MNF Data Format for Differential Time Data in *mloc*

Eric Bergman
Global Seismological Services
March 18, 2016
v1.5.0

This document defines the format for input of differential time data for use in *mloc*. The format is consider part of the MNF (*mloc* native format) family of data formats, and is assigned version 1.5.0

Background

The multiple event relocation code *mloc* was initially designed to use arrival time data only, but like most multiple event relocation codes, arrival time data is converted to differential times (i.e., the same phase at a specific station, observed from two or more events) to obtain improved resolution of the relative locations of events in the cluster. As the use (and evident value) of differential time data from waveform cross-correlation has increased in earthquake location algorithms, *mloc's* code was modified to allow explicit input of differential time data. This change was first implemented in v6.0 of *mloc*, in September 2006. At that time, the only source of waveform cross-correlation data for *mloc* was from Benjamin Kohl of SAIC and code was written to read the specific data format he provided.

Later, differential time data from waveform cross-correlation was provided by Michael Begnaud of LANL for a special study, and the code used to read Kohl's data format was hastily modified to work with the Begnaud's data. Currently, Will Yeck of NEIC is beginning to produce differential time data for use in *mloc* and it seems desirable to stabilize the file format for this kind of data, rather than create yet another subroutine to read a specific researcher's format.

Parameterization of Differential Time Data

There are several ways in which differential time data can be reported. The MNF v1.5 format supports only one form, "reduced relative arrival time". To understand this parameterization of differential time, it is helpful to think of measuring the time lapse between the arrival times at some station of two phase groups of the same type (say, teleseismic P) with cross-correlation and reporting the full time (in seconds), including the years, months, and days between the two events being compared. We call this the "relative arrival time". If we ignore the differences in years, months, and days for this calculation, and treat the two events as if they occurred on the same day, then the time difference between the two phase groups is considered "reduced" and we have the "reduced relative arrival time". This value will be less than 86,400 (in absolute value), the number of seconds in one day, but it may be positive or negative, depending on which event is taken as the reference or "template" event:

reduced_relative_arrival_time = reduced_target_arrival_time – reduced_template_arrival_time

where the template and target event arrival times have been reduced to the same day.

For use in *mloc*, differential time data provided in the reduced relative arrival time parameterization is converted into dummy arrival times for the template and target events:

template_dummy_arrival_time = template_origin_time + theoretical_TT
target_dummy_arrival_time = template_dummy_arrival_time + reduced_relative_arrival_time

where the theoretical travel-time is calculated for the phase of interest at the appropriate station, using the current best estimate of the hypocenter of the template event (with origin time in the reduced sense). The initial origin time of the template event used in this formulation can be provided internally in *mloc* in several ways, but the underlying differential arrival time datum is preserved.

This is a natural approach in *mloc* because all arrival time data are processed in "reduced" form, i.e., as if all events occurred on the same day. The dummy arrival times are combined with the other, genuine arrival time data available for the template and target events. Dummy times for differential data are only used for estimating the cluster vectors, so only their time difference is used and the potential bias from the assumed initial origin time of the template event and the theoretical travel-time calculation are removed by differencing.

Description and Features

Like other varieties of MNF, v1.5 uses several types of 'records' with defined formats and a character in column 1 that defines the type of record. Unlike the other MNF formats, which are event-oriented (i.e., records are grouped by event), v1.5 data files are cluster-oriented, and each differential time record carries references to two events in the cluster. Only 4 record types are defined:

Flag	Record Type	Minimum Length	Full Length
F	Format Version	14	14
D	Differential Time	87	149
EOF	End of File	3	3
#	Comment	1	149

Unlike all other record types, which are distinguished by the flag in column 1, the **end-of-file record** ("EOF") uses columns 1:3; it has no other arguments. It causes processing of a data file to end, so it would normally only be found once, at the end of the file. If an **end-of-file record** is placed in the middle of a file, processing will stop there.

The **differential time record** carries a "usage" flag (in column 3) that determines the way (or if) the information in that record will be used by *mloc*. None of the other record types carries a usage flag.

Comment records (flag "#") can be inserted anywhere in an MNF-formatted file, but should not be the first or last record. Obviously, any text that occurs after the first **end-of-file record** will not be processed and can be considered as a comment, regardless of the formatting, but the use of the **end-of-file record** in this manner is not recommended.

Information about the nature of the differential time data set can be carried in **comment records**.

File Structure

The first line in a data file must be a **format record**. This is followed by some number of **differential time records** (and **comment records**, optionally) and the file must end with an **end-of-file record**.

Defined Record Types

The concept of "optional" fields in the descriptions of record types is specifically in the context of use by *mloc*. Fields that are required by *mloc* are printed in bold below. In writing MNF-formatted data files it is advisable to pad lines to the full length of that record type, which is based on the defined fields, not the required fields, and it is not unwise to pad all lines to 149 characters, the full length of the longest defined record types, regardless of record type.

