SEISMIC DATABASE IN SUPPORT OF REGIONAL MONITORING RESEARCH IN ASIA

Julio Aguilar-Chang, Aaron A. Velasco, Hans E. Hartse, Los Alamos National Laboratory

Sponsored by U. S. Department of Energy
Office of Nonproliferation and National Security
Office of Research and Development
Contract No. W-7405-ENG-36

ABSTRACT

The Geophysics Group at Los Alamos National Laboratory (LANL) has successfully developed and implemented a research database in support of regional seismic monitoring research, with the goal of improving our ability to locate, characterize, and discriminate small magnitude events recorded at regional distances. The database contains a wide variety of data (approximately 20 Gigabytes), gathered from a variety of open sources, including the Incorporated Research Institutions for Seismology (IRIS), the United States Geological Survey (USGS), the prototype International Data Centre (pIDC), the International Seismological Centre (ISC), and individual contractors and collaborators.

The types of data that we have collected from these sources include global seismic event catalogs, raw seismic digital waveforms, regional arrival information, and network and station information. Contributions by contractors and collaborators have significantly added to this database. For example, the USGS has provided velocity models for China and GIS information such as topography, moho depth, and mine locations. Woodward-Clyde has provided source studies of events that we now use as depth ground truth. The seismic database also contains information from our own efforts, including processed seismic data, arrival time and amplitude measurements, ground truth information, travel-time tables for regional 1-D models, and processing algorithms.
This regional seismic database is a rich source of information for regional seismic research and has allowed for regional studies on a wide range of topics, including surface-wave studies, propagation studies, magnitude calibration studies, special event studies, location improvements, and discrimination enhancements. We have also created a subset of this large dataset to focus on nuclear test sites in Asia. Not only does the dataset provide seismological information for the nuclear test sites, it also allows for improvements in the detection, location, and discrimination of events near the test sites.

The dataset is also an integrated design, allowing for easy use. That is, reference event waveforms correspond to events in the earthquake catalogs, station information corresponds to the waveforms, and 2-D travel-time correction surfaces are derived from picked travel-time arrivals derived from the catalog events. Future enhancements will include more seismic data, the latest regional discrimination procedures, special event analyses and data, and calibration studies.

Key Words: seismic database, regional, Asia, nuclear test sites

OBJECTIVE

The Geophysics Group at Los Alamos National Laboratory (LANL) has successfully designed and implemented a research database that contains a wide variety of information. Much of the information in the database stems from our efforts to mine open sources of data, which allows us to use this information for our regional seismic research efforts. The database is focussed on regional data, although we include global information when necessary. This database is also being used as a research tool, with the goal of developing methods to improve the detection, location, and discrimination of events recorded at regional distances in Asia. Researchers use the seismic database to access information for reference events, earthquake catalogs, regional station information, digital waveforms, ground truth information, and amplitude measurements. Research that is finalized is then integrated into the seismic research database, making the database a rich source of regional seismic information.

RESEARCH ACCOMPLISHED

Database Design

The seismic data gathered resides in an Oracle™ relational database, which runs on a Sun server. The size of this database is approximately 20 gigabytes and consists of over 200 reference tables. These tables contain data from a wide variety of sources, such as information from the Unites States Geological Survey (USGS), the prototype International Data Centre (pIDC), the International Seismological Centre (ISC), and individual contractors and collaborators. The database is an integrated design, allowing for easy use. That is, reference event waveforms correspond to events in the earthquake catalogs, station information corresponds to the waveforms, and 2-D travel-time correction surfaces are derived from picked travel-time arrivals derived from the catalog events. Future enhancements will include more seismic data, the latest regional discrimination procedures, special event analyses and data, and calibration studies.

The global catalogs from the USGS, pIDC, and the ISC contain event origin information, which is a summary of hypocentral parameters and information describing a derived or reported origin for a particular event. Associated with event origins are tables that contain summary information of seismic arrivals. This is information characterizing a seismic phase observed at a particular station. The phase arrival attributes conform to seismological convention and are listed in earthquake catalogs.

There are auxiliary tables that support origin and arrival information. There is a table that contains network-station affiliations, by which seismic stations may be clustered into networks, as well as tables containing station location and station-channel information. Through the use of these tables, we can obtain station names and locations. The station-channel information available in these tables describes the orientation of a recording channel at a particular site; it also provides information about the various channels that are available at a station and maintains a record of the physical channel configuration at a site. Network and station magnitudes based upon measurements made on specific seismic phases are also available in other tables. All these data can be tied to SAC waveforms by means of a table that contains waveform header file and descriptive information, which provides pointers to waveforms stored on disk.

