EDUQUAKES

A Quick SEED Tutorial

by Adam T. Ringler and John R. Evans

INTRODUCTION

A number of different government-funded seismic data centers offer free open-access data (e.g., U.S. Geological Survey, National Earthquake Information Center, the Incorporated Research Institutions for Seismology (IRIS), and Data Management System), which can be freely downloaded and shared among different members of the community (Lay, 2009). To efficiently share data, it is important that different data providers follow a common format. The Standard for the Exchange of Earthquake Data (SEED) provides one such format for storing seismic and other geophysical data. The SEED format is widely used in earthquake seismology; however, SEED and its structure can be difficult for many first-time users (ourselves included). Below is a quick tutorial that outlines the basic structure of SEED format. This write-up is in no way intended to replace the comprehensive SEED manual (Ahern *et al.*, 2009), and instead of going into the details of any specific part of the SEED format we refer the reader to the manual for additional details. The goal of this write-up is to succinctly explain the basic structure of SEED format as well as the associated jargon, as most commonly used now, in a colloquial way so that novice users of SEED can become more familiar with the format and its application quickly. Our goal is to give the reader the necessary background so that when problems or questions about SEED format arise they will have some understanding of where they should look for more details or from where the problem might be stemming. As a secondary goal, we hope to help the reader become familiar with the SEED manual (Ahern *et al.*, 2009), which contains detailed information about all aspects of the SEED format.

BASIC STRUCTURE

SEED format logically breaks down into the part containing time-series data, called miniSEED, and the other stuff, which is called dataless SEED (Fig. 1). The dataless SEED file contains various network and station metadata, including the always-daunting instrument response information (described subsequently), as well as other information about the stations such as their physical location. By combining a dataless file with a miniSEED file, you end up with a full SEED volume (in the days of magnetic tape, they were stored as a contiguous whole). However, most people avoid the redundancy of putting these together into one file volume, and simply keep a separate dataless file (usually for a network or station and unchanging for extended periods) along with a number of miniSEED files containing time-series data by the hour, for a particular earthquake, or other logical separation. This way when corrections are made to the station metadata (e.g., the response information was wrong), the user simply needs to get the new dataless file, which is quite small.

