DEVELOPING REGIONAL SEISMIC DATABASES FOR IMPROVING EVENT LOCATION AND GROUND TRUTH DATA SETS IN ASIA

Michael L. Begnaud, Julio C. Aguilar-Chang, Aaron A. Velasco, Lee K. Steck

Los Alamos National Laboratory

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation
Contract No. W-7405-ENG-36

ABSTRACT

Seismic event location remains one of the most crucial elements in monitoring for nuclear explosions. Recent development of a two-dimensional (2-D) empirical travel-time technique allows for correcting for unmodeled velocity structure, thus improving typical seismic event location using station corrections and/or a 1-D velocity model. However, this technique requires ground truth information, which can be in the form of known explosion locations and well located teleseismic events. To develop a ground truth data set, we rely on existing catalogs and our own travel-time information.

Many global and regional seismic catalogs with earthquake information for Asia are used in location studies. Each catalog contains origin and arrival information that may or may not overlap with the other catalogs. In order to obtain the most accurate earthquake locations, all available arrival information should be combined into a single data set, including derived travel times from digital stations. We have developed a seismic location database for the China region, combining origin and arrival information from a number of global catalogs, including the prototype International Data Center (pIDC) Reviewed Event Bulletin (REB), United States Geological Survey (USGS) Earthquake Data Reports (EDR), International Seismic Centre (ISC), as well as several regional catalogs. Regional arrivals obtained from digital data in Asia are also included. We also include ground truth information from previous research efforts for nuclear test sites and regional mining information. This merged database will provide detailed arrival data from which to produce more accurate locations.

Global and regional origin and arrival catalogs are merged using ORLOADER, a software package developed at Lawrence Livermore National Laboratory for combining individual seismic event catalogs into a master database. The program keeps identification numbers unique and uses a hierarchy table to select preferred origins. The master arrival table will also be analyzed for redundant arrivals based on a pre-approved author hierarchy. Certain catalogs are known to have errors in arrival times from truncation and machine versus manual picking. Traits such as this form the basis for determining a suitable hierarchy of criteria to allow removal of duplicate information. The final location database will thus have the most accurate arrival information available and be used to select appropriate information for various seismic location studies and/or ground truth catalogs.

We will determine how this new database improves location in the China region by relocating the events and observing how many arrivals remained defined in the solution versus the number from the original catalogs. Preliminary analysis indicates that locations will be improved for those events with few original defining phases and/or a large azimuthal gap.

OBJECTIVE

With the multitude of earthquake catalogs available to researchers, a determination of which catalog to use for locations, depths, arrivals, etc. is necessary. Errors are observed in many global and regional catalogs that can propagate into earthquake relocations and cause numerous problems with related research. Separate catalogs also contain varying degrees of accuracy on such data as arrival picks and station coordinates.

Of crucial importance to monitoring for nuclear explosions all over the earth is the process of event location/relocation. The use of empirical travel-time correction surfaces has greatly improved the accuracy of regional seismic locations. However, adequate ground truth information is required to create these surfaces. Global earthquake catalogs may rely on many common stations for their data, but many do not include similar stations of interest. By combining arrival information from many catalogs and using only the "best" information, ground truth information should be improved, resulting in more accurate seismic locations.

We have developed a seismic location database for the China region, combining origin and arrival information from many common global catalogs [e.g., International Data Center (IDC) Review Event Bulletin (REB), International Seismic Centre (ISC), United States Geological Survey (USGS) Earthquake Data Reports (EDR)] as well as regional and local catalogs from various institutions and researchers. Also included is ground truth information from research efforts around nuclear test sites and mines in the region of interest.

RESEARCH ACCOMPLISHED

Data Acquisition and Merging

As a first step in the merging of global catalogs and the compilation of arrival time information, new seismic information is automatically downloaded in the form of daily and weekly seismic bulletins (ASCII text files) via FTP transfers from open sources, such as the USGS EDR and IDC REB. These text files are parsed through a number of programs to convert the seismic data from their original catalog formats to CSS3.0 format flat data files. For example, we use modified versions of programs EDR2DB and REB2DB (distributed by BRTT as the Antelope software) to convert the EDR and REB seismic bulletins to CSS3.0 style flat files. After the data are converted and loaded into the database, several SQL scripts are run to quality control (QC) the data, identifying and correcting any problems with the data. Such QC checks include verifying that all data were loaded, fixing any duplicate entries based on pre-determined database table constraints, and enabling database table constraints (primary and unique keys) to ensure data integrity.

