

**COLLABORATIVE RESEARCH: SEISMIC CATALOGUE COMPLETENESS
AND ACCURACY**

Frank Vernon¹, Gary Pavlis² and Terry Wallace³

¹IGPP, UC San Diego, ²Geology Dept., Indiana Univ., ³Dept. Geos., Univ. Arizona

Sponsored by Defense Threat Reduction Agency

Contract No. DTRA01-00-c-0069

ABSTRACT

The single most important goal of seismic monitoring of a nuclear weapons treaty is the detection and accurate location of seismic events of various origins (natural and man made) globally. However, the metric for evaluating the performance of the IMS or other monitoring networks is not easily established since global catalogues, such as the ISC and PDE, have been shown to be incomplete, with large regional variability. Further, the event parameters (location, origin time and magnitudes) are often quite different from catalogue to catalogue. To study these effects we are building a complete catalog of events from open data sources in the Middle East and Central Asia for the time period from 1991 through 2000. Data sources include the GSN, KNET, Geoscope, and Kaznet permanent networks, and data from temporary broadband deployments in Saudi Arabia, Turkmenistan, Pakistan, Tibet, and the Tien Shan. Current progress is centered on building catalogs for each individual dataset. At present over 20,000 events have been located or associated with external catalogs.

To make significant improvements in the locations, we are in the process of testing a new series of computer programs that implement the separated earthquake location method (SELM) for seismic catalogs at any scale. We first build a 3D grid of points inside the earth and associate every event within a region to one or more grid points using a simple distance cutoff criterion. For each grid point that has data associated with it we apply SELM with reference path anomalies computed as the difference between a 3D model and a 1D reference model. Projectors are used to extract only the 4D bias terms from the 3D model with the remaining residuals used to extract other components of the path anomaly vector for that grid point. This provides a mechanism to produce empirical 3D travel time grids that can be updated easily as new data is acquired.

In central Asia a perfect set of test events are the Omega explosions conducted as part of the Kazakhstan-American calibration experiment. In addition to the 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 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 out the importance of calibrated stations and the benefits of using all available data.

KEY WORDS: catalogs, seismicity, location

OBJECTIVE

Seismic event detection and location are the single most important research issues for adequately monitoring a Comprehensive Test Ban Treaty. 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 system 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 Production

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° to 60° N and 30° to 110° 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 seismicity will make the catalogue superior to any other product for the same region.

During the last year (2000-2001) our focus has been constructing a catalogue in Asia, centered on the Tien Shan. The stations used to construct the catalogue in those in KNET, a ten station broadband network in eastern Kyrgyzstan and the PASSCAL experiment deployed in stages from 1997-2000 in the Tien Shan mountains (23 stations in Kyrgyzstan and 15 stations in China). Figure 1 shows the distribution of the stations which occupy a box 5 x 5 degrees (approximately 300,000 km²). The total number of events

detected and located is more than 9000. Figure 2 shows the automated locations, with symbols scaled to event size and color coded to event depth.

The event magnitude scale must be calibrated to other catalogues, but preliminary analysis indicates that the detection capability in the 5 degree circle about the center of the array is approximately 2.0. This can be compared to the detection capability of KNET alone. Using the detection threshold technique of Harvey and Hansen (1989, the first step is to filter the broadband data to the traditional narrow frequency band appropriate for estimating m_b (approximately 1 Hz). After filtering, the ratio of the P wave amplitude to the pre-event amplitude noise (snr) is calculated. Using the value of m_b provided by standard catalogues such as the PDE, it is a linear function to scale the m_b by the snr to determine the minimum detection threshold. Eakins et al. (1999) determined that the detection capability for seismicity in the region of Lop Nor recorded at KNET is approximately magnitude 2.5. The experience from KNET is really just a manifestation of the fact that having seismic stations *close* the seismicity increases the detection. The fact that regional networks can greatly improve the monitoring capability for the given region is a central tenet for testing global monitoring catalogues.

