COLLABORATIVE RESEARCH: SEISMIC CATALOGUE COMPLETENESS AND ACCURACY

Terry C. Wallace,[1] Frank Vernon,[2] and Gary Pavlis[3]

University of Arizona, [1] University of California San Diego,[2] Indiana University[3]

Sponsored by Defense Threat Reduction Agency

Contract No. DTRA01-00-C-0069,[1] Contract No. DTRA01-00-C-0062,[2]
and Contract No. DTRA01-00-C-0095[3]

ABSTRACT

The basic problem in assessing catalogues is assigning errors to locations and separating errors introduced by Earth structure versus the ensemble of stations chosen for the location. We are constructing a catalogue for the region stretching from Saudi Arabia to western China for 1995 to the present. We will use all available data sources, including several temporary, portable seismic experiments and private or national seismic networks that are not easily available to the researcher. The data sources include GSN, KNET, Geoscope, and Kaznet permanent networks; PASSCAL deployments in the Tien Shan, Pakistan, India, Saudi Arabia; and national networks from Oman, Jordan and Iran. The total number of stations with continuous broadband data in the area bounded by 10^o to 60^o N and 30^o to 110^o E is in excess of 130. Although some of the data streams are intermittent, or only deployed for part of the window of time in which we wish to develop the catalogue, we can estimate the number of events expected on the basis of the performance of KNET and the Saudi experiment is in excess of 25,000. The fact that high-quality seismic stations are located near the seismic activity makes the catalogue superior to any other product for the same region.

We have developed a new set of computer codes to process a database of arrival time data to produce improved estimates of two fundamentally different entities: (1) a set of three-dimensional (3-D) grids of travel-time corrections for each seismic station defined in the database, and (2) a set of improved location estimates for each seismic event that is consistent with the 3-D travel-time surfaces computed in the same procedure. The procedure we use has three steps: (1) cluster association, (2) location and simultaneous estimation of travel-time surfaces, and (3) error estimation.

We have looked at events of special interest. In central Asia a perfect set of test events is the “Omega” explosions conducted as part of the Kazakhstan-American calibration experiment. In addition to the International Monitoring System (IMS) stations, at least some of the Omega events were recorded by 19 additional stations at regional distances. We have conducted a series of experiments to (a) determine optimal station coverage to produce ground truth GT-5, (b) develop empirical procedures for optimizing the phase information from regional records, and (3) test several joint hypocenter determination scenarios. Our tests highlight the importance of calibrated stations and the benefits of using all available data.

OBJECTIVE

Seismic event detection and location are the single most important research issues for adequately monitoring a Comprehensive Nuclear-Test-Ban Treaty (CTBT). Confidence in a CTBT relies on the assumption that any underground nuclear weapons test will be detected at an adequate number of seismic stations such that it can be located with sufficient accuracy that the test could be confirmed with an on-sight inspection. Another way to put this is that confidence in seismically monitoring a CTBT is equivalent to confidence in the *seismic catalogue* that is produced by the monitoring system. There are two principle components of a catalogue: (1) completeness in terms of representing all the seismicity within the region of monitoring interest (referred to here as *detection*), and (2) the accuracy of the source parameters for the events within the catalogue.

The present IMS has a low detection threshold and presumably high location accuracy in the Baltic shield region of Europe. This is due to the presence of various NORSAR arrays, which dramatically enhance the regional detection level. Other areas, in particular the southern hemisphere and Asia, have much higher detection thresholds. The quality of the locations produced by the prototype IMS has been assessed by comparing it to other catalogs of seismicity, in particular the NEIC's PDE. However, this comparison can be problematic because the PDE is not complete and suffers from sparse station distribution. Further, the locations in the PDE are not vetted against ground truth nor corrected for known heterogeneity in earth structure. A better gauge of the quality of the IMS catalogue is to measure it against a detailed local catalogue. However, this is a difficult task because individual catalogues are constructed with unique methodology, and thus are of uneven quality.