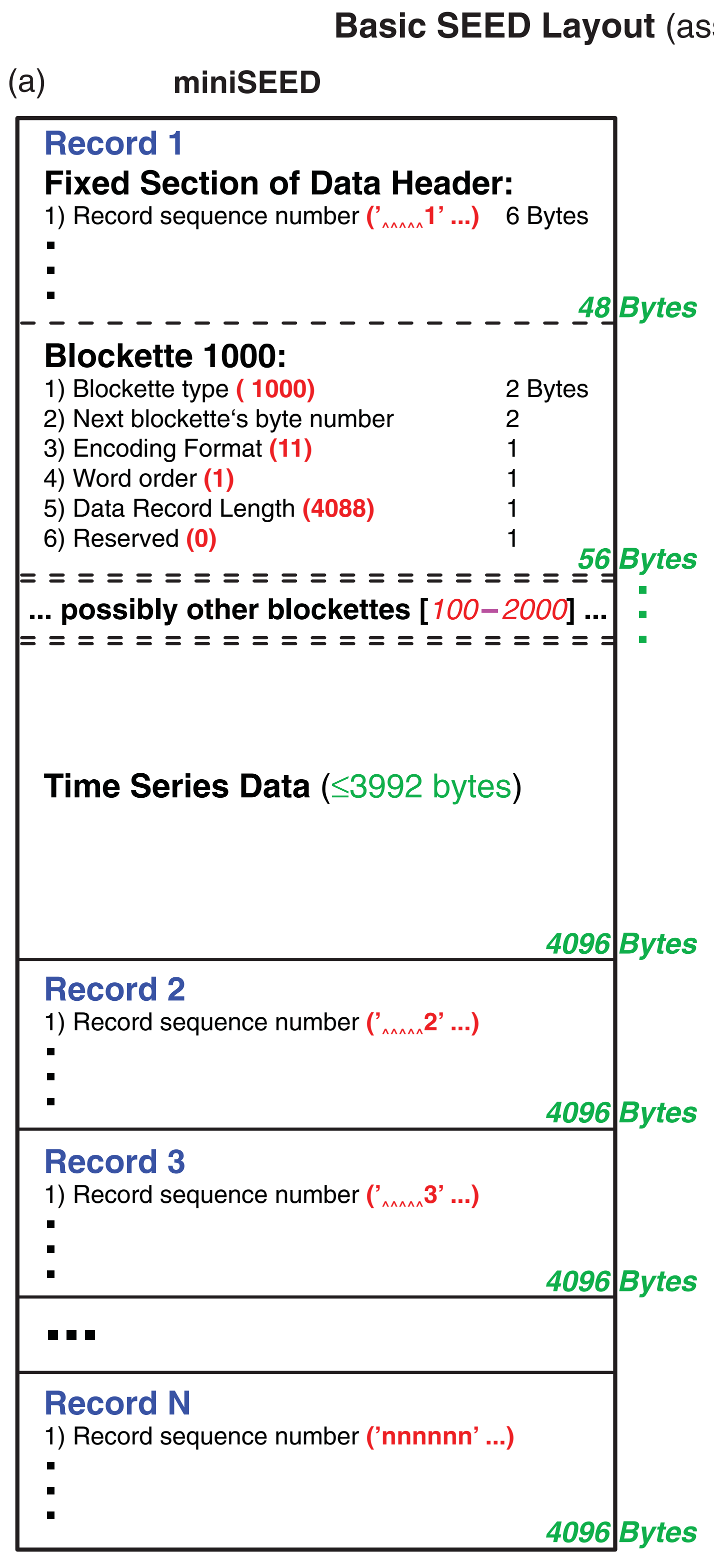
miniSEED

SEED files that do not contain detailed station metadata are called miniSEED and can be thought of as time series only data. Each miniSEED file breaks down into a collection of data records. These data records are a fixed number of bytes long, which is a power of two (e.g., 512 bytes or 4096). At the beginning of every record are 48 bytes of initial information about the record. This is called the fixed section of the data header (Ahern *et al.*, 2009, pp. 108). The fixed section of the data header provides information on decoding the data as well as some basic information about the data (e.g., network name, station name, etc.). By organizing the data this way, it is possible to have multiplex data-containing records from multiple different data sources (e.g., multiple stations). Each of the fields in this fixed section of the data header has a fixed length given as a number of bytes. The values in the fixed section of the data header can then be decoded according to their type as described in Figure 2. For example, the fourth field is the station identifier code, is in ASCII format, and is five bytes long (Fig. 2).

Following the fixed section of the data header, additional information is collected into blockettes. A blockette is a grouping of information. These optional data blockettes (Fig. 2) contain additional information about the data record that is not easily described elsewhere. For example, if there was an event detected in the record, then the record may contain an event detection blockette such as a Murdock event detection (Murdock and Hutt, 1983) Blockette (blockette type 201), which gives information about an event (Ahern *et al.*, 2009, pp. 113).

Additional important blockettes in miniSEED include blockette 1000 (Fig. 1, pp. 123), which contains information about the time-series data length (field 5) as well as information on how those data are encoded (field 3).

Finally, the remainder of the record contains encoded time-series data. The beginning of the data can be found in field 17 of the fixed section of data header. Data can be



▲ **Figure 1.** Example miniSEED and dataless Standard for the Exchange of Earthquake Data (SEED) layout. (a) Each miniSEED record contains a fixed section of data header, followed by blockettes, and finally encoded data. (b) Each dataless SEED record contains volume index control headers, followed by station metadata, which is further broken down into channel metadata. It is possible to combine miniSEED with dataless SEED to produce full SEED.

[ID]	Blockette Name	Size Range (Bytes)	pp. in manual
Volume Control Index Headers Headers			
[005]	Field-volume identifier	14 to 35	36
[008]	Telemetry volume identifier	29 to 113	37
[010]	Volume identifier	18 to 239	38
[011]	Volume stations header index	21	39
[012]	Volume time-span index	19 to 61	40
Abbreviation Dictionary Control Headers			
[030]	Data-format dictionary	17 to 66	42
[031]	Comment description	16 to 85	43
[032]	Cited-source dictionary	12 to 199	44
[033]	Generic abbreviation[s]	11 to 60	45
[034]	Units abbreviations	11 to 80	46
[035]	Beam configuration	33	48
[041]	FIR dictionary	25 to 49	49
[042]	Response, polynomial	108 to 132	51
[043]	Response, poles and zeros	145	53
[044]	Response, coefficients	75 to 99	55
[045]	Response, list	82 to 106	57
[046]	Response, generic	46 to [70 + 24n]	58
[047]	Decimation	54 to 78	59
[048]	Channel sensitivity/gain	63 to 108	60
[049]	Response, polynomial	108 to [132 + 24n]	61
Station Control Headers			
[050]	Station identifier	61 to 163	64
[051]	Station comment	19 to 61	66
[052]	Channel identifier	100 to 198	67
[053]	Response, poles and zeros	[142 + 48m + 48n]	71
[054]	Response, coefficients	[72 + 24m + 24n]	73
[055]	Response, list	[79 + 60n]	75
[056]	Response, generic	[43 + 24n]	76
[057]	Decimation	51	77
[058]	Channel sensitivity/gain	60 to [81 + (25 to 46)n]	78
[059]	Channel comment	18 to 124	79
[060]	Response reference	[17 + 6n]	80
[061]	FIR response	35 to [59 + 14n]	82
[062]	Response, polynomial	[105 + 24n]	84

