TOWARD A COMMON FORMAT FOR STRONG-MOTION DATA

A.F. Shakal, California Strong Motion Instrumentation Program,
Calif. Dept. Conservation/Div. Mines & Geology
and

R.D. Borcherdt, National Strong Motion Program,
United States Geological Survey

Abstract

The need for a common format for the earthquake engineering use of strong-motion data has become increasingly apparent during the last several years. An early goal of the Consortium of Organizations for Strong Motion Observation Systems is the development of a consensus format for data products. The current number and variety of formats arose largely through the nature of the growth of strong motion recording and processing. As new networks or processing facilities began many produced data in a format of their own convenience since no standard had been established. As a first step toward developing a common format, a Format Working Group met at PEER in January 1999.

The introduction of new standard formats should bear positive results in data exchange for both the data-producing organizations and the data users. A common format does not bar the use of individual formats by data producers, but rather provides for a common, minimum basic format for use of strong-motion data in earthquake engineering. Once a common format achieves adequate consensus, converters are planned to be made available at the COSMOS Virtual Data Center which will perform translations of data formats, so that from a user perspective, the data will appear to be all of one "virtual" format.

Introduction

Analysis and study of strong-motion data has been important since the first strong-motion records were obtained in the 1933 Long Beach earthquake. Early strong-motion records were processed by various means, including manual numerical integration and mechanical analyzers. As digital computers became more common in the 1960s, attempts to process the records using digital computers became common. Several important records had been recorded, including the 1940 El Centro record, which led to a variety of studies but the digitized data was not distributed in a standard format. Since those beginnings, a large number of records have been recorded. A large number of data formats have also come into use, making the translation of data between formats cumbersome and limiting. As more data becomes available, either more translation will be necessary, or the data will not be put to its fullest use. This makes conversion, sometimes called filtering, between formats more and more important.

Conversion or Filtering Between Formats

Before considering individual formats or possible standards, it may be beneficial to consider a framework for the release of data in a common format. This will also put into perspective the role of individual formats in this framework.

A schematic of the interaction between existing formats and a common format is shown in Figure 1. In this scenario, data produced by networks or researchers would be converted to a common format, either at their facility or at the COSMOS Virtual Data Center. Converters, or filters, would be built that perform the conversion between the original formats and the common format. Once the data can be converted in this way to the common format, everyone can use the data without further difficulty.

An additional aspect is derivative-product formats. For example, a post-converter could be made available that would convert the standard format to that needed, for example, by SAP, for structural analysis projects, or by Shake, for site response projects.

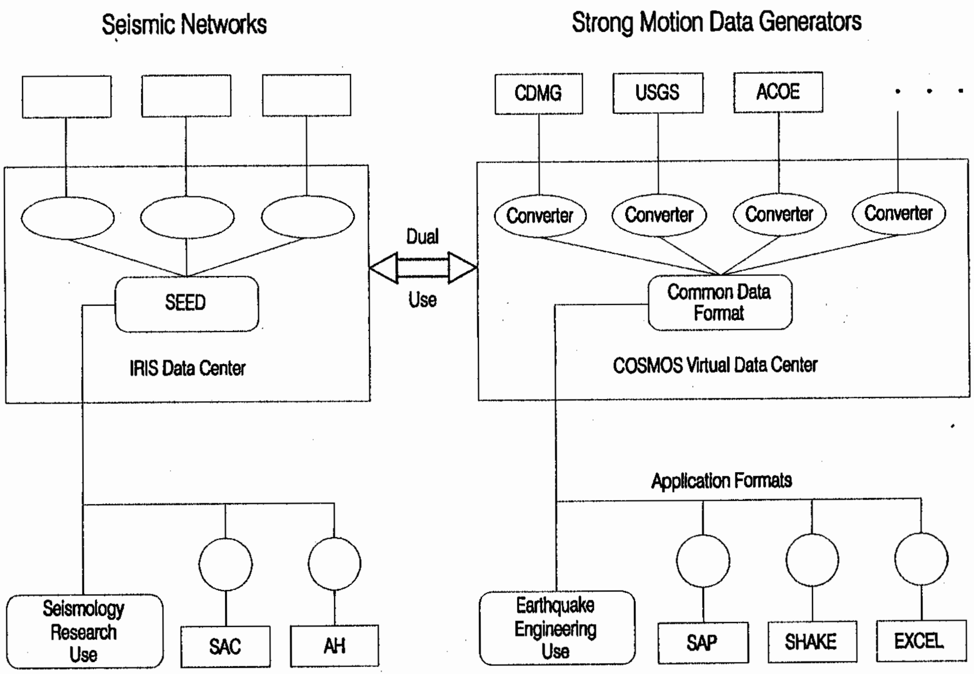

Figure 1. Schematic diagram showing the proposed interaction between strong motion network data production and a common strong motion format on the right, the parallel arrangement in place for seismic data on the left, and the connection between the two.