We are constructing a catalogue for the region stretching from Saudi Arabia to western China for 1995 to the present. We will use all available data sources which includes several temporary, portable seismic experiments and private or national seismic networks which are not easily available to the researcher. The catalogue construction will have two principle tasks: detecting and associating seismic phase information (the number of expected events within the time frame is in excess of 25,000), and locating these events and assigning realistic location errors. There are several fundamental research tasks that need to be performed before the catalogue can be completed. These include research on locating events which are primarily recorded at regional distances, evaluating both formal and systematic errors, and assessing "true" detection capabilities. Once the catalogue is constructed, it will provide an excellent test bed for research on strategies for spotlighting regions of enhanced monitoring interest, provide the data base for detailed regional velocity models (which can be incorporated into a general location scheme), and most importantly, assess the quality of catalogues routinely produced for monitoring purposes.

RESEARCH ACCOMPLISHED

Catalogue Construction and Accuracy

Overview. The geographic region which stretches from the Middle East to central Asia is an area of monitoring interest. It is also a region in which there have been a number of portable seismic experiments mounted in the last few years due to the interest in the structure and dynamics of the India-Asia collision. The total number of stations with continuous broadband data in the area bounded by 10^{o} to 60^{o} N and 30^{o} to 110^{o} E is in excess of 130. Although some of the data streams are intermittent or only deployed for part of the window of time in which we wish to develop the catalogue, we can estimate the number of events expected on the basis of the performance of KNET and the Saudi experiment is in excess of 25,000. The fact that high quality seismic stations are located near the seisimicity will make the catalogue superior to any other product for the same region.

Comparisons with other catalogues. The most rigorous test of catalogue completeness is to compare a local or regional earthquake list with a global catalogue that is produced with a standard technique. We have constructed a catalogue using KNET, KZNET, and the PASSCAL experiment GHENGIS, for a three year period and compared that to the PDE. There are nearly an order of magnitude more events in the local/regional catalogue, and the magnitude range is extended nearly 2 full unites (magnitude 4.0 for PDE compared to 2.0 for the regional/local catalogue).