▲ **Figure 2.** Figure of blockettes including the blockette number (column 1), blockette name (column 2), the number of bytes (column 3), and the location in the SEED manual (column 4; Ahern *et al.* 2009). (*Continued*)

Time Span Control Headers			
[070]	Time span identifier	10 to 52	101
[071]	Hypocenter information	56 to [125 + (16 to 64)n]	102
[072]	Event phases	53 to 93	103
[073]	Time span data start index	30 to [(19 to 40)n]	104
[074]	Time series index	49 to [112 + (11 to 32)n]	105
Fixed section of (waveform) data header		**48 total**	108
	1) Record sequence number **(^^^^^1** etc.**)**	6	
	2) Data header/quality indicator ['D' | '**R**' | 'Q' | 'M']	1	
	3) Reserved byte **(0)**	1	
	4) Station identifier code **('SSSSS')**	5	
	5) Location identifier **('LL')**	2	
	6) Channel identifier **('CCC')**	3	
	7) Network Code **('NN')**	2	
	8) Record start time **(BTIME)**	10	
	9) Number of samples **(UWORD)**	2	
	10) Sample rate factor **(200)**	2	
	11) Sample rate multiplier **(1)**	2	
	12) Activity flags **(UBYTE)**	1	
	13) I/O and clock flags **(UBYTE)**	1	
	14) Data quality flags **(UBYTE)**	1	
	15) Number of blockettes that follow **(UBYTE)**	1	
	16) Time correction **(0)**	4	
	17) Beginning of data **(48)**	2	
	18) First block **(1)**	2	
MiniSEED-relevant blockettes (and old-style waveforms in SEED file)			
[100]	Sample rate	12	111
[200]	Generic event detection	52	112
[201]	Murdock event detection	60	113
[202]	Log-Z event detection	reserved	114
[300]	Step calibration	60	115
[310]	Sine calibration	60	116
[320]	Pseudo-random calibration	64	117
[390]	Generic calibration	28	118
[395]	Calibration abort	16	119
[400]	Beam	16	120
[405]	Beam delay	6	121
[500]	Timing	200	122
[1000]	Data-only SEED (miniSEED time series)	8	123
[1001]	Data extension	8	124
[2000]	Variable-length Opaque data	[15 + n]	125

Figure 2. Continued.

encoded in a number of different formats including 16-, 24-, or 32-bit integers (encoding types 1, 2, or 3 in field 3). Additional common encoding types are STEIM 1 and STEIM 2 compression (encoding types 10 and 11). For more details on time-series encoding, we refer the reader to Ahern *et al.* (2009, pp. 123 and 171). Figure 1 illustrates the organization of a typical fixed header and dataless SEED and miniSEED files.

INTERMISSION—STATION, NETWORK, CHANNEL, AND LOCATION (SNCL): THE SEED STATION NAMING CONVENTION

To avoid confusion, we now briefly review the basics of station naming.

Station

A station name refers to a physical location where there are various instruments. These names are abbreviated with five or fewer letters, which comprise the station call sign. For example, ANMO is the permanent station at the Albuquerque Observatory in New Mexico. Historically, stations that ended with the letter O in the Global Seismographic Network (GSN) were stations adopted from the Seismic Research Observatory and were sensors deployed at depth. This convention is no longer kept. We should also note that the GSN is not a network in the SEED sense (it violates the two-or-fewer letter rule) and is made up of a number of other networks such as the II and IU networks. Whenever a station is relocated more than 1 km from its initial location, it will get a new station name, which gets registered at the International Seismological Centre (http://www.isc.ac.uk/registries/, last accessed August 2015). For example, station South Pole, Antarctica (SPA) was moved to a nearby quieter site and renamed to Quiet South Pole, Antarctica (QSPA).