Event Location Methodology Development

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. The procedure we use has three steps: (1) cluster association, (2) location and simultaneous estimation of travel time surfaces, and (3) error estimation. At the time of the deadline for this report the computer code to implement our concepts has been written but not yet fully tested. We described the details of the existing or planned implementation of each of these three steps in separate sections below.

Event cluster association. Our objective is to estimate a set of travel time corrections relative to a reference radially symmetric earth model for 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 subareas 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 in the database 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.

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 simultaneous methods like JHD where the whole system of equations is inverted all at once.

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 \mathbf{P}_R and \mathbf{P}_N . We apply these projectors using the relation

$$\mathbf{s} = \mathbf{P}_R \mathbf{s}_{3d} + \mathbf{P}_N \mathbf{s}_{data}$$

where \mathbf{s} denotes a $N_s N_p$ vector of path anomalies. That is, each component of \mathbf{s} , is a difference, $s_i = (t_{3d})_i - (t_{ref})_i$, between the travel time based on three-dimensional earth model and some 1D reference model. What we call \mathbf{s}_{3d} and \mathbf{s}_{data} use a different realization of 3D earth structure. Jordan and Sverdrup (1981) introduced the concept we use here in what they 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 \mathbf{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 \mathbf{s} . The projectors we construct are fundamentally different from the J&S approach. The basic difference is that in PMEL/SELM the projectors are constructed to annihilate the dependence of residuals on the hypocentral parameters instead of \mathbf{s} . This distinction is important to realize since to date the Jordan and Sverdrup approach seems to be the method of choice of various groups working on location problems within the DoD programs. That is, we are using a procedure that differs significantly from those based on the J&S paper.

The key feature SELM adds is that is not possible with the original J&S methods is that we utilize the term $\mathbf{P}_R \mathbf{s}_{3d}$ to add ancillary information to resolve the absolute location ambiguity. That is, we utilize a 3D model as a secondary reference model used only to resolve the absolute location ambiguity. The data are used directly to estimate the term $\mathbf{P}_N \mathbf{s}_{data}$ using PMEL. Because the \mathbf{P}_R and \mathbf{P}_N projectors are orthogonal the two procedures are completely decoupled. The method described in Jordan and Sverdrup's 1984 paper shares this decoupling feature but is incapable of applying the equivalent of $\mathbf{P}_R \mathbf{s}_{3d}$ exactly.

In the current implementation an ensemble of events and their related arrival times are associated with target points in space that define a GCLgrid (see above). The generalized PMEL/SELM program is then applied 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 $\mathbf{s} = \mathbf{s}_{3d}$. 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. A final detail is that we distort the GCLgrid by shifting the location of each node point from the target point to the hypocentroid of the final location. This distortion complicates the problem of utilizing a GCLgrid in other applications so we anticipate using an existing GLCgrid mapping algorithm we have developed to interpolate these distorted surfaces onto a regular GCLgrid.

Error Analysis. We have implemented a novel error estimation scheme based on previous work by Pavlis (1986, 1992). The approach we use utilizes the pair of orthogonal projectors described above to separate the errors caused by inadequacies in the velocity model into two terms. First, uncertainties in $\mathbf{P}_R \mathbf{s}_{3d}$ translates directly into a location bias and can be handled by the absolute error methods described by Pavlis (1986). Secondly, uncertainties in the overall estimate of the path anomaly to a given point in space, \mathbf{s} , fold into relative location errors. Pavlis (1992) shows that relative location errors can be treated as a term

involving variations in the constellation of stations that record a given event, which scales with the total error s , and a second term that scales with the spatial gradient of s . Future applications to real data should help us understand which of these terms dominate at different scales.

CONCLUSIONS AND RECOMMENDATIONS

We have constructed a catalogue for central Asia. This catalogue is rich in small events that do not appear in other global catalogues, and represent 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.

ACKNOWLEDGEMENTS

We would like to acknowledge Jennifer Eakins (UCSD) who has headed up the automated processing of the KNET and Tien Shan data.

REFERENCES

- 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.
- 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.

