Automatic Association of Earthquake Arrivals with Pattern Detection and Graph Theory based Approach

Ian W. McBrearty, Andrew A. Delorey, Paul A. Johnson
Los Alamos National Laboratory - Geophysics Group

Abstract

The association of earthquake arrivals becomes non-trivial as arrival detection methods become more sensitive, and earthquake rates are high. In challenging cases, earthquake arrivals across a seismic network from several sources may overlap in time, false arrivals may be detected, and all arrivals may be of unknown phase (e.g., P/S). We propose an explicit method to associate and locate earthquake arrivals applicable to such situations. To do so, we use a pattern detection metric based on the principle of reverse-time migration to reveal "candidate" sources, followed by a clustering and linear integer optimization routine to determine the associations and the minimum number of sources necessary to explain the data. We apply our technique to Northern Chile over 2007-2017, increasing the number of earthquake detections compared with the CSN catalog by $\sim$ 1.5 million, lowering the magnitude of completion from $\sim$ 2.75 to $\sim$ 1.65.

Introduction

Many seismic techniques have been developed to improve the detectability of earthquake arrivals from continuous time-series [1, 2, 3, 4]. However in practical cases, after detections have been made, determining the actual associations across a network and the correct number of sources producing the arrivals can still be ambiguous. We propose a solution using first-principles, based on the idea of reverse-time migration (i.e. backprojection). We apply our method to real data, and design it to handle:

- **Unknown number of sources in any time window**
- **Possible false arrivals**
- **Uncertain phase arrivals**
- **Uncertainties in velocity model**

The Association Problem and Reverse-Time Migration (RTM)

