Crazyseismic: A MATLAB GUI-Based Software Package for Passive Seismic Data Preprocessing

by Chunquan Yu, Yingcai Zheng, and Xuefeng Shang

ABSTRACT

We introduce an open-source MATLAB software package, named Crazyseismic, for passive seismic data preprocessing. Built-in core functions such as seismic phase travel-time calculation and multichannel cross correlation significantly improve the efficiency of data processing. Compared with conventional command-line-style toolboxes, all functions in Crazyseismic are embedded in one single graphic user interface (GUI). The human–machine interactive nature of GUI facilitates data quality control. The simplicity of the software allows users to process Seismic Analysis Code format seismic data with great ease and also provides a means by which users can tailor the software for their specific needs. We demonstrate the power of our software through two field examples: one for P-wave arrival-time picking and the other for receiver function calculation. The software can essentially be used for analyzing all major body-wave phases in seismology.

INTRODUCTION

In modern seismology, especially array seismology, seismic processing and data quality control are critical for the application of the sophisticated modeling, imaging, and inversion algorithms needed to understand the Earth's structures and earthquake source mechanisms. As a result of increased coverage and improved detection capability from both permanent and temporary networks, the volume of seismic data has increased exponentially in recent decades. For instance, the EarthScope Transportable Array has stepped across the conterminous United States with a total number of ∼2000 broadband seismic stations and has recorded thousands of terabytes of seismic data since 2004. Analyzing such amounts of data is prohibitive without an efficient processing workflow or tool.

In the past few decades, several seismic packages have been developed; the most popular is the Seismic Analysis Code (SAC; Goldstein *et al.*, 2003). Because of its rich and powerful library of functions for analyzing three-component seismograms, SAC has become the standard analysis toolbox in the global seismology community. However, its ability to visualize, process, and analyze a large number of seismic traces is limited by its original software architecture. Another powerful open-source package is Seismic Unix (SU; Cohen and Stockwell, 2000), which is capable of analyzing a large dataset. However, the SU is mainly used in exploration seismology with regularized data acquisition geometry that is often not present in global seismology. There are also some Python-based tools, such as ObsPy (Beyreuther *et al.*, 2010) and AIMBAT (Lou *et al.*, 2013). As the commercial platform MATLAB from Mathworks (see Data and Resources) became popular, MATLAB-based seismic packages were developed, such as SplitLab (Wustefeld *et al.*, 2008), Coral and SEIZMO (see Data and Resources). Most of these MATLAB-based packages are either command-line-style utility toolboxes or specified for particular purposes, for example, measuring shear-wave splitting (Wustefeld *et al.*, 2008). As an exception, MatSeis (Harris and Young, 1996) provides an interactive graphic user interface (GUI) for data visualization, processing, and analysis. The implementation of an interactive GUI is crucial for data quality control, but there is still much need for the improvement of overall performance and efficiency of currently available seismic packages.