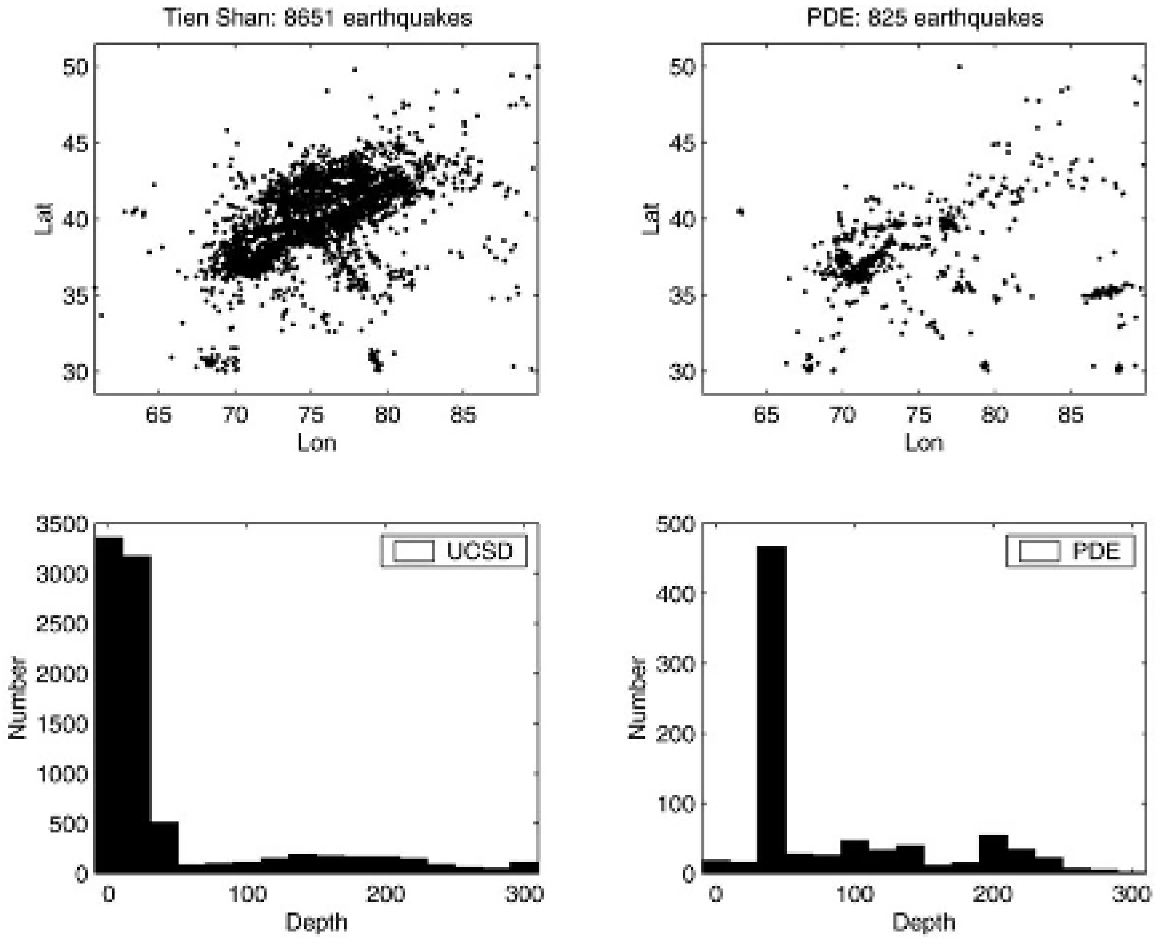

Figure 1: Comparison of regional/local catalogue to PDE

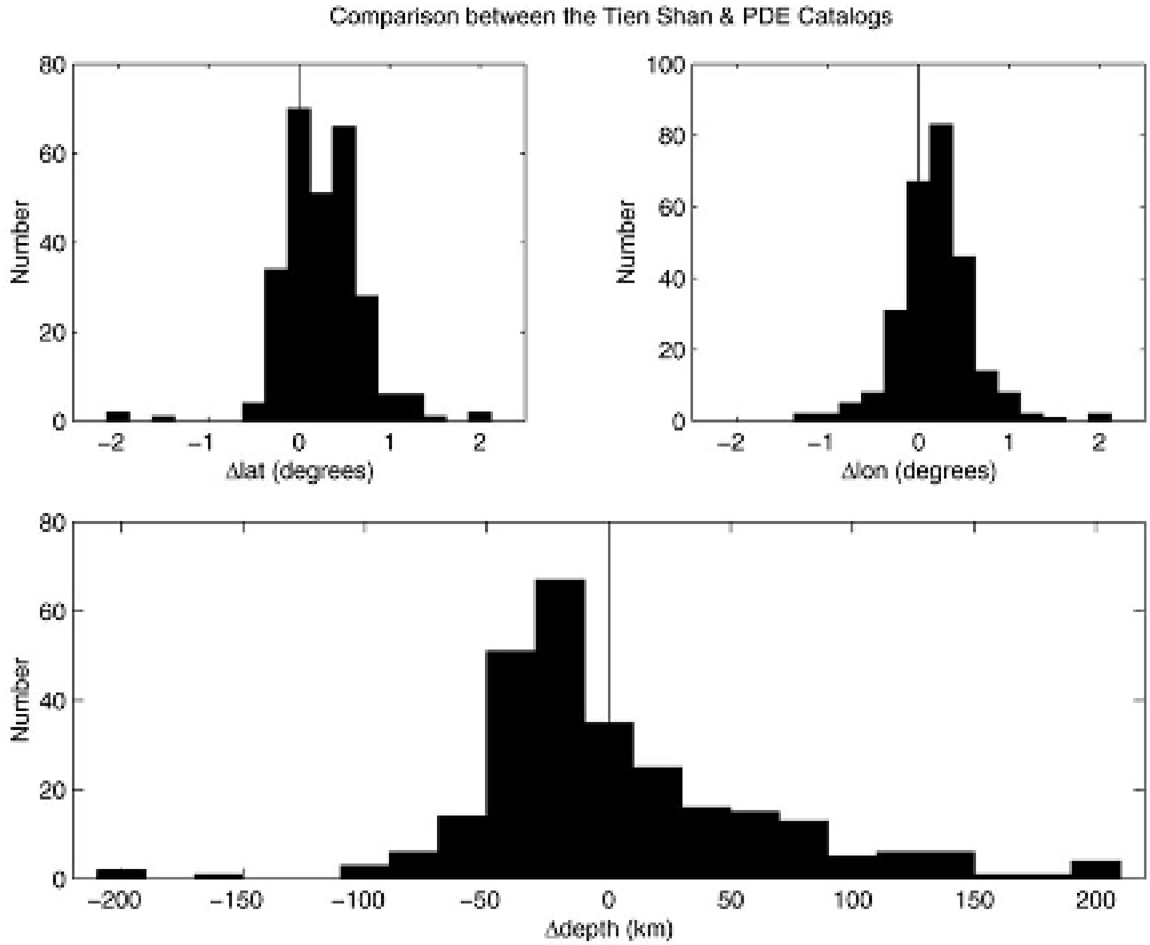

Figure 2: The differences between the local/regional catalogue parametrics and the PDE. If the locations are identidcal, then it will plot as a zero. The depth histogram has a large spike at –33 km associated subcrustal depth assigned to poorly located events in the PDE.

Event Location Methodology Development

Overview. We have developed a new set of computer codes to process a database of arrival time data to produce improved estimates of two fundamentally different entities: (1) a set of 3D grids of travel time corrections for each seismic station defined in the database, and (2) a set of improved location estimates for each seismic event that are consistent with the 3D travel time surfaces computed in the same procedure. In this section we describe this new algorithm and how it compares to other strategies that have been suggested for precision earthquake relocation.

Event cluster association. Our objective is to estimate a set of travel time corrections relative to a radially symmetric, reference earth model and to compute these corrections empirically in a 3D volume within the earth. The data we use for this purpose is arrival times of N_p seismic phases recorded at the set of N_s stations for which we have data. Consequently, the complete set of data objects we aim to produce is a set of N_s estimates of a 3D vector field (N_p travel time corrections per grid point) defined on some type of discrete grid. We chose to use a natural grid scheme for regional scale earth science applications that we will refer to as a geographical curvilinear grid (GCLgrid). A GCLgrid is a simple extension of what is commonly called a uniform field in the scientific visualization world. A uniform grid divides a box shaped region into a set of