The most important secondary conversion would be the conversion to and from the main format used by IRIS (Institutions for Research in Seismology) and other seismological networks, called SEED (Standard for Exchange of Earthquake Data). The use of SEED by IRIS for seismology research data in many ways parallels the framework proposed for strong-motion data conversion and use in earthquake engineering.

A common format allows convenient "dual use" of recorded data. If data recorded by strong-motion networks can be converted between SEED and the strong motion format, the path is paved for dual use of the data recorded. That is, low level data recorded by traditional strong motion networks, for whom the data may only be of secondary interest, can, through the use of a common format, easily be converted to SEED, and from that to SUDS, GEOS, or any other format, making the data convenient for effective utilization by a wide range of investigators. Similarly, data recorded by seismic networks with broadband instruments that record data of potential value in earthquake engineering can convert the data in a single translation, from SEED to the common strong motion format, and the data will be available for use in earthquake engineering. This linking between a common strong motion format and a common seismology format is important for the advance of both earthquake engineering and seismology research.

A necessary component for the model in Figure 1 to come into being is the development of a common format for strong motion data that is accepted by an adequate number of data producers. In order to move toward that goal, it is useful to review the existing formats and their evolution.

Standardized Processing - Caltech Blue Book Project

Until the late1960s, there was no standard processing or formatting of strong-motion records. The Coast and Geodetic Survey obtained many records, but a need developed for standardized processing, so that investigators would be analyzing the same time histories and spectral inputs in various engineering and or response studies. The records obtained from El Centro in 1940, Kern County in 1952, and Parkfield in 1966 made clear the need for uniform processing, particularly because of the unexpected amplitudes and spectral levels

A project initiated in 1969 at Caltech with National Science Foundation funding focused on computer processing of all records available at the time in a standardized way (e.g., Hudson et al., 1969; 1971). The San Fernando earthquake occurred during the course of the project and caused a large increase in the number of records and included the largest motions recorded up to that time.

The Caltech project was very productive, and the series of reports produced, all in blue cover, gave the project its unofficial name, "Blue Book". The complete results were released by means of printed reports as well as by computer tapes and cards, which allowed major progress by many investigators performing analytical studies of the data. The records processed during this period were the foundation of many studies in the following years.

Standard Data Products

The results of processing were released in the Blue Book project at several specific processing stages and the names for these have become traditional:

- Volume 1 Raw acceleration as digitized, usually given as acceleration-time pairs, and expressed in units of acceleration; no instrument correction or filtering applied. The background and original output format are described in Hudson et al. (1969).

- Volume 2 Processed or "corrected" acceleration, velocity and displacement; the final time-history product. The acceleration, velocity and displacement all have constant time steps, though they may be different. Important parameters describing the processing steps are also included. The original format of these files is given in Trifunac and Lee (1973).

- Volume 3 Spectral data, including response spectrum values for five damping values (0, 2, 5, 10, and 20%) and 91 periods (from 0.04 seconds to 15.5 seconds). Spectral acceleration (Sa), velocity (Sv) and displacement spectra (Sd), and pseudo acceleration (PSA) and velocity (PSV), and Fourier amplitude spectra are included in the file.

Several additional numbered products were generated during the Caltech project, but they are not commonly used today. Nearly all strong motion programs, worldwide, currently generate these three primary products. These three products are sometimes denoted as Phase 1, 2, 3 rather than Volume 1, 2, 3 data (e.g., Brady et al., 1980).

One of the benefits of the Blue Book project was that all the data was released with a standard format. As a result, when the project ended the vast majority of strong motion data, world wide, was available in a single format. For the first time, many investigators could study all of the data, and with a single program analyze records from many earthquakes, recorded by many agencies.

As a result of the growth of networks and processing centers, by early 1990s the situation was much changed from the early 1970s. There was now a variety of formats, some very close to the Blue Book format, and some quite different. The format used by CDMG is given in Shakal and Huang (1985), and Brady and Converse (1992) discuss a format used by the USGS.