Here, we introduce Crazyseismic, a new MATLAB GUI-based seismic software package, for passive seismic data preprocessing. Two main features make it distinct from other seismic packages. First, large-volume seismic data can be processed efficiently in a multichannel sense. For example, the alignment of a particular seismic phase and the determination of its relative arrival times at different stations can be simultaneously achieved through multichannel cross correlation. Second, all functions are unified into one single GUI so that users can interactively visualize and process data with great ease. Because Crazyseismic is built entirely on the MATLAB platform, new processing algorithms can be flexibly incorporated into this package.
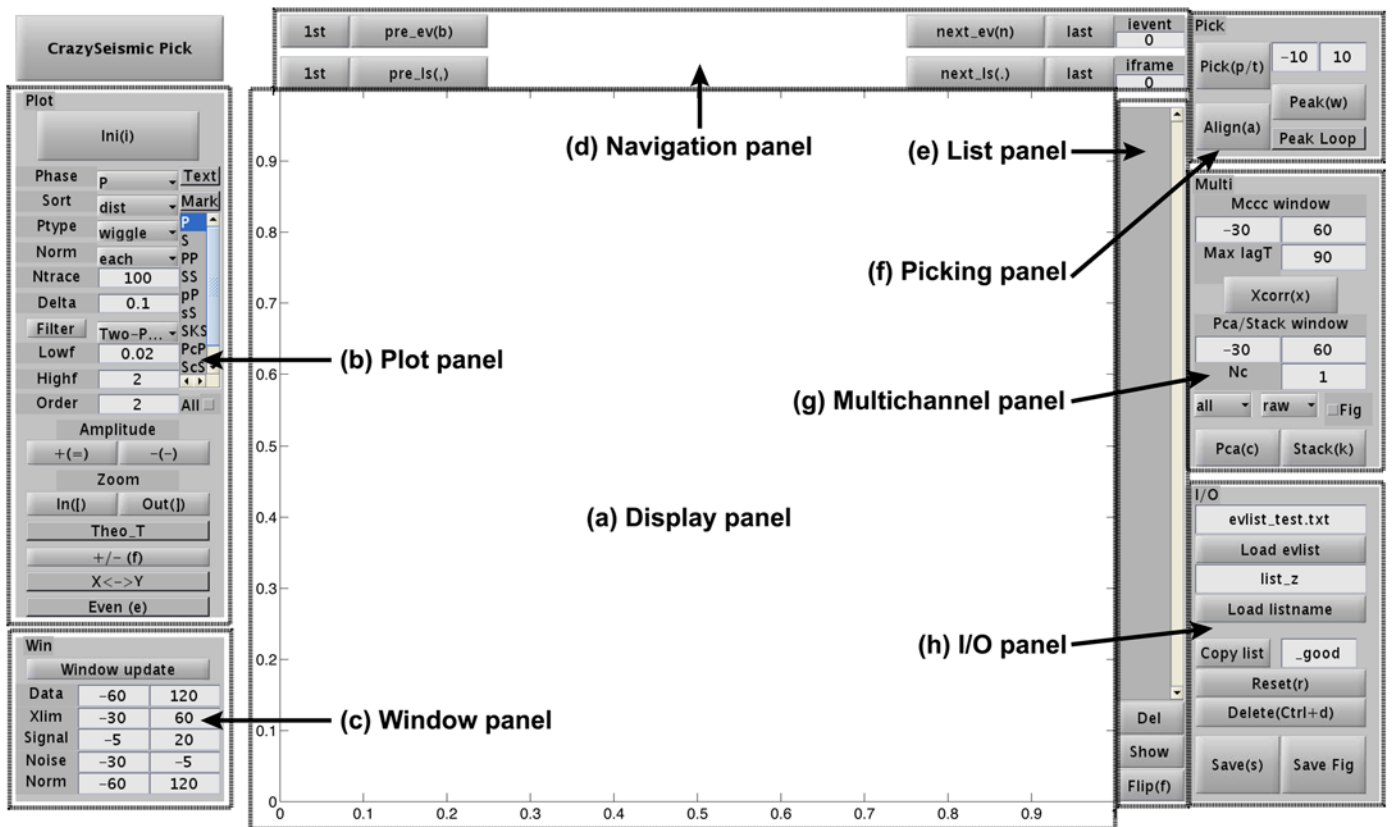
▲ **Figure 1.** Default Crazyseismic Pick graphic user interface (GUI). It comprises eight different panels. The central part is used for displaying waveforms. (Refer to the software manual for more details of parameters and functions.) The color version of this figure is available only in the electronic edition.

Crazyseismic is cross platform and can be executed on a variety of operating systems (e.g., Windows, Linux, or Mac OS) so long as MATLAB is available (signal processing and mapping toolboxes are required). Crazyseismic is open source and free to the public (see Data and Resources). It has been extensively tested since 2012 for MATLAB R2010a and newer versions. Here, we first give an overview of Crazyseismic, followed by a summary of its core functionalities. Then, we show two field examples of using this software for *P*-wave arrival-time picking and single-station receiver function calculation.

OVERVIEW OF CRAZYSEISMIC

Crazyseismic features one single GUI with embedded functions organized by control panels. Based on their major functionalities, we group the interface into nine different panels:

    (a) Display panel: displays seismic waveforms.
    (b) Plot panel: controls plotting parameters.
    (c) Window panel: controls time windows.
    (d) Navigation panel: navigates among events or frames.
    (e) List panel: lists station/file names.
    (f) Picking panel: picks phase arrival times.
    (g) Multichannel panel: controls multichannel cross correlation and principal component analysis (PCA) parameters.