constant sized boxes. A GCLgrid makes the closest equivalent approximation in a spherical earth. That is, we divide a region of the earth into a series of spherical shells of equal thickness and then subdivide the shells into sub areas of approximately equal solid angles. To improve the uniformity of the grid at regional scales we translate the intersection of the prime meridian and equator to a specified origin, define a baseline using a great circle path at a specified azimuth, and then define an equal angle grid of latitude and longitudes relative to this baseline and origin. This produces continent scale grids that have properties that do not depend upon the actual location of the grid. Simpler schemes based on the normal geographic reference, for example, commonly have serious distortions at high latitudes that are avoided by this method.

Once we define the geometry of a reference GCLgrid we process a list of associated events within a CSS3.0 database and link each event to one or more grid points using a database relation that is an extension of CSS3.0. The best recipe for assigning an event to a given point in space is an open question, but in the current implementation we use a simple set of rules based on distance. That is, we define a cylindrical region around each target grid point. The radius of this cylinder is initially set to a specified minimum size. We count the number of events within the specified volume. If a specified minimum number of events are not present within this volume the sphere is expanded incrementally until either the hit count exceeds the minimum or the radius exceeds a specified maximum. In the later case no links to the grid point are stored in the database. Otherwise we compute the hypocentroid (center of mass) of the ensemble of events and store the hypocentroid and indices that connect it to all events that form that ensemble.