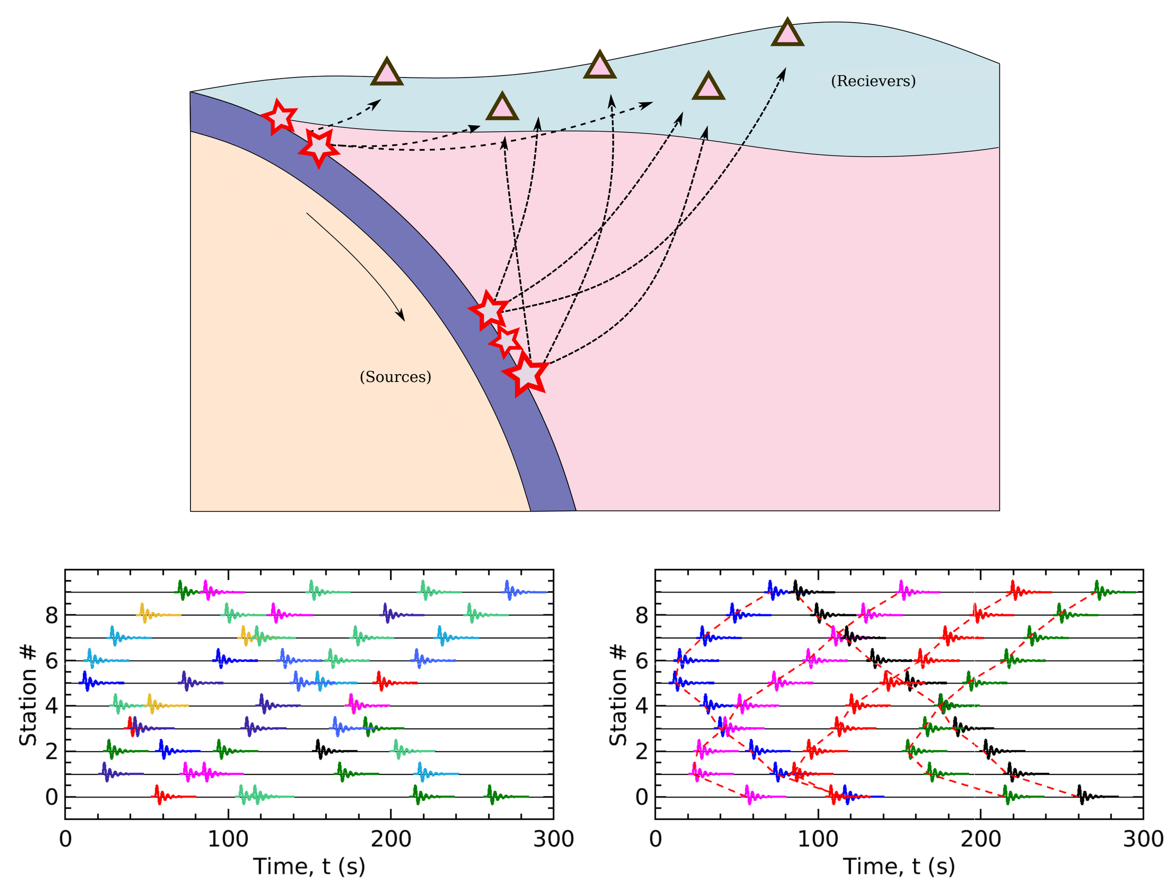


Figure 1: Schematic of multiple sources producing a complicated set of arrivals (top). Observed *unassociated* set of arrivals (bl). True association of arrivals - the target "solution" (br)

The principle of RTM is that when arrivals are "played" backwards in time into the Earth they should constructively interfere at the original source coordinate (and source origin time).

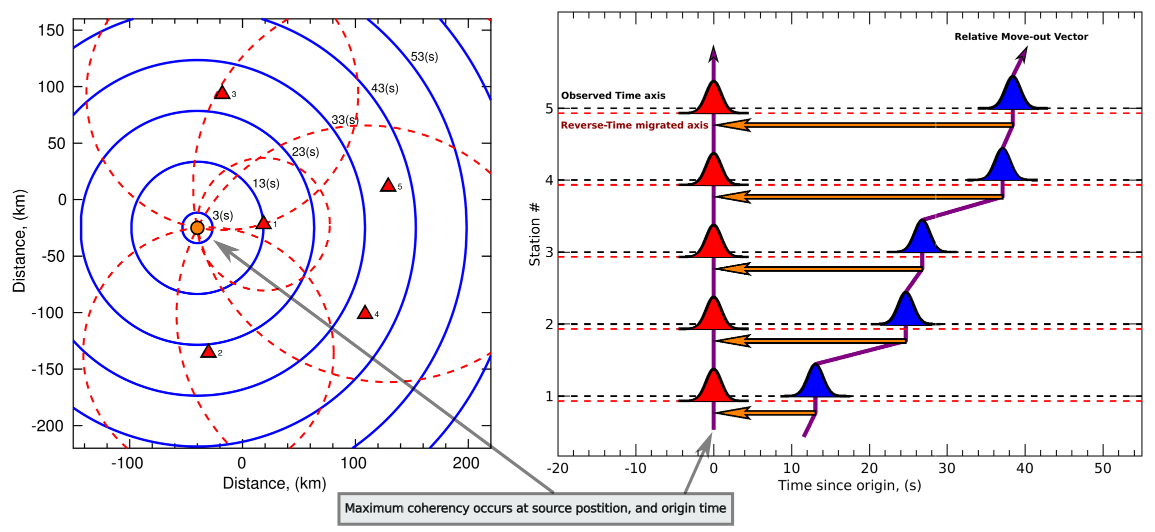


Figure 2: The forward (blue solid line) and "backward" (red dashed line) wavefields for one earthquake (left). The equivalent time migration procedure in the time domain. Arrivals (blue), Migrated waveforms (red), relative "move-out" vector (purple line). (right)

Method (Pattern Detection)

Our technique uses the principle of RTM to detect when known move-out vectors from real source locations occur in the time-domain. Rather than "backproject" to all points in space, we assemble a sparse set of "move-out vectors" which span all possible move-out vectors (in the sense of "nearest neighbor" distance), and run a continuous-time pattern detection routine to determine when/if any template move-out vector occurs. The "template" move-out vectors are obtained from the Self-Organizing Map algorithm [5].