This problem was an early focus of the Consortium of Organizations for Strong Motion Observation Systems (COSMOS), and in fact was one of the factors motivating the creation of the consortium. The creation of COSMOS, and the initial development of a Virtual Data Center described in these proceedings, puts strong motion data exchange on the threshold of new effectiveness and user convenience.

The Virtual Data Center being developed will allow convenient access to data recorded by different networks. But the current individual formats will limit and hold back the possible advances. There is a clear need for a common, standard format. For a common format to be established, certain basic properties are suggested.

Properties of a Standard Strong Motion Format

A successful common strong motion data format for earthquake engineering use needs to have a certain set of properties, and several of these properties are suggested below.

1. The format must include a minimum set of information about the record, the recording station, the recording instrument, and the causative earthquake.

2. Data in the format should be readily convertible to the most common formats used in seismology research (SEED, or mini-Seed), and thus to other seismological formats (SUDS, GEOS, SAC, etc.). The data should also be readily convertible to common engineering applications formats (SAP, Shake, etc.) as well basic applications tools like spreadsheets (Excel).

3. The format should include an adequate amount of text and information at the beginning to provide the user key information and assurances about the data

4. The format should have easily accessible information for use in metadata information collections and databases.

5. The format should share as much as possible with the legacy formats, and be consistent with past evolution. The format should also be convertible back to the old formats, so legacy programs can use newly recorded data.

6. The format should have adequate resolution, or precision, so there is no loss of accuracy in the data being written, to control noise source and propagation. The format needs to handle very small motion, since current earthquake engineering research often extends to very low levels of motion (e.g., studies of linearity).

7. The format should be flexible, to allow a range of inputs and allow tomorrow's data, if possible. It should accommodate physical parameters in addition to acceleration (relative displacement, pore pressure, etc.) at the Volume 1 level.

8. The format should, so far as possible, meet the needs of groups across the strong motion community, world wide, and allow the expression of aspects which are unique to specific seismic and political environments.

Early strong motion format design in the Blue Book project was quite deliberate, with the needs and convenience of the users firmly in mind. It is suggested that this focus on user needs and convenience should be preserved in arriving at a new common format.

Structure of Proposed Format

Consistent with the original Blue Book format design, it is suggested that the file format for Volume 2 data, the most commonly used data, have the two main sections: a header followed by data, with the header made up of three parts: text, integer, and real valued:

A. Header Section
   1. Text section
   2. Integer values section
   3. Real values section
B. Data Section
   1. Acceleration segment
   2. Velocity segment
   3. Displacement segment

A1. Text Section of Header

The purpose of the text header is to provide information that is intrinsically textual in nature, for use in titles and labels of plots, and for human reading. It is proposed that the length and information content in this section be at the discretion of the data generator, as long as the required basic information is present. The required fields, such as station name, earthquake name, etc., need to be present and in the prescribed locations. The first line of the text section should include the number of lines contained in the text section. For reference, a text header from the Blue Book series is included in Appendix A.

A2. Integer Section of Header

The integer section provides items of information about a record that are intrinsically integer. The data-generating agency may include as many parameters as desired, provided that a minimum set of defined parameters are present. The Blue Book format allowed for 100 integer parameters at the Volume 2 stage. In any case, the number of parameters should be given as one of the parameters. A list of proposed integer parameters is included as Appendix B.

A3. Real Section of Header

The real-values section provides information that is intrinsically real. As for the integer-values section, the data-producer may include as many parameters as desired, provided a minimum set of parameters are present. A proposed list of real-valued parameters is given in Appendix C. Many of the parameters also appear in previous formats, but some need to be augmented or redefined for the new digital data types.

The data follows the header section, and the three data segments comprise the greater part of the data file. The first line of each data segment contains basic information about the data points that follow in the file:

- the number of data points in the segment, N;
- the sampling rate (no. of points per second);
- the physical units of the data, and
- the numeric format in which the data is written, n points per line.

Note that including the numeric format here allows the format to be specified to match the precision or accuracy of the data itself. For example, some formats use a format of 10F8.3 (i.e., 10 values per line, each 8 digits wide, with 3 decimal digits of precision for acceleration). Thus, values could range from +/- 1.000g down to +/-0.001g. For modern digital recorders, this resolution is not adequate. But with a dynamically-specifiable fixed numeric format, the data producer can write the data in the most appropriate format, as 5F15.7 for example, preserving the resolution. The user's program must then read the format before it reads the data. With this approach, flexibility and preservation of accuracy is maintained.