The tables within the research database contain information on more than 400,000 distinct events, beginning in 1973 to the present. Of these, there are 13,662 events that have 128,120 related SAC waveforms. There have been over 3.5 million amplitude measurements made on 5,287 seismic phases representing 3,593 events with related SAC waveforms.

The research database is being used to store, query, manipulate, and produce reports on the following data:

- event location and confidence bounds
- information on seismic arrivals
- amplitude measurements
- waveform header file and descriptive information
- network and station magnitude
- station location and channel information

Database Content

Earthquake catalogs

We have collected earthquake catalog information from various sources. Most of our processing of waveform data requires an event location, which we take from available reference earthquake catalogs. We have gathered earthquake catalog information from global (United States Geological Survey: USGS) and regional (State Seismological Bureau of China: SSB) sources, and store the information as CSS 3.0 Oracle™ tables. We have also gathered information from the Reviewed Event Bulletin (REB) catalog from the prototype International Data Centre (pIDC). The earthquake catalogs are intended to provide a foundation for storing our regional waveform data. Thus, every waveform will have an origin for which it is associated. The catalogs may also be used for other purposes, such as event location and historical seismicity.

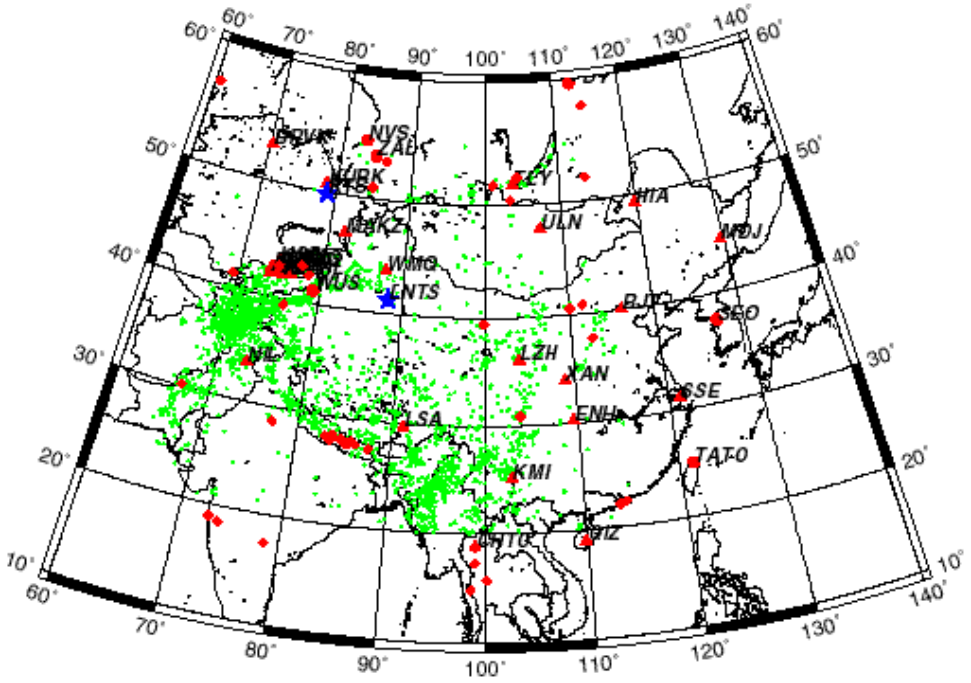

Figure 1: Map showing the Lop Nor (LNTS) and Kazakhstan (KTS) nuclear test sites (stars), USGS catalog events (dots), digital stations for which we have data (triangles), digital stations that are not in this database (circles), and USGS reporting stations that have arrival times in this database (diamonds).