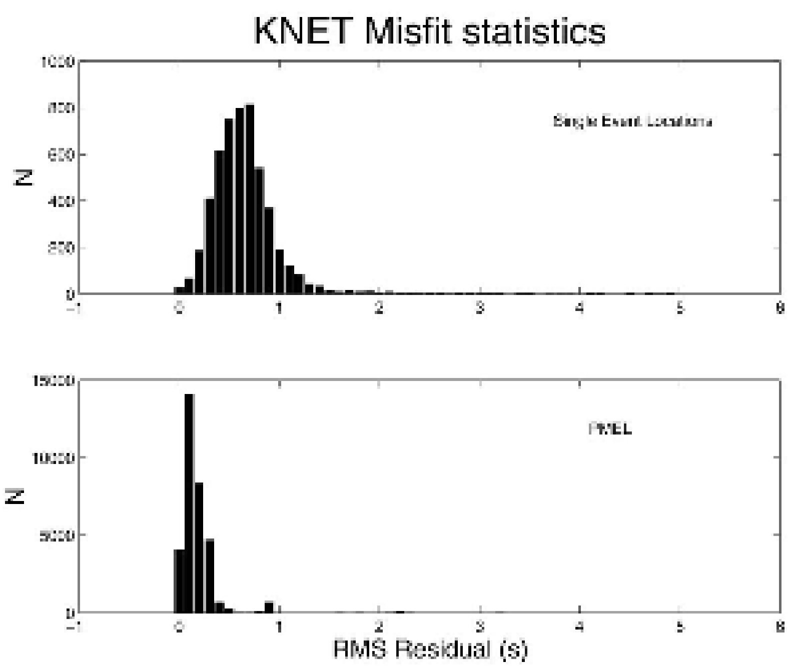

Figure 3: Misfits using PMEL vs. single event locations.

This simple approach is most appropriate for processing of existing catalog data for which waveforms may not be readily available. Current work is focused on utilizing waveform data as part of the clustering problem. That is, it is now widely known that sources that are closely located in space produce similar waveforms at common stations (reciprocal arrays). We are working on correlation methods based on a form of wavelet processing that builds on recent papers by Lilly and Park (1995), Bear and Pavlis (1997, 1999), and Bear et al. (1999). We use a set of special functions that Lilly and Park (1995) termed multiwavelets due to their close kinship to the Slepian tapers used in multitaper spectral analysis (Thomson, 1982). Multiwavelets provide a mechanism to produce a cross-correlation method that is a hybrid of Fourier and time-domain methods. We have successfully developed a computer program to use this technique for phased array processing of data from three-component broadband arrays with apertures of greater than 100 km. This program allows us to simultaneously estimate the following observables and to place objective error estimates on each measurement: (1) slowness vector of best fit plane wave, (2) travel time residuals at each station relative to the best fit plane wave, (3) particle motion estimates for each station, and (4) average particle motions for the entire array. These parameters are measured in multiple frequency bands and can averaged or examined independently to study their frequency dependence.

We are in the process of applying this technique to teleseismic and near-regional events recorded by the Kyrghyz network (KNET). We have found that with this approach we are able to measure P wave arrival times to subsample accuracy (timing uncertainty of the order of 0.01 s for teleseismic P waves recorded at 40 sps) and array slowness vectors to better than 1 part in 10^4. This may provide a superior way to utilize earthquake monitoring networks like KNET for regional location estimation. That is, it may be preferable to treat dense collections of broadband stations as phased arrays instead of utilizing individual arrival times in location estimation.

We are in the process of adapting the receiver array processing program to source arrays. That is, an ensemble of events located closely together in space observed at a single station commonly have coherent waveforms that can be processed by the same waveform multiwavelet correlation methods. There are two primary differences from receiver array processing and source array processing: (1) the key observables are different, we do not have to deal with slowness vectors but focus strictly on estimating precision time differences; and (2) the amplitudes from different sources can vary by orders of magnitude. The former makes the processing simpler than receiver array processing and the later is a major complication. For the later we are experimenting with a weighting scheme that measures coherence of each seismogram relative to the stack and downweights incoherent traces in an iterative procedure.

Location methodology. The approach we use for simultaneous estimation of improved earthquake locations and the ensemble of path corrections is the Separated Earthquake Location Method (SELM) described originally by Pavlis and Hokanson (1985). SELM is an extension of the Progressive Multiple Event Location (PMEL) method of Pavlis and Booker (1983). PMEL/SELM are multiple event location methods that are extensions of the classical Joint Hypocentral Determination (JHD) techniques (Douglas, 1967). The primary distinction between PMEL and JHD is that JHD solves the system of equations for an ensemble as a single, large matrix while PMEL separates the nonlinear location parameters from the linear, station correction terms. In our experience PMEL improves stability through iterative updates of earthquake hypocenter coordinates and through robustness introduced by automated deletions of events that have statistically significant differences in rms misfit compared to the ensemble average. This is not possible with JHD where the whole system of equations is inverted simultaneously.