A method is needed to merge the origin information so each event has only one preferred location. We use the database utility ORLOADER (developed at Lawrence Livermore National Laboratory) to merge all of the bulletin data into a global table. This utility analyzes origin information and combines different catalog origins into an event and preferred origin based on location and origin time. All of the database identification numbers are renumbered and kept consistent with the arrival information also included in the seismic bulletins.

The merged event and origin information is now used to tie all of the in-house digital waveform files (formatted as Seismic Analysis Code, or SAC, files) to the events in the global merged database. All of the waveform holdings are QC'd and duplicate and incomplete/missing waveforms and SAC header information are dealt with. The SAC headers contain information on the arrival, station, and event location associated with the waveform and are updated using Perl scripts to correspond with the global tables.

Once the waveforms are tied to the database, in-house Perl scripts are run to create WFDISC, WFTAG, ASSOC, and ARRIVAL database tables that contain data about our waveform holdings and Los Alamos National Laboratory (LANL)-generated phase picks. ORLOADER is then run to merge LANL phase arrivals and waveform information with the global tables. A schematic of the data acquisition and merging process is shown in Figure 1.

Developing the Location Database

The majority of earthquake location programs that utilize some type of CSS3.0 schema tables (database or flat file) require there to be one origin entry per event in the ORIGIN table. In addition, all of the arrivals should point to one origin via the ASSOC table. The process of merging catalogs results in EVENT and ORIGIN tables, where there are several origins for any given event. Many ARRIVA L tables, which have duplicate phase picks for the same event but different origins, are now merged. In order to correctly relocate any of these events, specific rules for arrivals must be adhered to, such one *P* phase used per station, per event. With the large number of duplicate arrivals, a method needs to be developed to select appropriate phases from robust catalogs.

Using merged global tables, we developed a method to create ASSOC and ARRIVAL database tables that have the proper phase names for use with location programs such as EvLoc (Bratt and Bache, 1988; Nagy, 1996) and MatSeis (Harris and Young, 1997). The ASSOC table contains only the preferred origin location for a given event (one event can have one or more origin locations; each location corresponds to a different author, such as USGS EDR, IDC REB; etc.), based on a pre-determined LANL-generated ranking table, which ranks preferred authors to origin locations based on that author's location and time estimate uncertainties. The ASSOC table also has the phase arrivals from all sources (global bulletins, regional catalogs, LANL picks) associated with the preferred origin of an event.

Seismic phases in the ASSOC table are renamed so that all phases associated with a given event have unique descriptive names [this is a requirement of the EvLoc (libloc) program used in the location effort]. The renaming of the seismic phases follows a pre-determined ranking scheme, in which LANL arrival picks are ranked highest, and takes place following the ranking based on phase pick author and phase names.

A common problem is the phase-naming convention for *P* and *S* arrivals. Many catalogs name the *P* phases just *P*, not necessarily a *Pn* if needed (same for the *S* arrivals). Originally, we assumed we could group all of the *P* phases, all of the *Pn*, etc. and remove duplicates. However, many times one catalog would name the same *P* phase a *P* and another would name it a *Pn*, even if the arrival times were the same or very similar for that station. Therefore, it is necessary to analyze the *P*, *Pn*, *S*, and *Sn* phases and choose the correct arrival to rename to *P* or *S*.

To account for these phase discrepancies, all of the *P*, *Pn*, and *Pg* phases for a given station/event are grouped by author rank and time. If, by chance, there were any combination of the three phases for the same station/event, the phase with the minimum time for that author (if it was the highest ranking) would be selected as the "*P*", including any azimuth and/or slowness measurements. We assume that the first P-type arrival was made from the first break in the waveform and, therefore, has the best chance of being an accurate pick. By renaming the phase to *P*, we allow the velocity model in the location program to determine if the time corresponds to a *P*, *Pn*, or *Pg* phase.

We did not want to actually remove the other P-type phases from the table, so they were renamed to append the rank number at the end of the phase, thereby removing the phase from the location procedure, but leaving it in the database for possible later use or comparison. For more complex phases such as pP, PmP, etc., this process of renaming the duplicate phase to add the rank number was all that was modified. The user could later choose whether or not to include more complex phases in the solution. An example of the phase renaming method is shown in Table 1. The renaming of phases is necessary in order to provide the best distinct phase arrival times to the programs used in the location effort. By doing this, we are compiling a database with the "best" phase arrivals, so that for a given event at a given station, there is only one *P*, *Pn*, and *Pg* phase.