Format Version Record ("F" in column 1)

Column	Description
1	**Line format flag "F"**
10-14	**Format version (a5)**

It is useful for readability to include extra text, such that columns 1:8 read "F   MNF v", but all that is required is the "F" in column 1 and the version number in columns 10:14. Version number is treated by *mloc* as a character string that can accommodate three-level versioning (e.g., ‘1.5.1’).

Differential Time Record (“D” in column 1)

Column	Description
1	**Line format flag “D”**
3	Usage flag (a1), optional
5:20	**Template Event Designator (a16)**
22:31	Template Event EVID (a10), optional
33:48	**Target Event Designator (a16)**
50:59	Target Event EVID (a10), optional
61:66	**Station (a6)**
68:75	**Phase (a8)**
77:87	**Reduced Relative Arrival Time (f11.4)**
89:90	Reading Precision (i2), optional
92:97	Uncertainty, s (f6.4), optional
99:103	Correlation Coefficient (f5.3), optional
105:112	Original phase name, optional (a8)
114:118	Agency, optional (a5)
120:127	Deployment or network, optional (a8)
129:133	Station, optional (a5)
135:136	Location, optional (a2)
138:140	Channel, optional (a3)
142:149	Author, optional (a8)

The usage field is used here the same way as it is used in MNF v1.3 phase records to flag certain readings so that they will not be used. Some common values that could be relevant to differential time data are:

Value	Meaning
x	outlier, do not use
d	duplicate reading, do not use
m	missing station coordinates, do not use
p	phase that is unknown or not used

The “event designator” field for the template and target events carries an identifier for each event based on an estimate of the date and origin time, to the nearest second. The format is:

yyyymmdd.hhmm.ss

with the character ‘0’ filling any blanks. Although the identifier could be read as numeric values for year, month, day, etc, *mloc* treats it as a character string that is only used to make the correct association with two corresponding events in the cluster being relocated. An equivalent event identifier is declared in the *mloc* command file for each event, using the ‘even’ command. Therefore it is desirable to ensure that the integer seconds part of the identifiers match. The evid (event id) fields are provided for the same reason. While evids are not always available, if they are they provide a more robust way to make the association with events in the cluster. *mloc* attempts to make the match using evids first.

The definition of ‘reduced relative arrival time’ is discussed above.

Reading precision is an integer that indicates the number of significant decimal places in the reduced relative arrival time datum. It is used in *mloc* to correctly format output. The values appropriate for differential time data are:

Value	Meaning
0	nearest second
-1	nearest tenth
-2	nearest hundredth
-3	nearest thousandth
-4	nearest ten-thousandth

If no value for reading precision is provided, *mloc* attempts to determine it from the formatting of the datum. No great harm will be done if this process is inaccurate, but specification of a reading precision is strongly recommended as good practice.

If an uncertainty for the estimate of reduced relative arrival time is provided, it will be read and may be used in *mloc*, at least initially, although it can be over-ridden in various ways. It is desirable to have an estimate of uncertainty from the cross-correlation analysis but *mloc* can proceed without one.

If the correlation coefficient field is non-blank, it is read by *mloc*. It can be used on input to filter out differential times with correlation coefficients below a threshold set by the user. It may also be helpful in deciding whether or not to flag a particular datum as an outlier.

The “original phase name” field provides a way for the user to change the phase name of a differential time datum without losing the information on the original phase ID. The same facility exists in the MNF v1.3 format for arrival time data, but this is likely to be a rare need with differential time data.

The next five fields, all optional, identify the seismograph instrumentation from which the measurement of reduced relative arrival time was made, in the new IASPEI Station Coding Standard. Known as the ADSLC code, this identification standard greatly expands the traditional concept of describing the source of seismic data by a single 3-5 character station code. See the documentation for MNF v1.3 for a fuller discussion of this, especially the rationale for carrying station code in two places in the record. Channel information may be of special significance with differential time data because it is common practice to carry out the cross-correlation analysis on more than one channel of a station and select the one with the highest correlation coefficient for use in relocation. Therefore differential time data sets may well include multiple estimates of reduced relative arrival time, differing only in the channel that was analyzed.

The author field identifies the source (usually a person) of the datum. The field is optional but it is strongly recommended that it be utilized. The character string employed to identify someone is completely arbitrary, but it is limited to 8 characters in *mloc*.

Comment Record ("#" in column 1)

Column	Description
1	**Line format flag “#”**
2:149	Comment, optional (a148)

The length of the comment field is rather arbitrary but it is convenient to limit the **comment record** to the length of the main ‘data’ record, in this case the **differential time record**. *mloc* ignores **comment records**.

End of File Record ("EOF" in columns 1-3)

Column	Description
1:3	**Line format flag “EOF”**