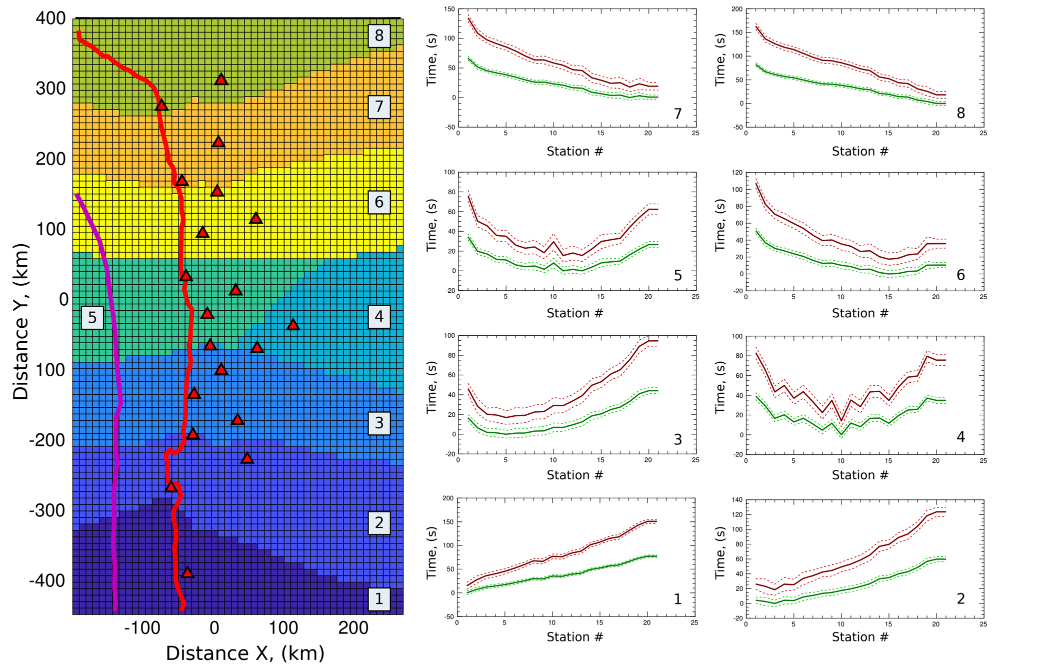


Figure 3: Map view of IPOC network (Northern Chile) and surrounding regions partitioned by nearest move-out vector template. Chilean coast (red line), trench (purple line). (left). Corresponding move-out templates of P (green), S (red) waves, with +/- 0.5 sigma of all real move-out vectors assigned to each template (right)

We compute continuous time "fitness" (i.e. likelihood of occurrence) of all templates with the following metric:

$$C_k(t) = \sum_{i=1}^{n}\Big[\sum_{j=1}^{m(i)}\sum_{r=1}^{s}\frac{1}{s}exp\Big(\frac{-(t-(\tau_{ij}-w_k^r(i)))^2}{2(\sigma^2+f(\tau_{ij},w_k^r(i))^2)}\Big)\Big]^{\perp} \tag{1}$$

for τ_{ij} arrivals indicating j^{th} arrival of i^{th} station, and $w_k^r(i)$ being the i^{th} station entry of k^{th} template with phase r. f denotes an uncertainty term proportional to arrival uncertainty and travel time uncertainty, and σ is the intrinsic "resolution" of the scan.

Method (Clustering and Linear Integer Optimization)

All times where $C_k(t)$ exceeds a threshold (γ) are treated as candidate source times. However, backprojection "coherency" is a necessary, but not sufficient condition for being a true source. We reduce the over-complete set of candidate sources by clustering similar sources, and then running a constrained linear integer optimization problem which enforces several important constraints, while maximizing the ability to "explain" the most arrivals with the least number of sources necessary. To achieve these goals, we construct an undirected graph g with edge weights defined by all (candidate source)-(arrival) pair *connectivity* values. Connectivity values are directly obtained from the interior terms in equation (1) for all candidate source times picked in $C_k(t)$. Spectral Clustering is used to split the large (sparse) adjacency matrix of g into sub-blocks containing sources of strong "similarity". Sub-matrices are then passed to the "Competitive Assignment" step, which yields the final result.