After phases have been renamed and distinct names given, we create synthetic arrival times based on the preferred origin. These times are then used to modify the time residual field in the ASSOC table so it is accurate for the given origin. The field for number of defining phases is also updated to be consistent with the actual number of phase chosen for location. The azimuth and slowness measurements are also updated. These updates permit the researcher to choose events for relocation that have a minimum number of defining phases, time residual, and/or azimuthal gap.

Relocation Results

We will utilize the location database to test effectiveness in relocating large sections of the LANL merged catalog. A first test will relocate approximately 156 events for a region around the Mw=7.5 Tibet event of 08NOV1997, with manually-picked LANL *P* and *S* arrivals as well as global catalog arrivals having a station-event distance ranging from 5.5 to over 30 degrees. The Tibet event provides a unique ground truth test due to the presence of a related surface rupture identified by Inferometric Synthetic Aperture Radar (InSAR) (Peltzer *et al*., 1999) and the determination of the Tibet event as associated with a vertical strike/slip fault (Velasco *et al*., 2000). We will test whether the location database produces relocations that better align with this surface rupture.

Figures 2 and 3 show preliminary relocations using *P* and *S* travel-time correction surfaces for USGS EDR origins/arrivals only and LANL location database origins/arrivals, respectively. The main shock is located at the center area of the surface rupture. Using the EDR arrivals only, the main shock does not transfer to the surface rupture. However, in other studies, the main shock does transfer to the rupture, using and correcting only the *P* phases (Steck *et al*., 2001).

For the location database, the number of events has increased as well as the number of stations and phases. Typically, improvement is readily observed for those events that, in standard catalogs, had few defining phases and/or large azimuthal gap. Relocating using the correction surfaces appears to produce events that align more on the rupture, and moves the main shock closer to the rupture. Again, the shear phases have inconsistencies through most catalogs and do tend to introduce more error in the locations. There is still considerable scatter in the relocations, but this probably is due in large part to the sparse nature of the data.

Other location methods have been tested with the location database, including application of the double-difference location algorithm HYPODD (Waldhauser and Ellsworth, 2000), using the location produced with the *P* and *S* corrections as the starting point (see *Steck et al*, this Proceedings).

CONCLUSIONS AND RECOMMENDATIONS

With the multitude of global, regional, and local catalogs, it is necessary to develop methods to incorporate all available arrival time data into one main catalog in order to create the most accurate set of travel times for use in seismic location. Merging datasets permits all available arrivals to be accessible to the researcher, where specific choices can be made about which authors and arrivals to use for the location process. Duplicate and/or redundant arrival information must be removed from the database so location algorithms have distinct phases for events and stations. The location database typically results in improved origin accuracy as well as an increased number of arrivals for an event. This method can improve locations considerably for events with few arrivals and/or large azimuthal gap.

REFERENCES

Bratt, S. R. and T. C. Bache (1988). Locating events with a sparse network of regional arrays, *Bull. Seism. Soc. Am.*, **78**, 780-798.

Harris, M. and C. Young (1997). MatSeis: a seismic GUI and toolbox for MATLAB, *Seis. Res. Lett.*, **68**, 267-269.

Nagy, W. (1996). New region-dependent travel-time handling facilities at the IDC: Functionality, testing, and implementation details, *SAIC Tech Rep. 96/1179*, 57 pp.

Peltzer, G., F. Crampé, and G. King (1999). Evidence of nonlinear elasticity of the crust from the Mw7.6 Manyi (Tibet) earthquake, *Science*, **286**, 272-276.

Steck, L. K., A. A. Velasco, A. H. Cogbill, and H. J. Patton (2001). Improving regional seismic event location in China, *Pageoph*, **158**, 211-240.

Velasco, A. A., C. J. Ammon, and S. L. Beck (2000). Broadband source modeling of the November 8, 1997 Tibet (Mw = 7.5) earthquake and its tectonic implications, *J. Geophys. Res.*, **105**, 28,065-28,080.

Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California, *Bull. Seism. Soc. Am.*, **90**, 1353-1368.

Table 1. Example of phase renaming for example event #100 at example station XYZ.