We currently have data from the USGS Preliminary Determination of Epicenters (PDE) catalog between Jan. 1973 and Dec. 1989, the USGS Earthquake Data Report (EDR) catalog between Jan. 1990 and Jan. 1998, the USGS weekly PDE catalog between Feb. 1998 and Feb. 1999, and the State Seismological Bureau of China (SSB) catalog between Jan. 1973 and Dec. 1989. The REB catalog event data covers from May 1993 to Nov. 1998. The EDR catalog contains origin and phase information, and is considered to be the final event location in these USGS catalogs. Event magnitudes for the USGS events are generally above m_b 4.0. Prior to 1990, we use the final PDE origin records, which do not include phase information. For all of 1998, researchers use the weekly PDE data, since the final locations are not yet available. Although these are not the final USGS locations, they allow us to work with new waveform data. We also have loaded into our research database origin information from the State Seismological Bureau of China (SSB) catalog between the years 1973 to 1989. This catalog includes many regional and local events that are not in the global USGS catalogs. The SSB events have local magnitudes (M_L) ranging between 2.0 and 4.5. The REB catalog data were obtained as exported database tables from the Center for Monitoring Research (CMR), who runs the pIDC. These catalogs offer their own quality control for which we have little information. However, our quality control has been to guarantee that the electronic files were converted correctly and input properly into the database. Furthermore, by using primary and unique constraints in the database, we have identified duplicate records and misassociated phases. We have carefully reviewed these duplicates and misassociations and have corrected them in the database.

Waveforms for reference and catalog events

These are segmented digital waveforms from selected events (nuclear explosions and earthquakes) in eastern Asia. The events corresponding to the waveforms are included in the event catalogs. The waveforms are obtained from the Incorporated Research Institutions for Seismology (IRIS) Data Management Center (DMC), and were retrieved using windows defined by the catalog epicenter information. These digital waveforms are used as reference waveforms for the evaluation of regional seismic events in Asia. The waveforms are also used as a historical reference waveform database that can assist with special event analysis.

The events for which we have waveform data occurred between 1974 and 1998. The waveform segments were retrieved from the IRIS database using the event origin information from the United States Geological Survey (USGS) Preliminary Determination of Epicenters (PDE) and Earthquake Data Report (EDR) event catalogs, the Reviewed Event Bulletin (REB) catalog from the prototype International Data Centre (pIDC), or using the Chinese State Seismological Bureau (SSB) catalog. Event magnitudes for USGS PDE/EDR events are generally above mb 4.0, and SSB events range between M_L 2.0 and 4.5. The data are broadband (20 sps), short-period (40 sps), and long-period (1 sps) three-component digital waveforms from stations located in China, Pakistan, Khyrghyzstan, Kazakhstan, Mongolia, Russia, Korea, and Thailand. Data from the 1980's are mostly from the Chinese Digital Seismic Network (CDSN). The waveforms are stored as Seismic Analysis Code (SAC) files, which are binary files with header information. The files can be read using the computer program SAC [Tapley and Tull, 1992]. Furthermore, any tool that reads CSS-3.0-style information (e.g., MatSeis) can read our information. We also store instrument calibration information, which have been calculated for most of the stations in our database.

We have made travel-time picks and amplitude measurements of regional phases on many of the waveforms, and store the travel-time information. Picks were made on broadband channels when available; otherwise, they were made on short-period channels. The associated amplitude measurements from these picks are stored in a custom table. Regional phase amplitudes are measured using time domain techniques following the techniques of Hartse et al. (1997).

Travel-time tables and 2-D correction surfaces

We have developed 1-D travel time tables for regional velocity models in the region. The travel-time tables are stored as flat files, but are read by an event locator that writes results to the research database (see Steck et al., this issue for a review). The velocity models are derived from a number of sources, including the USGS, which are used as base models to compute 2-D travel-time correction surfaces to improve seismic event location. This is a research product that was derived from preliminary information in the seismic research database, and the results were later integrated into the database.

Preliminary ground truth events

We have digital waveform data for 10 Lop Nor nuclear explosions ranging in m_b between 4.7 and 6.5. We have also created a list of 26 events with well-constrained depths located within 1500 km of the Lop Nor test site in western China. Woodward-Clyde has provided source studies for these events using a variety of techniques, and we have looked at a number of other sources (see Steck et al., this issue). The events can be considered as depth ground truth events and are meant to be helpful for identifying reference events. This ground truth earthquake data is not based on epicentral location (such as GT2, which is a ground truth event known to within 2 km). Rather, the list of events is based on well-determined depth estimates using a variety of methods. It is important to emphasize that, in the absence of realistic strategies for gaining ground truth information, this approach of establishing ground truth may be the best means available to verification seismologists for calibrating a region. Furthermore, much can be learned about the performance of seismic techniques from comparative analysis, and such knowledge can be used to design improvements in techniques and/or to identify new data requirements.
Regional magnitude calibration information for station WMQ

Priestley and Patton (1997) developed a list of magnitudes calculated for events recorded at station WMQ, and we have incorporated this information into the database. Regional magnitudes based on regional phases have been shown to be useful for seismic discrimination. A list of regional phase magnitudes calculated for seismic station WMQ and the developed magnitude scale can be used to assist with short-period to long-period discrimination. All of the data and methods are outlined by Priestley and Patton (1997). They studied 27

nuclear tests and 52 earthquakes. One nuclear event occurred at the Lop Nor nuclear test site, while 26 occurred at the East Kazakhstan test site.

