PUKE files

December 17, 2013

Eric Bergman
Global Seismological Services
bergman@seismo.com

The multiple event relocation code **mloc**, which implements the Hypocentroidal Decomposition (HD) method of Jordan and Sverdrup (1981), creates a variety of output files, some by default, others optionally, that have been designed to serve various needs. The **puke** file (filename suffix ".puke") is an optional output file (controlled by the "puke" command) that is well-suited to transmitting the arrival time data along with the hypocentral data in a form that maintains information on which specific data were used for both the hypocentroid and the cluster vectors. It also includes the location and elevation coordinates for the station associated with each reading. This resolves many sources of confusion and error when station information is supplied separately from arrival time information.

The **puke** format is derived from a similar file format, called a "pick" file, that E.R. Engdahl developed as a means of transmitting the data from his single event relocation ("EHB") catalogs to tomographers and others who needed access to the arrival time data, including the flags that determined which data were used. The pickfile format would not suffice for the information appropriate to an HD analysis, and so the **puke** format was developed to serve the same function for **mloc**.

The **puke** format as defined in **mloc** version 9.9.8 (release date December 16, 2013) is documented here.

Version number of a puke-formatted file

The version and release date of **mloc** that was used in any given relocation is carried in the first line of the **summary** file (filename suffix ".summary"). **Puke** files do not carry any version information, but they can be associated with the corresponding version of **mloc**. As with many of **mloc's** output files, the interpretation of certain data fields is dependent on whether the **puke** file is the result of a calibrated or uncalibrated relocation, and which type of calibration was used.

Basic structure

A puke file consists of a series of event blocks, each followed by a blank line. Each event block consists of a single hypocenter line of 147 characters followed by a series of phase reading lines, 107 characters long.

Hypocenter line format

Columns	Variable (format)
1:4	Calibration code (a4)
6:9	Origin year (i4)
10:11	Origin month (i2)
12:13	Origin day (i2)
15:16	Origin hour (i2)
17:18	Origin minute (i2)
19:23	Origin seconds (f5.2)
25:28	Origin time uncertainty (f4.2)
30:36	Geographic latitude (f7.3)
38:45	Geographic longitude (f8.3)
47:51	Focal depth, km (f5.1)
53:56	+ uncertainty in depth (deeper), in km (f4.1)
58:61	- uncertainty in depth (shallower), in km (f4.1)
63:67	Standard error of observations in cluster analysis, in seconds (f5.2)
69:73	Azimuth of semi-minor axis (f5.1)
75:78	Semi-minor axis length, km (f4.1)
80:84	Azimuth of semi-major axis (f5.1)
86:89	Semi-major axis length, km (f4.1)
90:93	Number of defining phases for hypocentroid (i4)
94:97	Number of stations with defining phases for hypocentroid (i4)
99:103	Open azimuth based on defining phases for hypocentroid (f5.1)
105:109	Closest (defining) station distance (deg) for hypocentroid (f5.1)
111:115	Furthest (defining) station distance (deg) for hypocentroid (f5.1)
116:119	Number of defining phases for cluster vector (i4)
120:123	Number of stations with defining phases for cluster vector (i4)
125:129	Open azimuth based on defining phases for cluster vector (f5.1)
131:135	Closest (defining) station distance (deg) for cluster vector (f5.1)
137:141	Furthest (defining) station distance (deg) for cluster vector (f5.1)
143:145	Preferred magnitude (f3.1)
146:147	Magnitude scale (a2)

Notes on the hypocenter line

Calibration Code

I have recently developed a code (GTCNU) to carry information about the calibration status of an event to replace the widely-abused “GTX” formulation introduced by Bondar et al. (2004). This subject is documented elsewhere. **mloc** calculates this code based on the nature of the relocation and calibration data available. The calibration code includes a scale parameter (location accuracy level).

Uncertainty of Origin Time

In seconds. For uncalibrated clusters and direct calibration, it is taken from the covariance matrix of the relocation, including uncertainty of hypocentroid and cluster vector. For indirect calibration it is based on the uncertainty of the calibration shift plus the uncertainty of relative origin time from the cluster vector.

Uncertainty in Focal Depth

Uncertainty in focal depth can be read from the input file, from the command file, or from the relocation. Uncertainty in focal depth is carried in two fields because it is not uncommon for the uncertainty to be asymmetric. This can arise when inferring depths from teleseismic depth phases and there is uncertainty about the correct identification of the phase (pP, sP, or pwP). It can also arise in waveform analyses where the error vs depth curve is not symmetric. If focal depth has been a free parameter in the relocation it will be symmetric, and it will over-ride any specification in the input file and command file.

