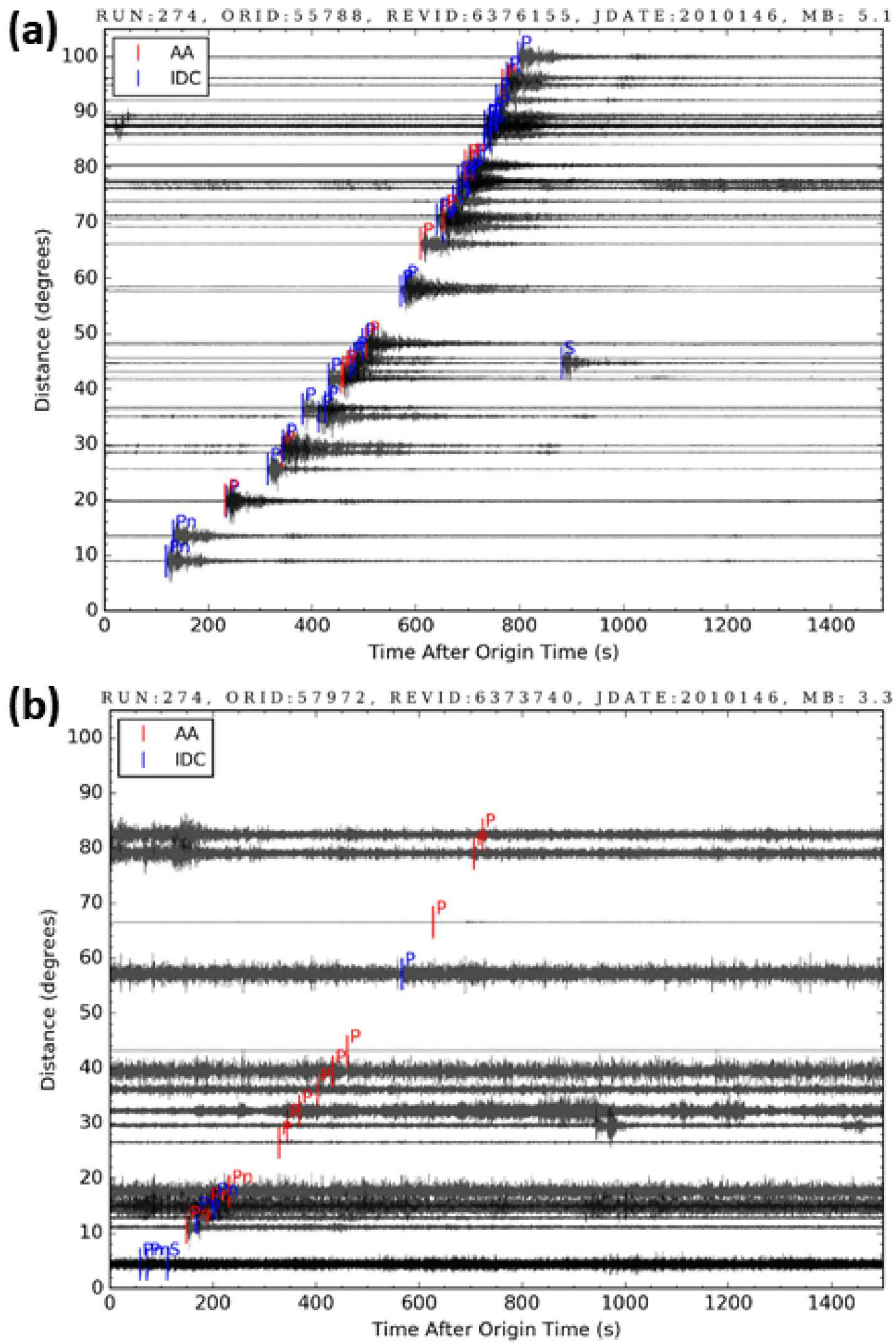

Figure 8. (a) Waveform section for an m_b 5.1 event. IDC and Auto Analyst's detections are indicated with blue and red vertical bars, respectively. (b) The same as in (a) but for an m_b 3.3 event. Note the contrast in SNRs between the waveforms in (a) and (b).