    (h) Input/output (I/O) panel: controls input/output parameters.
    (i) Deconvolution panel: deconvolves one component from the other.

Depending on the purposes of specific studies, the panel (i) may or may not be necessary for basic data processing. We include two versions of Crazyseismic GUIs in the software package: Pick and Decon. The Pick GUI comprises the panels (a–h) and is mainly designed for phase arrival-time picking on single-component seismic data (Fig. 1). The Decon GUI contains all nine panels and is suitable for processing seismic data on two components. The additional deconvolution panel is used to compute deconvolution between two components, for example, to obtain receiver functions. Both GUIs are interactive. Users can sort and display waveforms in a variety of ways, such as by station locations, epicentral distances, and signal-to-noise ratios. We can conveniently select/deselect or delete/retain seismic traces to have an efficient way of data quality control. Moreover, the built-in multichannel cross-correlation function allows users to measure relative phase arrival times accurately and automatically. Exceptions do occur when the waveforms are not similar to each other. In this case, the cross correlation may give incorrect time picks, but we can easily correct them using the interactive GUI. The general workflow of Crazyseismic is summarized in
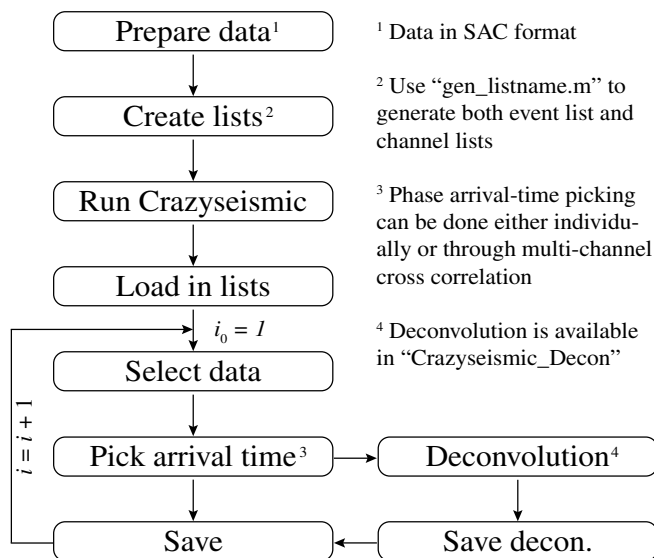
▲ **Figure 2.** General workflow of data processing using the Crazyseismic software package.

Figure 2. Our current version accepts the widely used SAC binary format as the input seismic data format, but it can be extended to other data formats. Crazyseismic reads an event list in which the full paths of seismic events are listed. In each event directory, the software further reads station channel lists and loads corresponding SAC files. The event list and channel lists can be simultaneously generated using the "gen_listname.m" program within the Crazyseismic package. Seismic data are usually processed event by event. The output for each event is a seismic phase information file, which contains phase arrival times, polarities, ray parameters, and so on. (Refer to the software manual for more details.)

CORE FUNCTIONALITIES

In this section, we summarize the key points of four core functionalities in Crazyseismic.