Overall File Structure

It has been suggested that a strong motion data file contain data for only one component of an accelerogram. That is, the accelerogram from a 3-channel freefield accelerogram would be contained in three files, one for each component. For a structural record, which may have 20 channels, for example, the data will be released as 20 separate files, each containing one channel. Several parameters specific for structures are included in the parameter specifications; these would be blank or zero for non-structural stations.

The format of Volume 1 and Volume 3 files can be specified in a similar way, and the headers would be largely coincident with the Volume 2 header. The Volume 2 format is used as an example; the principles agreed to for that format would carry forward to the other files.

More fundamental files than the user-oriented data products discussed here are also needed. These files contain the data in the most raw and elemental way; the data in these files would still be in digital counts as recorded, for example. In the context of the products discussed here, they could be referred to as Volume 0 files, since they are precursors to the Volume 1 files. These files would be maintained by the data-generating agencies. Their format need not be common, since they are specific to the instruments and network configurations of a specific network. Nonetheless, they can be archived as the most basic record, much like accelerogram film itself is preserved for analog recorders, even though the film has been digitized and the data has been processed and released. These fundamental files will not interfere with the common data format for earthquake engineering users, which is the focus of this effort.

Summary

A common format for strong motion data is needed to allow effective utilization of the data being recorded by the various networks in the United States and worldwide. This does not require data-generating agencies to stop using the formats they are currently using, but only requires that a common standard be agreed to, and that converters be available to convert data from their formats to the common format. From a user perspective, this step will represent a breakthrough in the convenient and rapid of new data. Definition of this format, with adequate flexibility and power to handle past data and needs of the present and immediate future, is an important task to accomplish. Some suggestions toward that goal are included here from an initial Workshop. The next steps include broad input from a larger user workshop, and finalization of a proposed format by a COSMOS and its member agencies.

Acknowledgements

The ideas and suggestions from the Working Group were important to the progress to date. Participants included N.Abrahamson and W. Savage (PG&E), R. Graves (Woodward Clyde), A. Tumarkin (UCSB), W. Joyner (USGS), M. Huang and V. Graizer (CDMG) and the authors.

References

Brady, A., V. Perez and P. Mork (1980). The Imperial Valley earthquake, October 15, 1979, Digitization and processing of accelerograph records, U.S. Geological Survey, Seismic Engineering Data Report, Open File Report 80-703.

Converse, A.M. and A.G. Brady (1992) BAP: Basic strong-motion accelerogram processing software, USGS Open File Report 92-296A.

Hudson, D., A. Brady and M. Trifunac (1969). Strong-motion earthquake accelerograms, digitized and plotted data, Vol. I, Parts A-Y, Calif. Inst. Technology, EERL reports.

Hudson, D., A. Brady, M. Trifunac and A. Vijayaraghavan (1971). Strong-motion earthquake accelerograms, Corrected accelerograms and integrated velocity and displacement curves, Vol. II, Parts A-Y, Calif. Inst. Technology, EERL Reports.

Shakal, A. and M.J. Huang (1985) Standard tape format for CSMIP strong-motion data tapes, CDMG/SMIP Report OSMS 85-03, 26 pp.