SELM adds an important second feature that we exploit. Pavlis and Hokanson (1985) show how to construct a pair of complementary, orthogonal projectors, which we will refer to as $\boldsymbol{P}_R$ and $\boldsymbol{P}_N$. We apply these projectors using the relation

$$\boldsymbol{s} = \boldsymbol{P_R}\, \boldsymbol{s}_{3d} + \boldsymbol{P}_N\, \boldsymbol{s}_{data}$$

where $\boldsymbol{s}$ denotes a N_sN_p vector of “path anomalies. That is, each component of $\boldsymbol{s}$, is a difference, $s_i=(t_{3d})_i - (t_{ref})_i$, between the travel time based on three-dimensional earth model and some 1-D reference model.

In the current implementation, which we will henceforth refer to as *dbpmel*, an ensemble of events and their related arrival times are associated with target points in space (GCLgrid structure) using the clustering algorithm described above. The actual algorithm uses a relational database table analogous to assoc that links event ids to grid points. The generalized PMEL/SELM program is then applied to the ensemble of events assigned to each target point in space. For points with sufficient data we compute the solution as described above. For points with insufficient data our method guarantees that our estimate of $\boldsymbol{s}=\boldsymbol{s}_{3d}$. That is, the computed path anomalies automatically revert to those computed from the reference 3D model when no data is available, but otherwise use all available data to refine the estimated path anomalies. The end result is a set of path anomaly estimates (defined on a GCLgrid in space) derived empirically as a best fit to all the available data.

The vectors we call $\boldsymbol{s}_{3d}$ and $\boldsymbol{s}_{data}$ use a different realization of 3D earth structure. Jordan and Sverdrup (1981) introduced the concept called the “hypocentroidal decomposition theorem. That is, for an ensemble of events that are localized in space the absolute location of the group is fundamentally indeterminate but precise estimates of $\boldsymbol{s}$ can dramatically reduce the scatter in the relative location of events within the group. The J&S method resolves this ambiguity by constructing a projector that annihilates the dependency of the solution on $\boldsymbol{s}$. Furthermore, recent work by Wolfe (2002) shows that the double difference technique (Waldhauser and Ellsworth, 2000) is actually equivalent to applying the J&S method to a weighted set of equations with the weights dependent on the number of pairs of events differenced. That is, both double difference and the J&S method explicitly or implicitly use projectors designed to annihilate the dependence upon s. The projectors we construct take the opposite approach. In *dbpmel* the projectors are constructed to annihilate the dependence of residuals on the hypocentral parameters instead of $\boldsymbol{s}$. This allows us to utilize the term $\boldsymbol{P_R}\, \boldsymbol{s}_{3d}$ to add ancillary information to resolve the absolute location ambiguity. That is, we utilize a 3D model as a secondary reference model. The 3D model is used only to resolve the absolute location ambiguity. The data are used directly to estimate the term $\boldsymbol{P}_N\, \boldsymbol{s}_{data}$ using the PMEL algorithm. Because the $\boldsymbol{P}_R$ and $\boldsymbol{P}_N$ projectors are orthogonal the two procedures are decoupled within the limits of linearization. The method described in Jordan and Sverdrup's 1984 paper shares this decoupling feature but is incapable of applying the equivalent of $\boldsymbol{P_R}\, \boldsymbol{s}_{3d}$ exactly because it forces the solution to be independent of s. Double difference solutions can accomplish almost the same thing by computing all theoretical travel times in a 3D reference model. We would argue, however, that our approach has two advantages over a 3D implementation of double differencing. First, 3D travel time computation remains a major computing problem. In our implementation 3D model estimates need to only be computed once per grid point instead of for all travel time calculations. Given that results presented here on regional catalogs with the order of 10^4 events have run times of the order of cpu days, this is not a factor that can be dismissed. Secondly, Wolfe (2002) shows that a problem with applying double differences to regional catalogs is that the entire catalog gets linked together in a complex way that depends on the distribution of events and stations. This means that if new data become available a relocation with double differences would have to always start from all the available data to keep results internally consistent. Our PME/SELM implementation, in contrast, is localized through the gridding scheme. It could, in principle, be automated to update each grid point as a each new event is recorded. Such an approach would allow a scheme for continuous updates of empirical, 3D travel time grids as new data are acquired.