Ray Tracing and Travel-Time Calculation

Seismic data processing usually requires information of theoretical phase travel times. The TauP toolkit is a widely used package for calculating body-wave travel times in the seismology community (Crotwell *et al.*, 1999). However, the TauP toolkit is written in Java. To be compatible with the Crazyseismic software package, we use and extend a MATLAB version seismic travel-time calculator developed by Zheng *et al.* (2007). The code utilizes the tau method (Buland and Chapman, 1983) and calculates travel times and ray paths similar to the TauP toolkit. The theoretical phase travel times predicted by these two packages agree with each other (Fig. 3). The current version of Crazyseismic supports the following seismic phases: direct phases (*P*, *S*); depth phases (*pP*, *sP*, *sS*, *pS*); core diffracted phases (P_{diff}, S_{diff}); multiple surface reflections (*PP*, *PPP*, *SS*, *SSS*); converted phases (*PS*, *SP*); core reflections (*PcP*, *ScS*, *PcS*, *ScP*, *ScSScS*); outer core phases (*PKP*, *PKKP*, *SKS*,
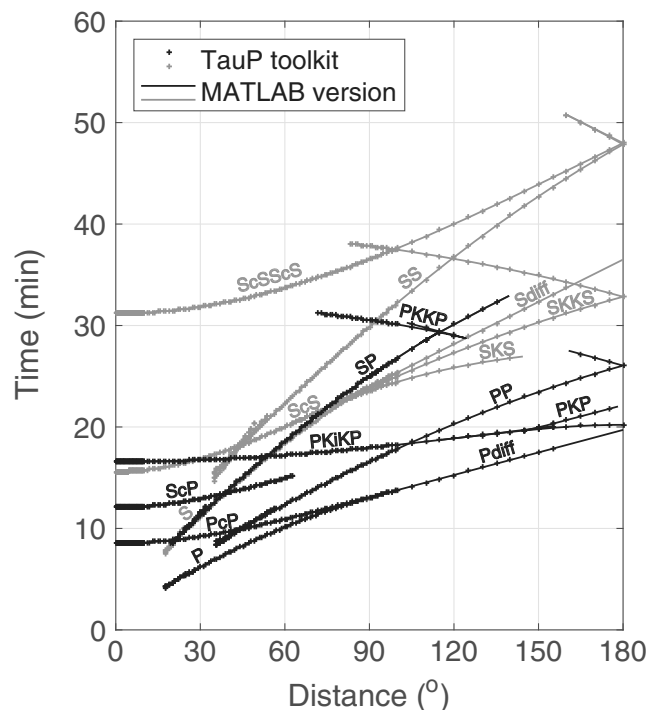
▲ **Figure 3.** Comparison of theoretical phase travel times computed using the TauP toolkit (crosses) and the MATLAB version seismic travel-time calculator (lines). Dark colors mark phases ending with *P*. Gray colors mark phases ending with *S*. The event depth is fixed at the free surface. The reference 1D velocity model ak135 is used for the computation.

SKKS); and inner core phases (*PKiKP*, *PKIKP*). An advantage of the MATLAB version seismic travel-time calculator is its capability to compute travel times related to any hypothetic discontinuities, whereas the TauP toolkit requires true discontinuities in the model. In addition, users can easily modify the code to meet their own specific demands.
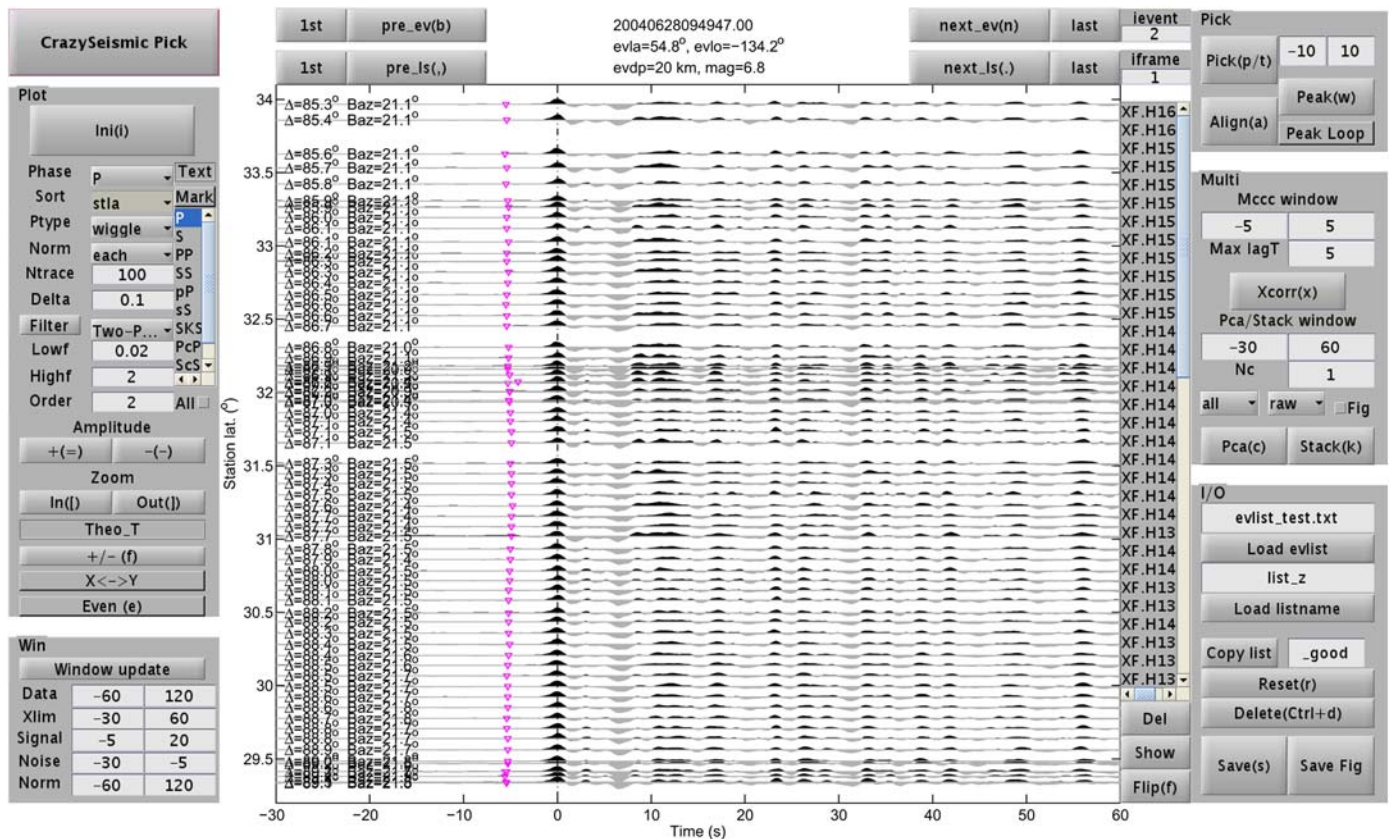
Multichannel Cross Correlation and Phase Arrival-Time Picking