Appendix A - Example Header

(For purposes of illustration, a header from a data file produced during the Blue Book project is shown. The record happens to be the Pacoima Dam record from the San Fernando earthquake of 1971.)

```
CORRECTED ACCELEROGRAM IIC041  71.001.0    COMP S16E   FILE  1 CORRESPONDING TO
FILE   1  OF UNCORRECTED ACCELEROGRAM DATA OF VOL. I-C, EERL 71-20
SAN FERNANDO EARTHQUAKE
FEBRUARY 9, 1971 - 0600 PST
IC041   71.001.0 R                                                    18
STATION NO. 279   34 20 06N,118 23 48W                                38
PACOIMA DAM, CAL.                                                     17
COMP S16E                                                             9
SAN FERNANDO EARTHQUAKE   FEB 9, 1971 - 0600 PST                      48
EPICENTER  34 24 00N,118 23 42W                                       31
INSTR PERIOD   0.0510 SEC  DAMPING   0.544                            42
NO. OF POINTS   3002  DURATION   41.822 SEC                           43
UNITS ARE SEC AND G/10                                                22
RMS ACCLN OF COMPLETE RECORD   1.1934 G/10                            42
ACCELEROGRAM IS BAND-PASS FILTERED BETWEEN   0.070  AND  25.000  CYC/SEC
  2091  INSTRUMENT AND BASELINE CORRECTED DATA
AT EQUALLY-SPACED INTERVALS OF   0.02   SEC.
PEAK ACCELERATION    -1148.06055  CMS/SEC/SEC AT    7.7400 SEC
PEAK VELOCITY         -113.23398  CMS/SEC     AT    3.0400 SEC
PEAK DISPLACEMENT       37.66193  CMS         AT    7.7800 SEC
INITIAL VELOCITY         1.22996  CMS/SEC       INITIAL DISP.       0.42512 CMS
SAN FERNANDO EARTHQUAKE   FEB 9, 1971 - 0600 PST

IIC041  71.001.0    PACOIMA DAM, CAL.    COMP S16E

        1         1         3        41        71         1         0         5       279        34
       20         6       118        23        48        34        24         0       118        23
       42         2         9      1971       600         0       164      3002        23        17
        0         0         0         0         0         0         0         0         0         0
        0         0         0         0         0         0         0         0         0         0
     3002      3004      2091         2        10        10         1         0        48        48
       10        10         2      1046         5       419         0         0         0         0
        0         0         0         0         0         0         0         0         0         0
        0         0         0         0         0         0         0         0         0         0
        0         0         0         0         0         0         0         0         0         0
  0.05100   0.54400  41.82199   1.19340   0.10000   0.0       0.0       0.0       0.0       0.0
  0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0
  0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0
  0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0
  0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0
  1.00000   0.01020   0.01000  41.82199 123.19958   1.00000   1.00000  27.00000   2.00000  41.79999
  0.02000   0.07000   0.02000   0.0       7.740001148.06055   3.04000-113.23398   7.78000  37.66193
  1.22996   0.07000  25.00000   0.20000   0.20000   0.42512   0.0       0.0       0.0       0.0
  0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0
  0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0       0.0
      -78      -260      -122       259       -72      -208       306      -258      -341       371
     -295      -341       362      -249      -183       369       311      -144       -71       146
      -36        40       527        28      -715       183      1029       719       119      -254
     -594      -433        16       432       330        23        24      -149       -60       225
      409       -40      -814      -224       233       104       931       797      -425      -943
```

Appendix B - Integer Parameters (Interim)

	Parameter Description
1	Accelerogram channel number
2	Processing stage index of data
3	Station channel number
4	Number of channels in record (i.e., recorder)
5	Total number of channels at station (i.e., in recording system)
6	Original data sampling rate (samples per second if digital; -1 if analog)
7	Event identification code: 0 for mainshock, 1,2,3,.. for aftershocks, etc
8	Recorded medium code (index for film, tape, solid-state, etc.)
9	Code for network/agency operating network that recorded this record
10	Code for network/agency that owns the instrumentation at this station
11	Code for network/agency that processed this record
12	Station number assigned by Network/Agency
13	Unique station number assigned by COSMOS
14	Station type code (index for FF, Bldg, Bridge, Special array, other)
15	COSMOS data format version number for this file (3 digits)
16	Number of elements in this Integer header (e.g., 100)
17	Number elements in Real Header (e.g., 50 for Vol.1; 100 for Vol.2)
18	Accelerograph recorder type code
19	Sensor Type code
20	Units of raw Vol.1 data (index for g, cm/sec, etc)
21	Full-scale output of sensor (nominal; actual sens. n Real parameter 6)
22	Gain of this channel (1,2,4,8, etc)
23	Sample word length as originally recorded (number of bits)
24	Effective number of bits, if different
25	Orientation code for sensor for this channel
26	Orientation of sensor with respect to Reference direction
27	Number of raw acceleration points
28	Number of letters in earthquake name
29	Number of letters in station name
30	Number of letters in earthquake title line
31	Azimuth of Reference North of a structure, if applicable
32	Azimuth of accelerograph connector, for a FF accelerograph
33	Number of Vol.1 acceleration points in record (same as param. 27 in cases of equally-spaced Vol.1 data)
	Record start time:
34	Hour (GMT)
35	Minute
36	Second

37 Fraction of second (integer, msec)
38 Julian day of year
39 Month
40 Day of month

41 Year (inferred if necessary, 4-digit)
42 Time quality indicator (0-10)
43 Time source indicator
44 Event number of this record (in recorder)
45 Recorder serial number
46 Number of recorders in recording system at station
47 Recorder System serial number (if appropriate)