Phase pick author	Rank	Original phase			Renamed phase		
		time 1	time 2	time 3	time 1	time 2	time 3
LANL	1	Pn	Pg		P	Pg	
REB	2	Pg	P	Pn	Pg2	P2	Pn
EDR	3	Pn	P	Pg	Pn3	P3	Pg3

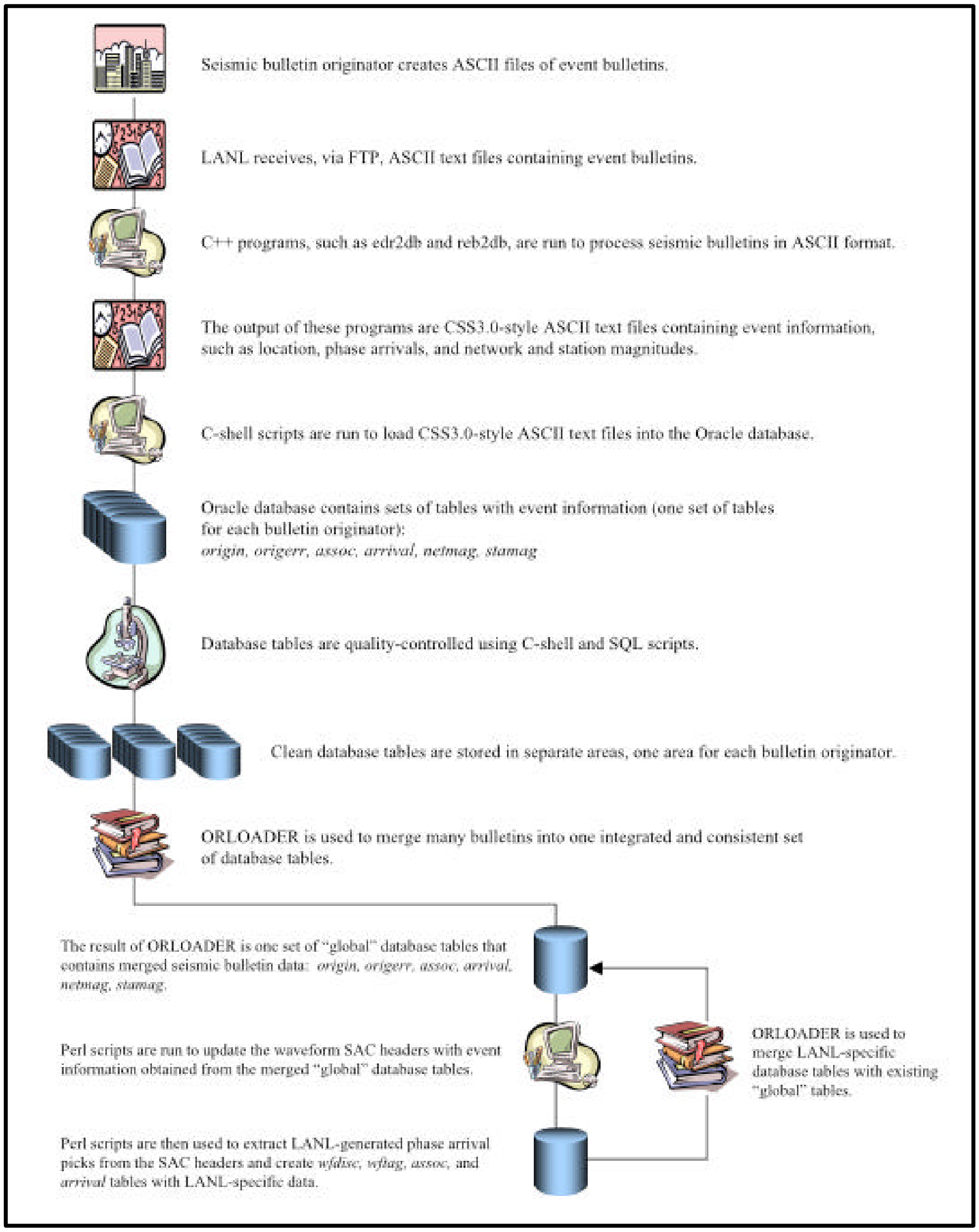

Figure 1. Flow chart for LANL seismic catalog merging process.