Picking arrival times of seismic phases is essential in seismic tomography. For teleseismic events, seismic waveforms recorded by arrays, such as the EarthScope Transportable Array, often look similar. Under these circumstances, we can use multichannel cross correlation (VanDecar and Crosson, 1990) to determine relative phase arrival times at different stations. Again, to be compatible with Crazyseismic, we developed a MATLAB version multichannel cross-correlation code. It determines relative phase arrival times using least squares. In Crazyseismic, applying multichannel cross correlation can automatically line up seismic waveforms. It can be implemented for either all traces or a selected subset of traces of the event.

Principal Component Analysis and Extracting Source Wavelet

Coherent features of waveforms recorded by a seismic array can be extracted using PCA (Jolliffe, 2002). PCA uses a singular

▲ **Figure 4.** An example showing aligned seismic waveforms after multichannel cross correlation. Arrival times are centered at the maximum peak of the direct *P* wave. Theoretical arrival times are marked by inverted triangles. Seismic waveforms are filtered by a two-pass Butterworth filter with a passband between 0.02 and 2 Hz. The time window used for multichannel cross correlation is −5 to 5 s with respect to picked *P*-wave arrival times. The color version of this figure is available only in the electronic edition.

value decomposition algorithm to decompose seismic waveforms into orthogonal components. The first principal component, corresponding to the largest singular value, gives the most coherent feature contained in all seismograms recorded by a seismic array. Compared with simple stacking, the first principal component also preserves the waveform amplitude of each channel. Higher-order principal components contain more detailed and station-dependent information. Thus, PCA offers a robust way to enhance seismic signal-to-noise ratios and to estimate apparent source wavelets of earthquakes.

Deconvolution

Crazyseismic offers two deconvolution methods: (1) the frequency-domain water-level deconvolution (Langston, 1979; Ammon, 1991) and (2) the iterative time-domain deconvolution (Ligorria and Ammon, 1999). These two methods are widely used in receiver function calculation. The frequency-domain water-level deconvolution is numerically fast to compute but the deconvolution result depends on the choice of the water-level value. The iterative time-domain deconvolution, on the other hand, does not have a regularization term like the water level and can produce spiky/impulsive deconvolved traces with less ringing effects. However, it is more difficult to evaluate noise levels on time-domain deconvolved traces.

