geoconvention
Calgary • Canada • May 15-17 2023

Benchmarking and Validation of Automated Real-Time Seismic Event Detection

Aaron Booterbaugh, Evgeny Naumov, Sepideh Karimi, Ben Witten, Adam M. Baig
Nanometrics

Introduction

Regional and local networks of seismic equipment have expanded greatly in industrial settings over the last decade in response to regulatory environments in different jurisdictions, efforts to maintain a social license to operate, and to gain an increased understanding of faulting and stress in different areas. The amount of seismic data that needs to be winnowed for events, located, and characterized has also expanded. The load on processing such data, often with reporting turnaround times of minutes for first-pass locations and magnitudes, requires automated processes to provide these outputs and to triage the data to optimize analyst time.

In this paper we describe the components necessary to obtain automated locations in real-time with preliminary magnitude characterization. In broad terms, the approach is 1) detect events using AI-enhanced workflows, 2) obtain phase picks and locate the events and 3) detailed investigation of event distributions (locations, errors, and magnitudes) to highlight outlying events for prioritizing manual review. Much has been discussed on different aspects of each of these points, and rather than add to that volume of literature, we focus on how we've benchmarked the assemblage of (mostly proprietary) algorithms. In this way we hope to provoke a discussion on what constitutes an acceptable standard for comparison of automated data processing workflows.

Comparing Different Analyses

Manual processing of events usually serves as the comparison against which automated processing is compared. This is not quite perfect, as there is error inherent in triggering events and picking even by human analysts. In figure 1 below, we show the results of how events are analyzed against a manual catalog of events that is taken as "ground truth" when comparing two generations of automated processing ("old" and "new"). This site is anonymized but is generally triggering with a very high sensitivity level to ensure every possible event can be captured. The left subfigure compares the number of false positives and shows that this number drops from over 8000 to around 1400. In other words, over 6000 triggers can be eliminated from the QC load of the analysts by adopting better automated processing routines.

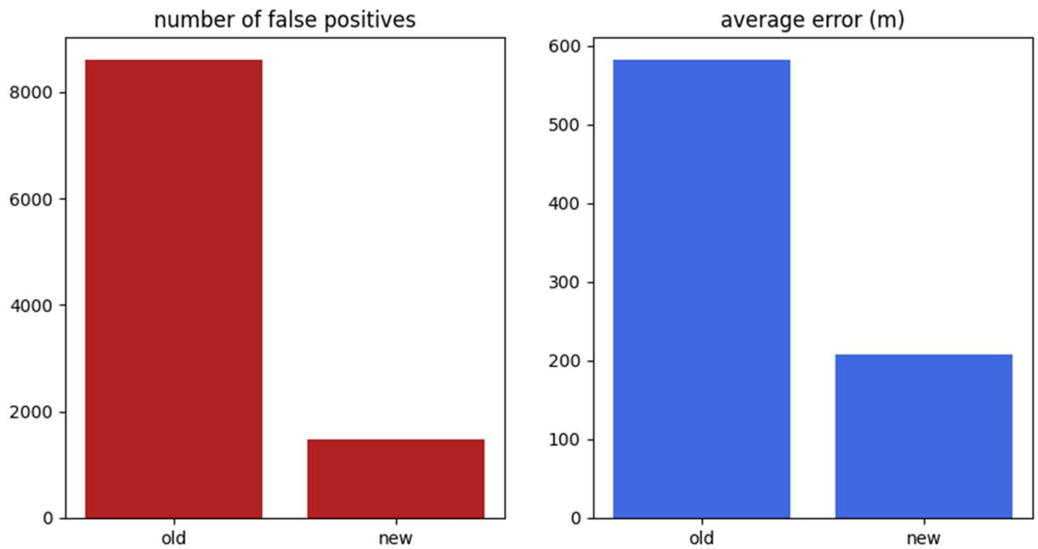

Figure 1. A comparison of different generations of automated processing from (left) the numbers of false positives to (right) the Euclidean distance (error) between the automated solutions and the manual locations.

On the right hand side of figure 1, we show how average error for the events behaves from the first to second generation of the events. We take this average error as a proxy for pick consistency: the events in general have around a 3 times lower error dimension than the previous generation, indicating indirectly that the consistency of the picking is much greater. Here, error is defined as the Euclidean distance between the manual events and the automatically processed.

Discussion

Advances in AI and machine learning allow for strong improvements in automated processing of seismicity. The quantification of such improvements though can be challenging, especially on the granular level of comparing one set of picks against another set of manual picks complete with the imperfections of human subjectivity. In this paper we show how using basic aggregated statistical distributions of event parameters and false positive rates can show how these improvements can be objectively quantified.