Additional parameters only appear for Vol.2 and Vol.3 files:

48 Number of Vol.2 acceleration data points
49 Desampling factor from Vol.1 to Vol.2
50 Decimation factor, if any, in long-period filtering of acceleration

51 Decimation factor, if any, for long-period filtering of velocity
52 Decimation factor, if any, for long-period filtering of displacement
53 Number of velocity points
54 Number of displacement points
55 No. of periods for which response spectra are computed
56 No. of damping values for which response spectra computed (typically 5)
57 Low-cut filter type, Vol.2 (1=Ormsby, 2=Cos bell, 3=Butterworth, ..)
58 Order of low-cut filter (if Butterworth, etc)
59 High-cut filter type, Vol.2
60 Order of high-cut filter
61 Frequency domain/time domain filtering flag (1 if filter applied in time domain, 2 if applied in frequency domain)
62-100 --

Appendix C - Real-Valued Parameters (Interim)

Description

1 Natural period of transducer (seconds)
2 Damping of transducer (fraction of critical)
3 Length of record (in seconds, Vol.1)
4 RMS value of record (in g)
5 Units of Vol.1 acceleration (fractions of g)
6 Sensitivity of transducer (cm/g for film recorder; for acceleration sensors, mvolts/g; for other sensors, volts per motion unit)
7 Peak acceleration (Vol. 1) for this channel (in g)
8 Time of peak acceleration value (seconds after start)
9 Natural frequency of transducer (in Hz)

Record Resolution Parameters:
A) For digitized film records (i.e., integer parameter 8 = -1):
10 Digitizer y-step (acceleration) size, in microns (cm/10000)
11 Digitizer y-step, in milli-g
12 Digitizer x-step (time) size, in microns (cm/10000)
13 Actual average time step of digitized record, in milliseconds
14 Standard deviation of time step, in milliseconds
15 Minimum time step size, as digitized (milliseconds)
16 Maximum time step size, as digitized (milliseconds)

B) For digital records (i.e., integer parameter 8 > 1):
10 LSB, in millivolts or microvolts (i.e., mv or uv/count; see param. 12)
11 LSB, in milli g or micro g (i.e., mg or ug/count; see param. 12)
12 LSB units code (1=mv/mg; 2=microvolts/micro g)
13 Sample interval, delta t (in msec)
14-16 --

17 Full scale output (volts) of sensor
18 Pre-event memory (PEM) length for recorder (seconds)
19 Post-event time for recorder (seconds)
20 Latitude of station (North positive)
21 Longitude of station (West negative)
22 Elevation of station (meters)
Earthquake location parameters
23 Latitude of earthquake epicenter (North positive)
24 Longitude of earthquake epicenter (West negative)
25 Depth of hypocenter (km)
Earthquake magnitudes
26 ML

27 Ms
28 Mw
29 mb
30 Other magnitude

Parameters values for Vol.2 and Vol.3 files:

31 Scaling factor for converting acceleration from Vol.1 units to cm/sec/sec (e.g., 98.0665)
32 Time step of Vol. 2 acceleration data (sec)
33 Length of Vol.2 acceleration time series (secs)
34 Termination frequency of high-cut filter (Hz), if Ormsby
35 Roll-off width of high-cut filter (Hz), if Ormsby
36 Length of Vol. 2 output (secs), after resampling to dt given in param. 32
37 Roll-off corner frequency (3 dB point) of low-cut filter (Hz); for non-Ormsby filter
38 Roll-off width of low-cut filter (Hz), if Ormsby
39 Time of peak acceleration (Vol. 2), seconds after start
40 Peak acceleration value (Vol. 2), cm/sec/sec
41 Time of peak velocity (secs after start)
42 Peak velocity (cm/sec)
43 Time of peak displacement (secs after start)
44 Peak displacement (cm)
45 Initial velocity value (cm/sec)
46 Roll-off corner frequency (3 dB point) for high-cut filter (Hz); for non-Ormsby filter
47 Velocity time step (sec)
48 Displacement time step (sec)
49 Initial displacement (cm)

Duration measures:

50 Bracketed, seconds over 5%g
51 5-95% Duration
52 5-75% Duration
53 Other duration
54 Arias Intensity
55 CAV (m/s)
56-100 --