Because the result is iteratively updated, the computation time is relatively lengthy. If the signal-to-noise ratio of data is high, both methods should give consistent results.

DEMONSTRATION USING FIELD EXAMPLES

We demonstrate the power of Crazyseismic through two field examples. The first example uses the Pick GUI to pick teleseismic *P*-wave arrival times. The second example uses the Decon GUI to compute *P*-wave receiver functions. Both examples use part of the dense seismic array data recorded by Project Hi-CLIMB (Himalayan–Tibetan Continental Lithosphere during Mountain Building) in Tibet (see Data and Resources). The choice of this array is guided by the high quality of the data and small interstation spacing (Chen *et al.*, 2010). The array consists of ~70 broadband seismic stations located roughly in the north–south direction, across the Lhasa and Qiangtang terrane.

Example 1: Relative *P*-Wave Arrival-Time Picking

First, we prepare the data in SAC format and create both event lists and SAC file lists as shown in Figure 2. Then, we run the Pick GUI and load in corresponding lists through the I/O panel. After this, seismic waveforms are shown in the central window of the GUI. Users can modify various control
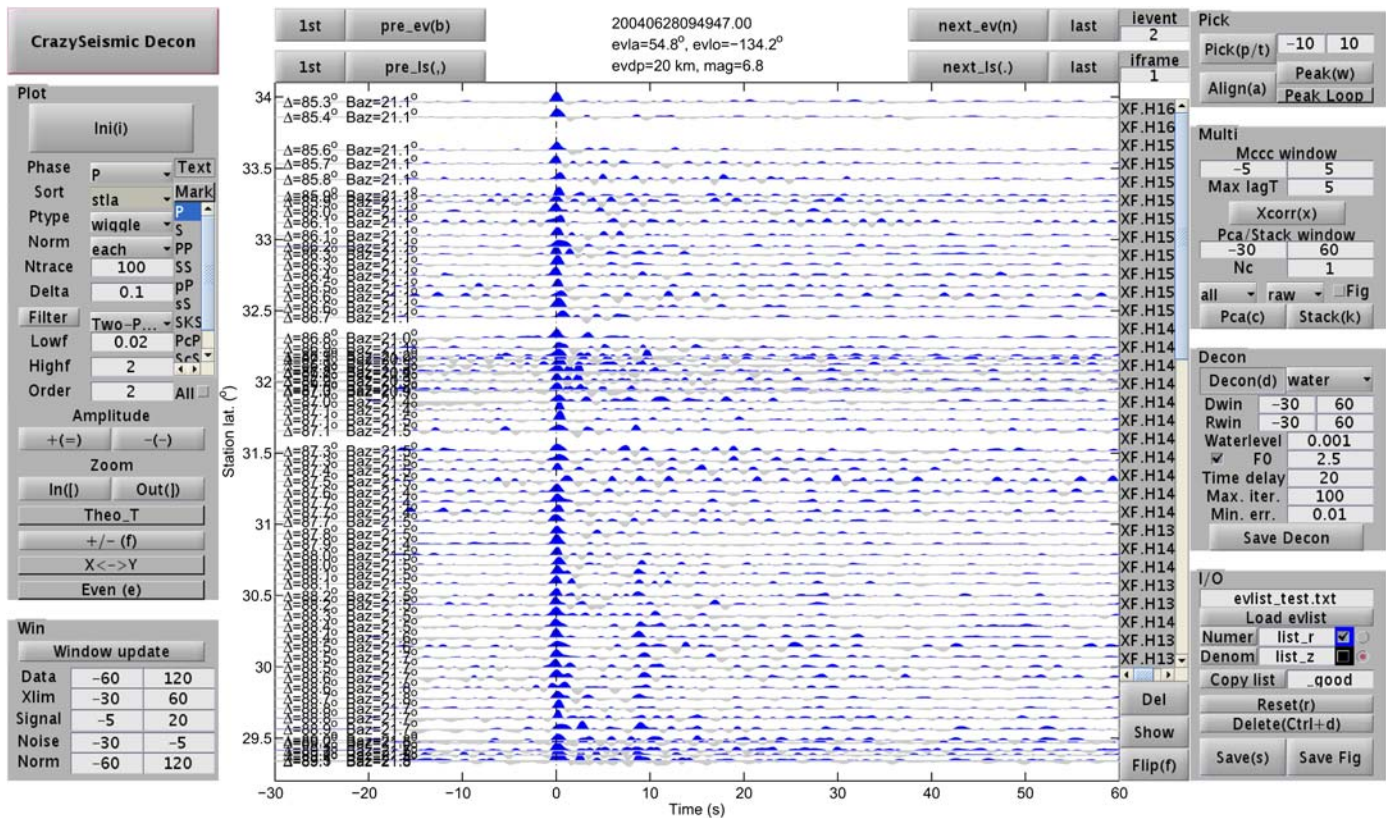
▲ **Figure 5.** An example showing radial *P*-wave receiver functions using Crazyseismic Decon GUI. Receiver functions are computed in the frequency domain using the water-level deconvolution method. A water level of 0.001 and a Gaussian factor of 2.5 are applied. The color version of this figure is available only in the electronic edition.