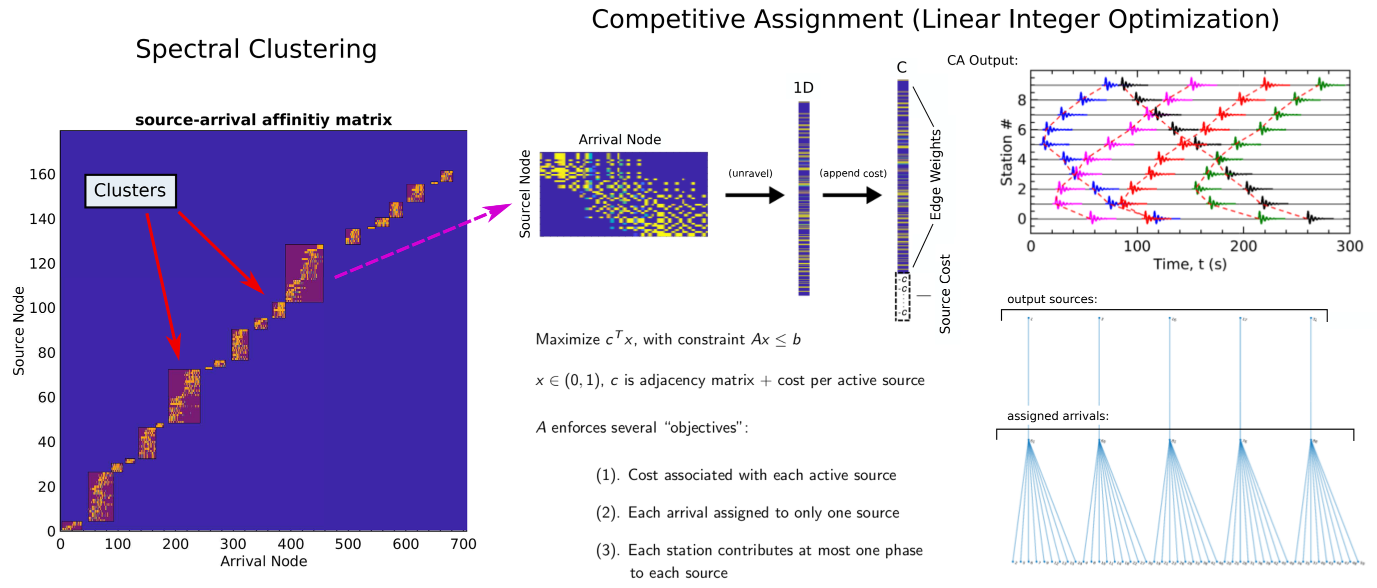


Figure 4: An example of the source-arrival adjacency matrix, and its characteristic block diagonal structure. Red rectangles mark regions grouped by spectral clustering (left). A schematic of the "competitive assignment" method to determine optimal source-arrival assignments, and the resulting output graph, which satisfies listed constraints (right)

Application and Results

We apply our method to Northern Chile seismicity recorded on the IPOC network between 2007 - 2017. Our objective in this study was to reduce the magnitude of completion and improve our understanding of seismicity in this region. We apply our method to a large set of (unknown) phase arrival time measurements obtained with our Network Coherency Method (NCM) picking algorithm [2]. We use $\sim$ 800 template move-out vectors (of P and S waves), and set $sig = 3\ sec$, $\gamma = 4.5$, $source\ cost = 5$. We detect and catalog on average 500 events a day, resulting in $\sim$ 1.8 million detections over this time interval. We estimate a magnitude of completion of $\sim$ 1.65 (Figure 6).