Improving Locations and Errors

A problem-free seismic network with excellent azimuthal coverage of the seismic area of interest and an accurate model of the velocity structure does not guarantee precise earthquake locations. This is because location errors can result from uncertainties in the time of the seismic wave's arrival or misidentification of the seismic phases. New algorithms, based on seismic waveform cross-correlation, can account for uncertainties in arrival times and in some cases correct for misidentified phases (e.g., Jordan & Sverdrup 1981; Pavlis, 1986; Got et al., 1994). These methods make use of the fact that earthquakes from the same region (< 3 km) who's seismic waves travel the same path from source-to-station have similar waveforms. By correlating similar waveforms, we can often obtain sub-sample accuracy of seismic wave arrival times. This, in turn, can reduce the relative errors in earthquake locations to tens or hundreds of meters, values that are often an order of magnitude smaller than the errors in the original catalogues (Kilb & Rubin, 2002).

We demonstrate the power of these relocation methods with a simple example using Tien Shan catalogue data of aftershocks from the August 27, 1998 M6.1 earthquake in the Xinjiang province of China. Here, we correlate a master event with all other events (Figure 4). A more sophisticated approach, such as that of Got et al. (1994), compares each event to every other event. Using Got's method we can identify clusters of similar waveforms from events that may be ~co-located but have different focal mechanisms. We are currently streamlining and updating the algorithms and computer codes of Got et al. 1994 to interface with our ANTELOPE formatted databases. Our aim is to confirm the accuracy of the final locations by comparing results derived from the PMEL algorithm with the results derived from the Got algorithms.