parameters on the GUI for different viewing purposes. In this example, we load in the vertical-component seismic waveforms and set parameters as shown in Figure 4. We remove noisy or problematic traces. This can be done in the list panel by clicking the corresponding station names and pressing either the DEL button shown on the GUI or the delete button on the keyboard.

Because of the similarity of seismic waveforms, we then pick phase arrival times using multichannel cross correlation by clicking the Xcorr button in the multichannel panel. The time window for multichannel cross correlation and maximum time lags are defined in boxes above the Xcorr button. The outcome is a lineup seismic traces. Absolute *P*-wave arrival times are obtained by shifting all traces systematically left or right by first selecting all in the list panel then using the left/right arrows in the keyboard (hold Ctrl/Shift for minor/major adjustments). Another efficient way of lining up seismic traces is to use the automatic peak function in the pick panel. This function can align seismic waveforms by their maximum peak in the defined window.

After all these steps, users can save the results by clicking the Save button in the I/O panel. The output file contains phase arrival times, ray parameters, station parameters, event parameters, and so on. This output file will be the basis for advanced seismic analysis or inversion, such as seismic tomography.

Example 2: *P*-Wave Receiver Function Calculation

The second example is to compute radial *P*-wave receiver functions for the same event as shown in Example 1. First, we load in both radial- and vertical-component seismic waveforms through the I/O panel. Users can also choose to show two components simultaneously or any component using checkboxes shown in the I/O panel. Arrival times for these two components will be automatically synchronized. Then, we adjust deconvolution parameters in the decon panel and click the Decon button to compute receiver functions. In our case, we use the frequency-domain water-level deconvolution method, with a water-level threshold of 0.001 of the maximum power for the vertical component and a Gaussian factor of 2.5 (corresponding to a low-pass filter with corner frequency ~1.2 Hz). The outcomes of the deconvolution are radial *P*-wave receiver functions shown in Figure 5. Deconvolution results can be saved into the SAC format by clicking the Save Decon button.

The basic processing steps for the above two examples can essentially be repeated for all events recorded by Hi-CLIMB or other arrays. As an estimation of the efficiency of Crazyseismic, it takes approximately one hour for a proficient user to pick *P*-wave arrival times and calculate *P*-wave receiver functions for ~100 teleseismic events recorded by Hi-CLIMB.

CONCLUSIONS AND DISCUSSIONS

We introduce the Crazyseismic software package for passive seismic data preprocessing. Crazyseismic is based on the MAT-LAB GUI, which is a high-level programming language and an interactive environment widely used by researchers in many disciplines. Major features of this software include its simplicity and high efficiency. All functions are unified into one single GUI interface, so that users can process data by simply using mouse click or hot keys. The embedded phase travel-time calculator and multichannel cross-correlation function not only accelerate arrival-time picking but also improve the accuracy of relative arrival-time measurements. In our current version, Crazyseismic works for a majority of body-wave phases. Essentially, it can be extended to any seismic phases defined by users.

Crazyseismic is an open-source software package. Users can modify the package to meet their specific demands. Moreover, the idea of Crazyseismic can be applied to develop similar software in other fields of seismology and other disciplines in geosciences. We hope that Crazyseismic can be a valuable software package for the seismological community and beyond.