Stations contributing to the Tien Shan dataset

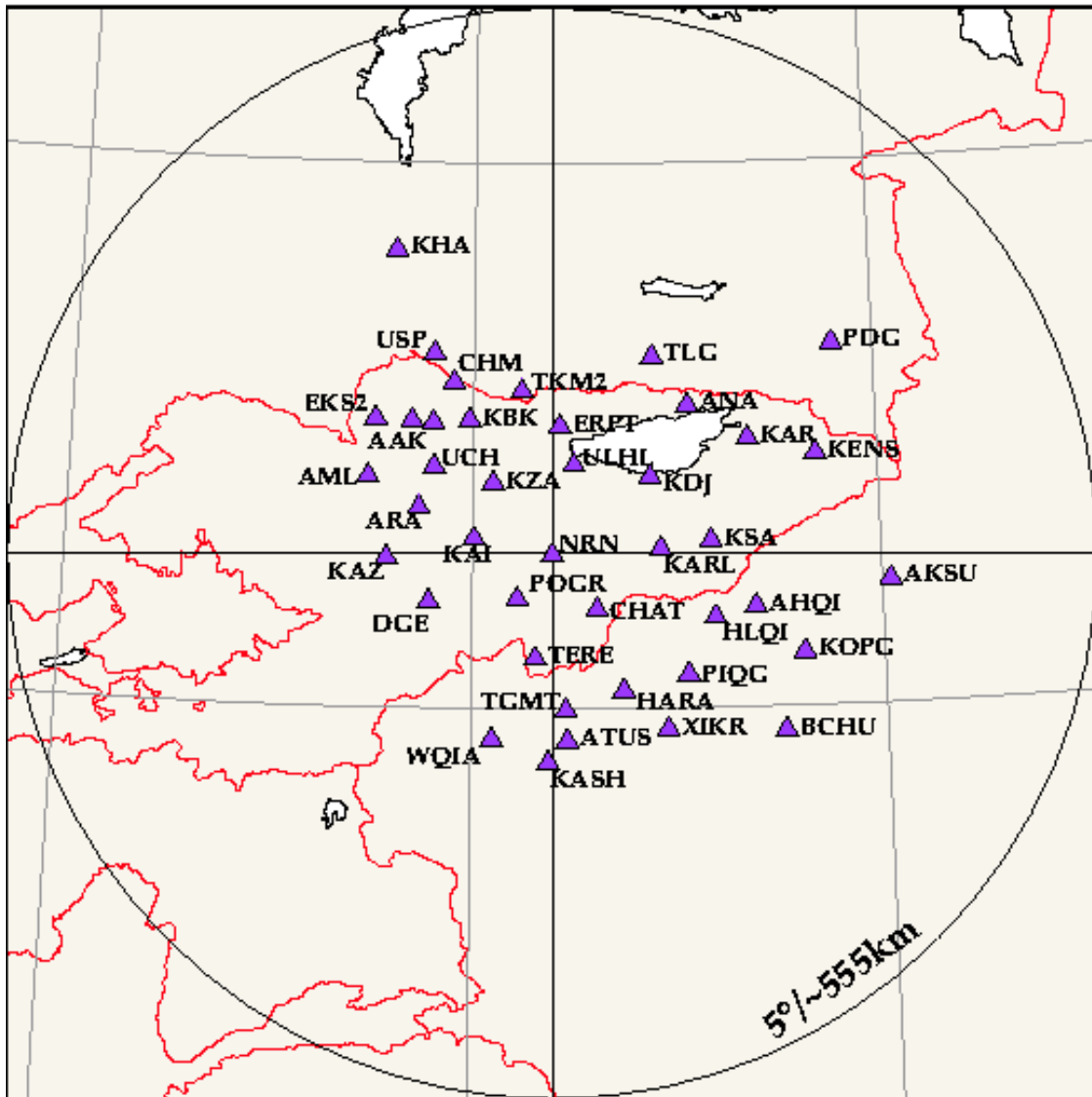


Figure 1. Seismic Stations used to construct the central Asia Catalogue.