For uncalibrated clusters and direct calibration with a free depth solution, the uncertainty is taken from the covariance matrix of the relocation, including uncertainty of hypocentroid and cluster vector. For indirect calibration it is based on the uncertainty of the calibration shift plus the uncertainty of relative depth from the cluster vector. If no estimate of uncertainty in focal depth is available, the default value of 99.9 is printed. In the next version of the format, this will likely be changed to a blank field.

Confidence Ellipse

Confidence ellipses (90% confidence level) for the epicenter are given in four columns which give the azimuth and length of each semi-axis (half-length) for the ellipse. The shorter semi-axis is given first. Azimuth is in integer degrees, clockwise from North. Semi-axis lengths are given in decimal km.

For an uncalibrated HD analysis (calibration code will be type “U”), the confidence ellipse of the cluster vector is given. This represents the uncertainty of location in a relative sense, with respect to the other events in the cluster. The formal uncertainty of an uncalibrated hypocentroid is not provided in a **puke** file, but it can be found in the .summary file. For a calibrated relocation, the confidence ellipse in a **puke** file gives the full uncertainty, that is, the sum of uncertainty for the hypocentroid and the cluster vector. If direct and indirect calibration have both been used, indirect calibration takes precedence.

Open Azimuths

The largest open azimuth for the readings of this event used to estimate the hypocentroid (of the cluster) and cluster vector are given. There is little significance to the value with respect to the hypocentroid, unless nearly all the close-in readings used for direct calibration are associated with a single event. Normally, the data for direct calibration of the centroid are distributed among many events, so the azimuthal distribution can only be understood through review of all the data.

Epicentral Distance Ranges

The epicentral distance, in degrees, of the nearest and farthest reading from this event used for the hypocentroid and also for the cluster vector are given. When direct calibration has been used, it is common for many events to contribute no readings for estimation of the hypocentroid.

Magnitude and Magnitude Scale

A single representative (or preferred) magnitude is carried through the **mloc** analysis to assist in interpretation. It has no bearing on the relocation. MNF input files can carry multiple estimates of magnitude, but one will be specified (or selected by default) as the preferred magnitude. If no magnitude estimate is available, a magnitude of 0.0 will be printed. In the next version of the format, this will likely be changed to a blank field. In some cases a magnitude may be available but not a scale, in which case the magnitude scale field will be blank.

Phase line format

Columns	Variable (format)
1:5	Station code (a5)
7:14	Station latitude (f8.4)
16:24	Station longitude (f9.4)
26:30	Station elevation, in meters (i5)
32:37	Epicentral distance, degrees (f6.2)
39:41	Azimuth, event to station (i3)
43:50	Phase name (a8)
52:55	Arrival time year (i4)
56:57	Arrival time month (i2)
58:59	Arrival time day (i2)
61:62	Arrival time hour (i2)
63:64	Arrival time minute (i2)
65:70	Arrival time seconds (f6.3)
72:77	Reading error, seconds (f6.2)
79:86	Observed travel time, seconds (f8.2)
88:95	Travel time residual, seconds (f8.2)
97:104	Source (author) of the reading (a8)
106:106	Defining phase for hypocentroid (“y” or “n”) (a1)
107:107	Defining phase for cluster vector (“y” or “n”) (a1)

Notes on the phase line

Arrival time

The **puke** format carries no information on the precision of arrival times. This will probably be changed in the future.

Reading error
In most cases reading errors are "empirical reading errors", based on an analysis of the spread of residuals for a given station-phase. If there is only one instance of a certain station-phase, a default value is used. See the .rderr file for more information.

Observed travel time
The difference between the arrival time and the origin time given in the hypocenter line.

Travel-time Residual
The **puke** file does not carry any information about the travel-time model used to calculate theoretical travel times, and thus, residuals. This information can be found in other output files, notably the .summary file and the .log file. In the case where a travel time could not be calculated for a reading the residual will be 999 seconds.

Source of the reading
The concept of source or authorship for a reading has been somewhat casually observed for a long time, but it is becoming more firmly established in the main seismological databases. It is not required and is not always available in a reliable form. If present it should be treated as informational, but not forensic.

Defining phase for hypocentroid
For calibrated locations using the direct method, the selection of data to be used to estimate the hypocentroid is critical. Usually only a small subset of the available readings are used, those at very short epicentral distance. For uncalibrated relocations or for indirect calibration, the choice of data for estimating the hypocentroid is not very significant. Typically, it will be teleseismic P arrivals.

Defining phase for cluster vector
In general, any phase at any epicentral distance can (and should) be used to estimate the cluster vector because that estimation process is based only on travel-time differences. The main reason why a reading would not be used is if it has been identified as an outlier through the analysis of the spread of residuals for that station-phase. Sometimes, it may be desirable to exclude arrivals in certain epicentral distance ranges (e.g., the Pn/P transition or around the PKP caustic) because of confusion concerning phase identification.