Accurate earthquake locations are essential for providing reliable hazard assessments, understanding the physical mechanisms driving extended earthquake sequences, and interpreting fault structure. Techniques based on waveform cross-correlation can significantly improve the precision of the relative locations of event pairs observed at a set of common stations. Here we describe GrowClust, a relative event relocation algorithm that can provide reliable relocation results for earthquake sequences over a wide range of spatial and temporal scales. The GrowClust method uses input differential travel times, cross-correlation values, and reference starting locations, and applies a hybrid, hierarchical clustering algorithm to simultaneously group and relocate events within similar event clusters. Requiring no explicit matrix inversion, the method is computationally efficient and capable of handling large data sets, and naturally applies greater weight to more similar event pairs. Additionally, GrowClust outputs robust location error estimates that are useful in interpreting the reliability and resolution of relocation results. As an example, we apply the GrowClust method to two prominent, recent sequences in western Nevada: the Spanish Springs and Sheldon earthquake swarms. The sequence migrates linearly along strike towards M4.2 mainshock. Mainshock rupture zone is nearly devoid of seismicity, with aftershocks concentrated along its perimeter. Aftershock sequence is extended in duration and encompasses spatially distinct fault sections.

Key features of the GrowClust program:
- stable, computationally efficient algorithm (implementation in Fortran90)
- intuitive data weighting scheme that is robust to data outliers (L1-norm optimization)
- multi-scale capability to handle both large and small cross-correlation data sets (no matrix inversion)
- nonparametric internal processes that automatically estimate location uncertainties
- flexible and user-friendly input/output and program interaction
- initial release will be fully open-source, with user feedback driving future improvements

Hybrid Clustering/Relocation Algorithm:
1. Each event starts in its own distinct starting cluster.
2. For each event pair, compute a waveform similarity coefficient that measures the quantity and relative strengths of cross-correlation data linking this pair.
3. Sort event pairs by similarity coefficient and process each in turn, starting with the most similar. Processed event pairs fall into one of three classes:
   a. Both events are members of distinct, single event clusters: merge into a single cluster and relocate with respect to centroid (L1-norm, grid search approach).
   b. Either event is a member of a distinct, multi-event cluster: test relocation of clusters with respect to one another. Accept relocation results and merge into a single cluster (if warranted by the data and algorithm control parameters).
   c. Both events are members of same cluster: skip pair, already relocated.
4. Continue until all event pairs are processed. Save relocation results for output, compute final data residuals, and proceed to nonparametric uncertainty analysis.
5. Iterated bootstrap resampling of input cross-correlation data: generates bootstrap distribution of event locations. Event uncertainties are obtained from MAD of bootstrap distribution (more robust and stable nonparametric error estimate than bootstrap standard errors).

GrowClust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish Springs and Sheldon, Nevada, earthquake sequences

Daniel T. Trugman1, Peter M. Shearer1, Kenneth D. Smith2
1. Scripps Institution of Oceanography, UC San Diego, La Jolla, CA
2. Nevada Seismological Laboratory, U. Nevada - Reno, Reno, NV

Abstract
Earthquake location is the most fundamental task in seismology. However, the current methods employed through routine processing of phase arrival data are often highly scattered and imprecise. GrowClust is an open-source computer program that performs relative earthquake location using waveform cross-correlation data (differential times and cross-correlation values) for pairs of events observed at a set of common seismic stations. Building on previous work in this field1,2, the GrowClust method is both computationally efficient and multi-scale in its capacity to handle sets large and small. GrowClust’s relocation algorithm uses a hybrid, hierarchical clustering technique to simultaneously group events into clusters based on waveform similarity and relocate each event with respect to its cluster centroid. This relocation approach requires no explicit matrix inversion, allowing for improved robustness to data outliers through optimization against the L1-norm of the differential travel-time residuals. The GrowClust program implements a nonparametric resampling technique to estimate location uncertainties, allowing users to directly examine the resolving power of the output relocation results.

Sequence #1: 2012-2014 Spanish Springs, NV
The 2012–2014 Spanish Springs, NV sequence contained more than 1600 events occurring over a compact spatial scale of several kilometers. The sequence was well-recorded by near-source stations operated by the NSL, and its M4.2 mainshock was widely felt in the Reno area.
- Seismicity illuminates distinct faulting structures with near-vertical dip (consistent with the mainshock focal mechanism).
- Vigorous foreshock sequence migrates linearly along strike towards M4.2 mainshock.
- Mainshock rupture zone is nearly devoid of seismicity, with aftershocks concentrated along its perimeter.
- Aftershock sequence is extended in duration and encompasses spatially distinct fault sections.

Sequence #2: 2012–present Sheldon, NV
The ongoing Sheldon, NV sequence occurred within the remote northeastern corner of the state, where station coverage is sparse and hence the initial catalog locations are poorly constrained. The sequence is of particular scientific interest due to its persistent, swarm-like seismicity and large moment release (26 M4+ and 261 M3+ events through July 2016, with the largest being M4.8).
- Seismicity is concentrated in two subparallel clusters.
- Both clusters contain two distinct fault strands.
- Both clusters occur on east-dipping fault structures (consistent with regional moment tensor analyses).
- Distinctive patterns of spatial migration along strike during the extended, swarm-like sequence.
- Analysis of GrowClust relocation uncertainties indicate that the major western cluster is better resolved than its minor, eastern counterpart.

Summary
- GrowClust is a new and open-source algorithm that uses waveform cross-correlation data to perform precise relocations of earthquake hypocenters.
- The GrowClust method is fast, flexible, robust to data outliers, and provides location uncertainty estimates that aid in the interpretation of relocation results.
- Relocation results for the recent Spanish Springs and Sheldon, NV sequences demonstrate the method’s viability in illuminating key features of the seismicity that are obscured by catalog location scatter.
- Future implementations will allow for travel-time computations through 3D structure and parallelize the uncertainty analysis subroutines for maximum algorithm efficiency.

References