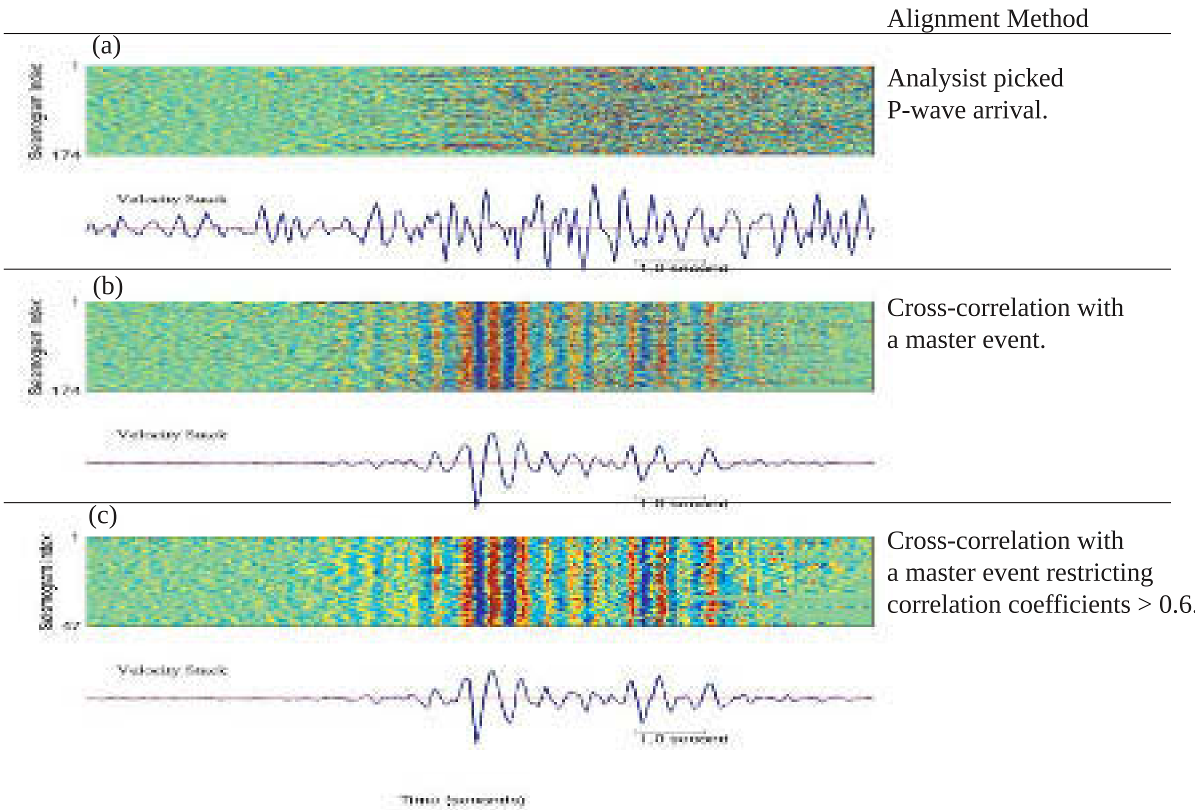

Figure 4: Simple example of waveform cross-correlation with a master event. (a) Vertical seismic waveforms from 174 earthquakes aligned on the analysts picked P-wave arrival (upward motion is red; downward motion is blue). As the noisy data stack shows, this alignment is not optimal. (b) After cross-correlation with a master event, the stack has a clear S-wave signature, and as many as 12-20 fringes are identifiable in the color map, indicating similarity in the seismic waveforms. (c) As in (b) but restricting all pair-wise correlation coefficients to be above 0.6. This limits the dataset to 47 earthquakes.

<u>CONCLUSIONS AND RECOMMENDATIONS</u>

We have constructed a catalogue for central Asia using regional networks and temporary PASSCAL deployments. The catalogue has 10 times as many events as the PDE for the same spatio-temporal footprint, and the difference in location scales with event size. Further, our catalogue is rich in small events that do not appear in other global catalogues such as the REB and ISC, and represents the power of regional monitoring with national resources or portable experiments. At present, the locations are from automated processing; in the next year we will relocate these events using the progressive multiple event methodology, which was also developed and enhanced during this contract year.

<u>ACKNOWLEDGEMENTS</u>

Although the Authors are the Principal Investigators on the project, a number of others are working on the various tasks. Deborah Kilb has been working on the catalogue completeness analysis and is using her experience on waveform cross-correlation to help with the travel errors. Tammy Baldwin has been working on the use of PMEL with sparse stations. Marie Renwald helped with the production of this paper.

REFERENCES

Bear, L. K., G. L. Pavlis, and G. H. R. Bokelmann (1999). Multiwavelet analysis of three-component seismic arrays: Application to measure effective anisotropy at Pinon Flats, California, *Bull. Seism. Soc. Amer.,* **89**, 693-705.

Bear, L. K. and G. L. Pavlis (1999). Multichannel estimation of time residuals from seismic data using multiwavelets, *Bull. Seism. Soc. Amer.*, **89,** 681-692.

Bear, L.K. and G.L. Pavlis (1997) Estimation of slowness vectors and their uncertainties using multiwavelet seismic array processing, *Bull. Seism. Soc. Amer*., **87**, 755-769.

Douglas, A. (1967). Joint epicentre location, *Nature*, **215**, 45-48.

Jordan, T. H. and K. S. Sverdrup (1981). Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, *Bull. Seism. Soc. Amer.,* **71**, 1105-1130.

Lilly, J. M. and J. Park (1995). Multiwavelet spectral and polarization analyses of seismic records, *Geophys. J. Int.*, **122**, 1001-1021.

Pavlis, G. L. (1992). Appraising relative earthquake location errors, *Bull. Seism. Soc. Amer.,* **82**, 836-859.

Pavlis, G. L. (1986). Appraising earthquake hypocenter location errors: a complete, practical approach for single event locations, *Bull. Seis. Soc. Amer.*, *76*, 1699-1717.

Pavlis, G. L. and J. R. Booker (1983). Progressive multiple event location (PMEL), *Bull. Seis. Soc. Amer.*, **73**, 1753-1777.

Pavlis, G. L. and N. B. Hokanson (1985). Separated earthquake location, *J. Geophys. Res.*, **90**, 12,777 12,789.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis, *Proc. IEEE,* **70**, 1055-1096.

Waldhauser, F. and W. L. Ellsworth (2001). A double-difference earthquake location algorithm; method and application to the northern Hayward fault, California, *Bull. Seism. Soc. Amer.,* **90**, 1353-1368.

Wolfe, C. J. (2002). A note on earthquake relocation using difference operators, *Bull. Seism. Soc. Amer.,* in press.