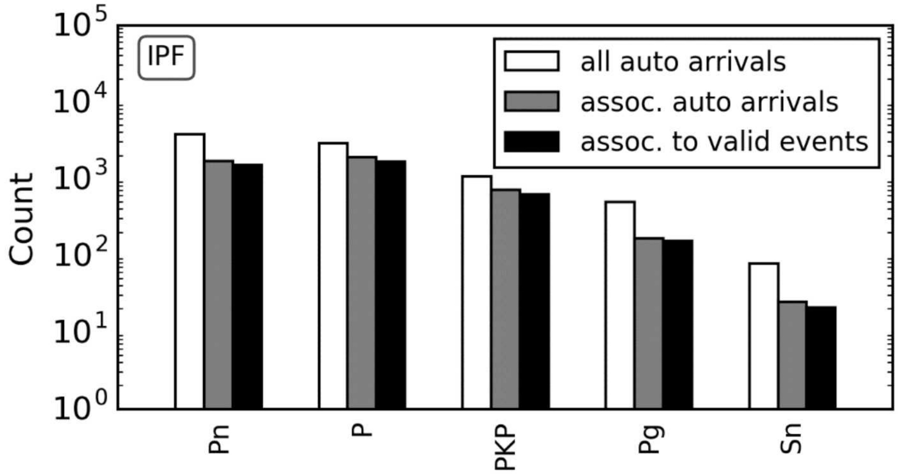

Figure 9. Bar plot showing the number of Auto Analyst's arrivals for each phase indicated on the horizontal axis. White bars indicate the total number of AA arrivals, gray bars the number of AA arrivals associated to event hypotheses, black bars the number of AA arrivals associated to valid events.

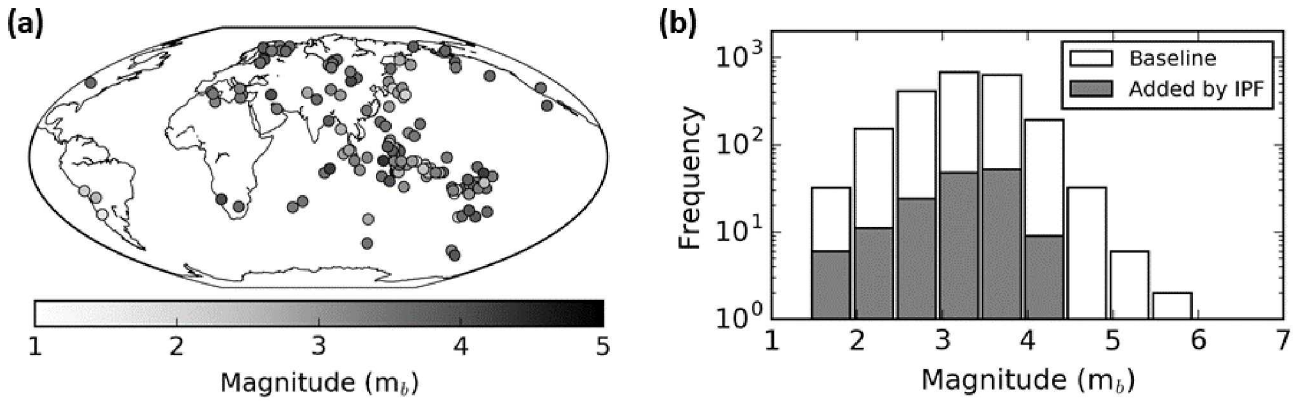

Figure 10. (a) Map showing the locations of the 150 additional events built by the IPF, which are not available in the baseline bulletin. Each circle in the map is shaded according to the magnitude of the event it represents. (b) Magnitude distribution for the 150 additional IPF events (gray) and the 2128 baseline events (white).