DATA AND RESOURCES

Seismograms used in this study were collected from Project Hi-CLIMB (Himalayan–Tibetan Continental Lithosphere during Mountain Building). Data can be obtained from the Incorporated Research Institutions for Seismology-Data Management Center at www.iris.edu (last accessed October 2016). Coral is retrieved from http://earthweb.ess.washington.edu/creager/ (last accessed October 2016). SEIZMO is retrieved from http://epsc.wustl.edu/~ggeuler/codes/m/seizmo/ (last accessed October 2016). Crazyseismic can be downloaded from http://web.gps.caltech.edu/~yucq/software.html (last accessed October 2016). The commercial platform MATLAB is from Mathworks website http://www.mathworks.com (last accessed October 2016). ⧉

ACKNOWLEDGMENTS

REFERENCES

Ammon, C. J. (1991). The isolation of receiver effects from teleseismic *P* waveforms, *Bull. Seismol. Soc. Am.* **81,** no. 6, 2504.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.* **81,** no. 3, 530–533.

Buland, R., and C. Chapman (1983). The computation of seismic travel times, *Bull. Seismol. Soc. Am.* **73,** no. 5, 1271–1302.

Chen, W.-P., M. Martin, T.-L. Tseng, R. L. Nowack, S.-H. Hung, and B.-S. Huang (2010). Shear-wave birefringence and current configuration of converging lithosphere under Tibet, *Earth Planet. Sci. Lett.* **295,** nos. 1/2, 297–304, doi: 10.1016/j.epsl.2010.04.017.

Cohen, J., and J. Stockwell Jr. (2000). *CWP/SU: Seismic Unix Release 33: A Free Package for Seismic Research and Processing*, Centre for Wave Phenomenon, Colorado School of Mines, Golden, Colorado.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999). The TauP toolkit: Flexible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.* **70,** no. 2, 154–160.

Goldstein, P., D. Dodge, M. Firpo, and L. Minner (2003). SAC2000: Signal processing and analysis tools for seismologists and engineers, in *The IASPEI International Handbook of Earthquake and Engineering Seismology*, Academic Press, London.

Harris, J. M., and C. J. Young (1996). *MatSeis: A Seismic Toolbox for MATLAB*, Sandia National Labs., Albuquerque, New Mexico.

Jolliffe, I. (2002). *Principal Component Analysis*, Second Ed., Springer, New York, New York.

Langston, C. A. (1979). Structure under Mount Rainier, Washington, inferred from teleseismic body waves, *J. Geophys. Res.* **84,** no. B9, 4749–4762.

Ligorria, J. P., and C. J. Ammon (1999). Iterative deconvolution and receiver-function estimation, *Bull. Seismol. Soc. Am.* **89,** no. 5, 1395–1400.

Lou, X., S. van der Lee, and S. Lloyd (2013). AIMBAT: A python/matplotlib tool for measuring teleseismic arrival times, *Seismol. Res. Lett.* **84,** no. 1, 85–93.

VanDecar, J., and R. Crosson (1990). Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, *Bull. Seismol. Soc. Am.* **80,** no. 1, 150–169.

Wustefeld, A., G. Bokelmann, C. Zaroli, and G. Barruol (2008). SplitLab: A shear-wave splitting environment in Matlab, *Comput. Geosci.* **34,** no. 5, 515–528.

Zheng, Y., T. Lay, M. P. Flanagan, and Q. Williams (2007). Pervasive seismic wave reflectivity and metasomatism of the Tonga mantle wedge, *Science* **316,** no. 5826, 855–859.

Chunquan Yu[1]
Xuefeng Shang[2]
Department of Earth, Atmospheric and Planetary Sciences
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139 U.S.A.
yucq@gps.caltech.edu
xuefeng.shang@gmail.com

Yingcai Zheng
Department of Earth and Atmospheric Sciences
University of Houston
Houston, Texas 77004 U.S.A.
yc.zheng@gmail.com

[1] Now at Seismological Laboratory, California Institute of Technology, 1200 E California Boulevard, MC 252-21, Pasadena, California 91125 U.S.A.
[2] Now at Shell Oil Company, Houston, Texas 77079 U.S.A.