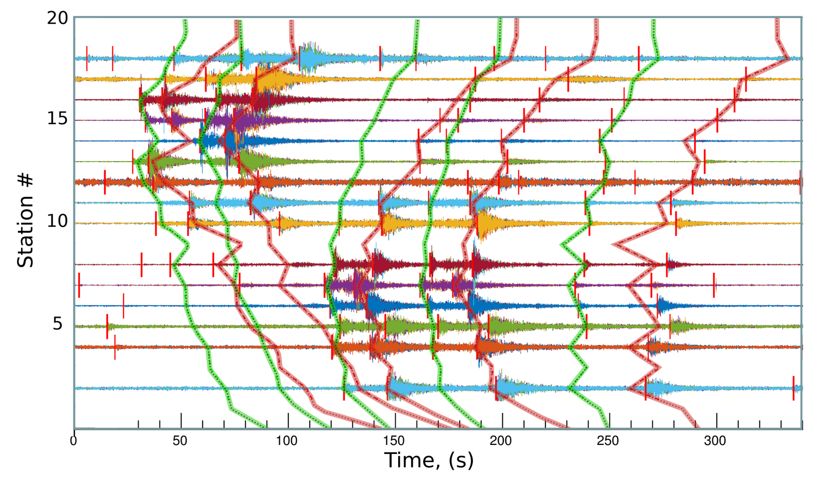
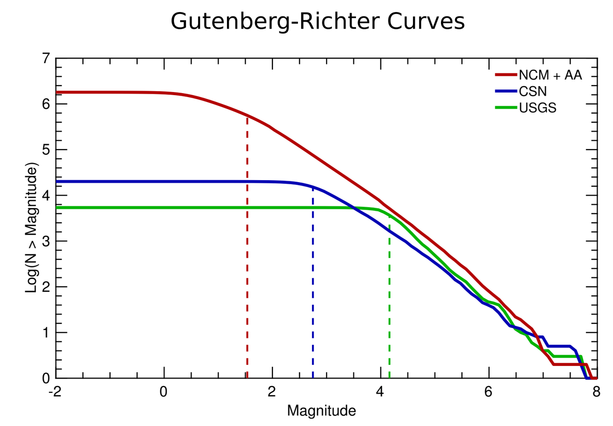


Figure 5: An example set of detections on a sequence of five earthquakes. Predicted arrival phases marked P wave (green), S wave (red). Vertical red bars denote arrival time picks.



Catalog Comparisons

	(%) theirs detected in our catalog	(%) ours detected in their catalog
USGS (2007/01 - 2017/4)	98.4	0.35
CSN (2010/6 - 2016/6)	99.4	1.27
Bloch et al., (2014) (2010/6 - 2012/2)	88.5	0.77

- we note catalog comparisons are preliminary; further analysis must be done to verify these results

Figure 7: Gutenberg-Richter plots for three catalogs (USGS, CSN, and our own, NCM + AA), are shown. Approximate magnitudes of completion denoted by dashed vertical lines (left). Percentage of pair-wise matching events between our catalog and several others are given (right)

Conclusions

We have developed an automated technique to process arrival time data and determine the optimum number of sources, phase assignments, and associations across a network. We have found our technique to be effective on synthetic and real data cases, and believe it may complement existing arrival detection algorithms. Several aspects of the method are designed to be computationally efficient, and robust to various types of noise and uncertainty, making it a practical algorithm for general use. Recently, we have applied this method to Northern Chile, in an attempt to gain more insight into seismic processes occurring at this highly active subduction zone. Future work will assess what we observe in the catalog, and validate our detections with more scrutiny.

References

[1] Steven J Gibbons and Frode Ringdal.
The detection of low magnitude seismic events using array-based waveform correlation.
Geophysical Journal International, 165(1):149–166, 2006.

[2] Andrew A Delorey, Kevin Chao, Kazushige Obara, and Paul A Johnson.
Cascading elastic perturbation in japan due to the 2012 mw 8.6 indian ocean earthquake.
Science advances, 1(9):e1500468, 2015.

[3] Clara E Yoon, Ossian OâĂŹReilly, Karianne J Bergen, and Gregory C Beroza.
Earthquake detection through computationally efficient similarity search.
Science advances, 1(11):e1501057, 2015.

[4] Thibaut Perol, Michaël Gharbi, and Marine Denolle.
Convolutional neural network for earthquake detection and location.
Science Advances, 4(2):e1700578, 2018.

[5] Teuvo Kohonen.
The self-organizing map.
Neurocomputing, 21(1-3):1–6, 1998.

[6] W Bloch, J Kummerow, P Salazar, P Wigger, and SA Shapiro.
High-resolution image of the north chilean subduction zone: Seismicity, reflectivity and fluids.
Geophysical Journal International, 197(3):1744–1749, 2014.

Acknowledgements

This work was supported by institutional funding from Los Alamos National Laboratory. The authors acknowledge the important contributions of Dr. Joan Gomberg (USGS, Research Geophysicist) to the development of the method and its application to Chile.

Contact Information

- Email: imcbrearty@lanl.gov

LA-UR-18-20709