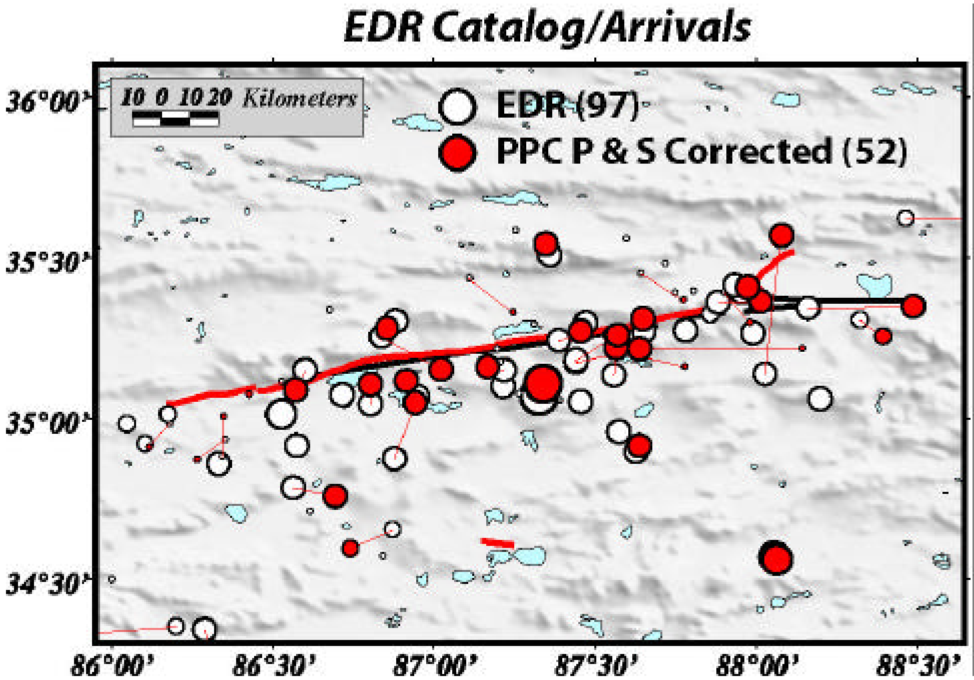

Figure 2. Relocations of Tibet events around a surface rupture using EDR catalog locations and arrivals only. Red line is surface rupture, black line is mapped fault. Both P and S phase travel time correction surfaces were used in the relocation. The main shock (largest event near center of rupture) does not relocate on the rupture (the event was determined to be a vertical strike/slip mechanism, see text).

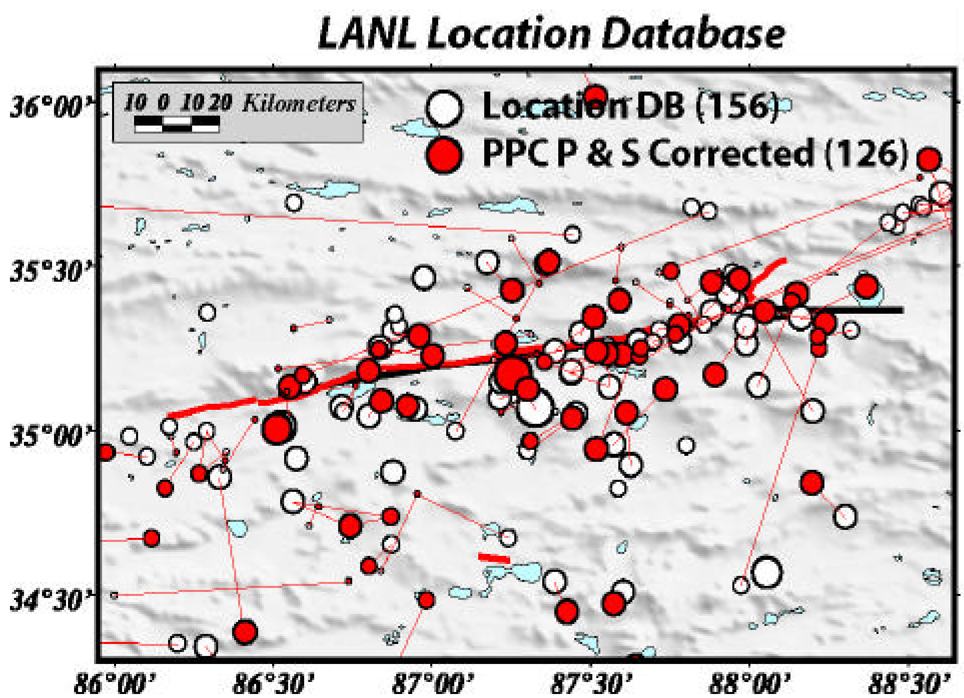

Figure 3. Relocations of Tibet events around a surface rupture using LANL location database origins and arrivals. Refer to Figure 2. The number of events in the area has increased relative to the EDR catalog. The main shock relocates much closer to the surface rupture and many other events align more along the rupture and produce more event clusters.