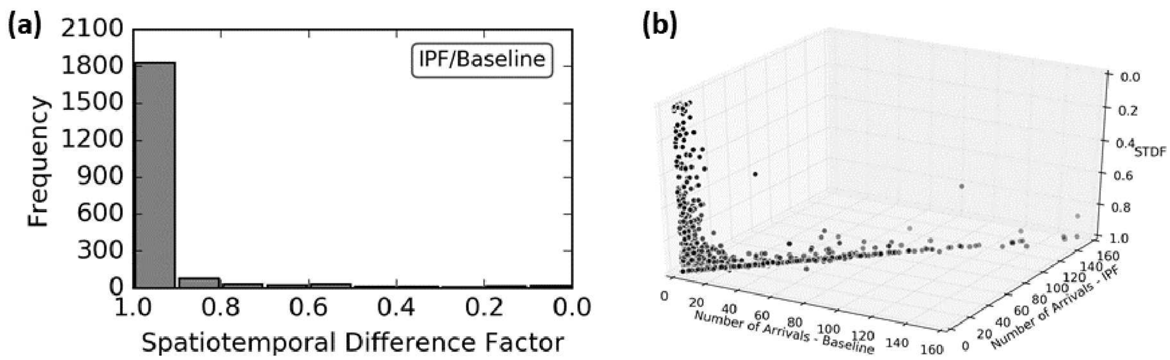

Figure 11. (a) Distribution of the STDF values for the 2065 common events between the IPF and the baseline bulletins. (b) 3D plot showing the calculated STDF values on the vertical axis, with the number of associated arrivals for the IPF and the baseline on the two horizontal axes, respectively, for the 2065 events that are common between the two bulletins. Note that the majority of the population of data points lies in the farther half of the 3D plot.

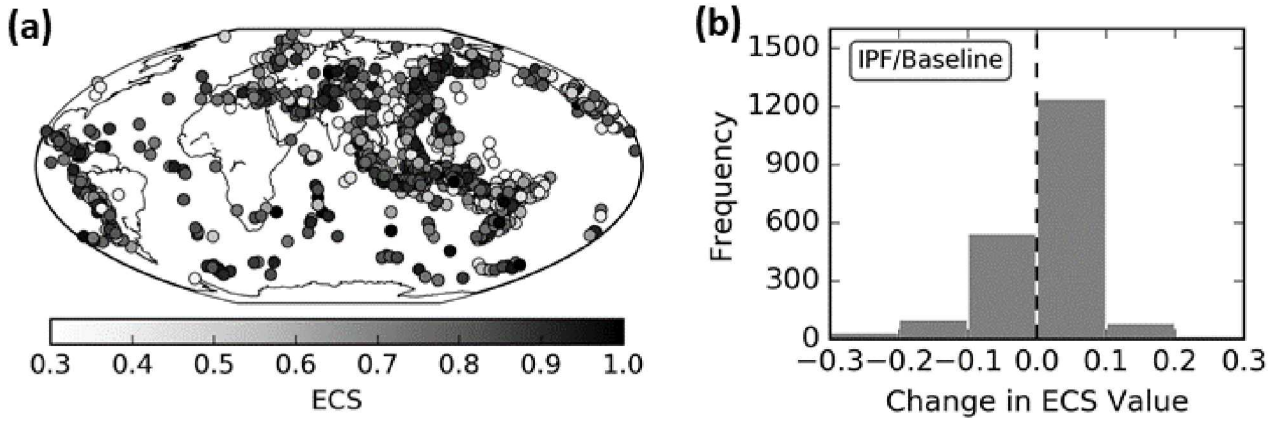

Figure 12. (a) Locations of the IPF events. Each circle is shaded according to the ECS value (with respect to the best matching event in the ground truth bulletin) for the event it represents. (b) Distribution of the changes in ECS values between the IPF and baseline bulletins for the 2065 events that are common between the two bulletins.