**Reviewed events recorded by Tien Shan network
(09/28/1997 - 02/23/1999)**

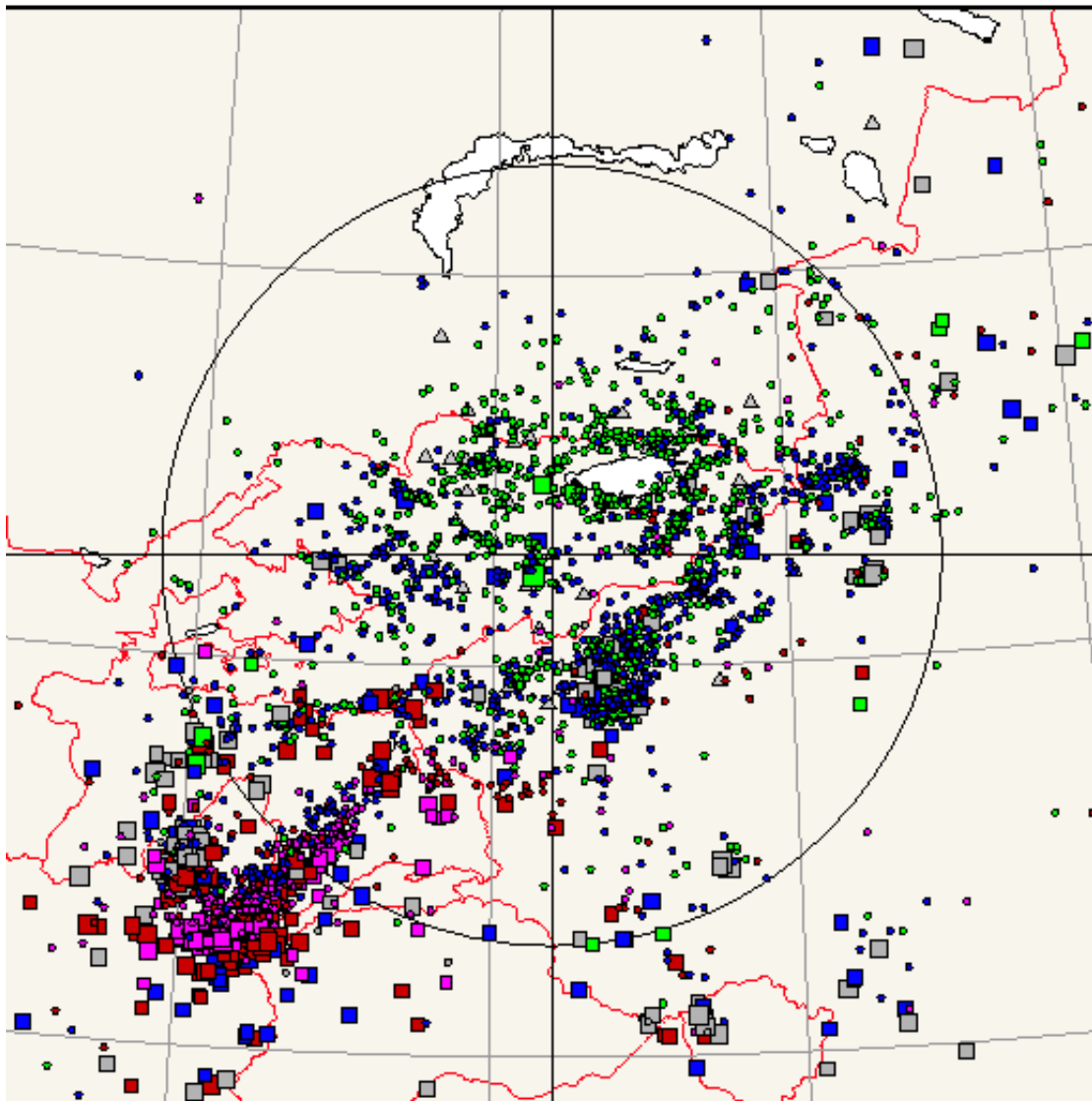


Figure 2: Location of events processed recorded at the stations shown in Figure 1. The epicenter symbols are scaled to event size, and the color is scaled to event depth