Regional station information

Our research database contains station information for all waveform data that has been gathered. We must have accurate station information in order to effectively utilize our waveform data. The station information is used for regional waveform analysis and event location. We have obtained station information from the IRIS database and from the Lawrence Livermore National Laboratory (LLNL) effort. Other information for global catalogs was obtained from the USGS Earthquake Data Report (EDR), which include many stations at teleseismic distances. We have compared the latitude and longitude of each station from different sources and have discovered some discrepancies. For example, stations sometimes have different elevations between the EDR and IRIS sources. Thus, we have kept the two sources separate. Unfortunately, we have little information about the accuracy of the station locations.

We have visually scanned the instrument response information gathered for each station. At this time, the instrument files should have accurate information. The accuracy of this information is dependent on the source of the station information. Some stations listed both by the IRIS database and the EDR differ in latitude, longitude and elevation. For station information, see the on-line Federation of Digital BroadBand Seismographic Network Station Book (FDSN Station Book).

Contextual (GIS) information

The USGS has provided a contextual dataset for China that includes moho depth, mine locations, faults, tectonic regions, and seismicity information. The dataset is stored in a GIS format that can be read and manipulated by users. Much of the interface to this information was developed by colleagues at Sandia National Laboratories, and compliments the seismic information in our dataset.

Utility for Research

The research database can be utilized to perform research on a vast amount of data. For example, in an effort to improve our ability to locate seismic events in western China using only regional data, Steck et al. (this issue) have developed empirical propagation path corrections (PPCs) using the earthquake catalogs within the database. They used travel-time observations available from the USGS Earthquake Data Reports (EDR), the International Seismic Centre (ISC) catalogs, the prototype International Data Centre Reviewed Event Bulletin (REB), and our own travel-time picks from regional data. They also included ground truth events mentioned above that are stored within the database. Another example is the use of extensive regional amplitude measurement for the development of amplitude corrections in 1-D and 2-D (Taylor et al., this issue; Phillips et al., this issue).

CONCLUSIONS

The Oracle™-based regional seismic database designed and implemented by the Geophysics Group at Los Alamos National Laboratory is a rich source of information that has allowed for research on a wide range of regional seismic topics. These include surface-wave, propagation, magnitude calibration, and special event studies, location improvements, and discrimination enhancements. We have also created a subset of this large dataset to focus on nuclear test sites in Asia. Not only does the dataset provide seismological information for the nuclear test sites, it also allows for improvements in the detection, location, and discrimination of events near the test sites.

Researchers have used the seismic research database effectively to store and manipulate large quantities of seismic information. The relational nature of the Oracle™ database helps in the integration of the different types of seismic data needed for research. For example, event information from published catalogs is stored in a table that is linked to tables containing seismic arrival information and pointers to waveforms stored within the operating system. Event origin and seismic arrival data can also be linked directly to phase amplitude information and to network and station magnitude data.

A major advantage of implementing the CSS-style relational database is the ability to use software developed to support CTBT-related research projects, such as MatSeis, a MATLAB-based graphical user interface and toolbox for working with seismic data, developed at Sandia National Laboratories (Harris et al., 1998).

REFERENCES

Hartse, H.E., S.R. Taylor, W.S. Phillips, and G.E. Randall, Regional event discrimination in central Asia with emphasis on western China, *Bull. Seism. Soc. Am*., 87, 551-568, 1997.

Harris, J. Mark, Christopher J. Young, Judy I. Beiriger, New Features in MatSeis, *20th Annual Seismic Research Symposium on Monitoring a CTBT, Abstract Proceedings*, 85, 1998.

Priestley, K. F., and H. J. Patton, Calibration of mb(Pn), mb(Lg) scales and transportability of the Mo : mb discriminant to new tectonic regions, *Bull. Seism. Soc. Am., 8*7, 1083-1099, 1997.

Tapley, W. C. and J. E. Tull, SAC- Seismic Analysis Code, Users manual, revision 4, 1992.