Network

A collection of stations generally operated by one group. Networks are abbreviated with a one- or two-letter network code assigned by the International Federation of Digital Seismic Networks (FDSN). For example, the IRIS/USGS portion of the GSN is IU (operated by the Albuquerque Seismological Laboratory), whereas the II network is operated by IRIS and the International Deployment of Accelerometers, at the University of California at San Diego. A full list of recognized network codes is available from FDSN (http://www.fdsn.org/networks/, last accessed August 2015).

Channel

The channel for a given time series is a single data stream. The three characters of the channel name include the frequency band, the sampling rate (e.g., L for long-period 1 samples/s), the instrument code (e.g., H for high gain), and the orientation code (e.g., N for a north–south axis). In this example, the channel name would be LHN. Another example could be BNZ, for a broadband (band code B; 10–80 samples/s) accelerometer (instrument code N) with vertical axis (orientation code Z). A collection of common channel-naming conventions is located in Table 1, whereas a more complete list is contained in appendix A of the IRIS SEED manual (Ahern *et al.*, 2009). It is important to note that the orientation codes are often not precise. In most cases, there will be slight orientation errors that are described in the station metadata. Therefore, it is always important to rotate your data, especially the horizontal axes, according to the metadata and not to rely on the orientation code alone.

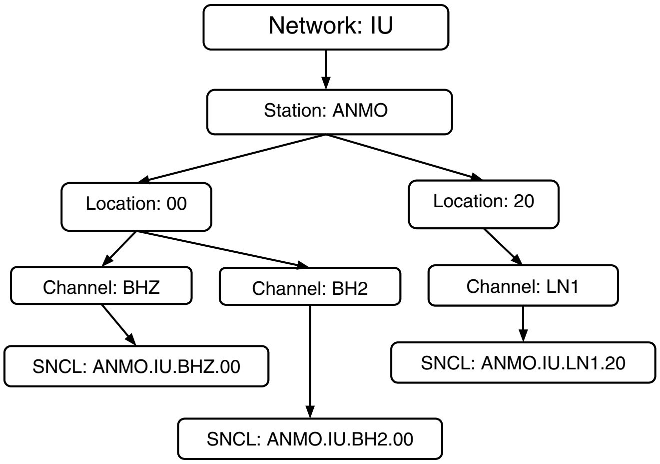
Table 1
Common Channel-Naming Conventions

Sample rate prefix
E: greater than 80 samples/s (including 100 and 200 samples/s)
S: 10–80 samples/s
H: greater than 80 samples/s (broadband)
B: from 10 to 80 samples/s
L: 1 samples/s
V: 0.01 samples/s
Instrument code
H: high-gain seismometer
L: low-gain seismometer
G: gravimeter
M: mass position
N: accelerometer
P: geophone
D: pressure
Orientation
Z, N, E: traditional vertical, north–south, east–west
T, R: transverse and radial
1, 2, 3: orthogonal with nontraditional orientation

Sample rate prefixes (rows 2 through 7; initial letter in channel code), common instrument codes (rows 9 through 15; middle letter in channel code), and common orientations (rows 17 through 19; last letter in channel code).

Location

The final field to complete a full SNCL name (pronounced "snickle" for station site, network, channel, and location) is the location. The location code is two letters or numbers that help to distinguish similar instruments or similarly named channels at the same station. This code becomes important for stations like QSPA where there are more than five broadband seismometers close to one another. So if you wanted long-period vertical data from the primary borehole sensor (location code 00), your SNCL name would be QSPA.IU.LHZ.00. Other examples of SNCL combinations are given in Figure 3. It can also be important when there are similar data streams at different sample rates from the same instrument. For example, triggered low-gain vertical data (HNZ) as well as continuous low-gain vertical data from the same sensor can be distinguished with different location codes. At some stations, there will be only one instrument, and in these situations sometimes network



▲ **Figure 3.** Flow diagram of a station, network, channel, and location grouping. The station ANMO has multiple location codes (e.g., 00 and 20), and a single location code can have multiple associated channels (e.g., location code 00 has both BHZ and BH2).

operators will use a blank location code or they might use the blank location code for the primary instrument. In these cases, some data-requesting mechanisms will accept "—" as the blank location code.

It can be handy to use the IRIS metadata aggregator (MDA) to quickly verify information for a specific SNCL. The location of the MDA is (http://www.iris.edu/mda, last accessed August 2015). For more details on the band codes and instrument codes, we refer the reader to pages 133 through 140 of Ahern *et al.* (2009).

DATALESS

In addition to the miniSEED waveform data, you will also need to understand the basics of the dataless SEED, which contains additional information about networks and stations. The dataless SEED is organized solely using blockettes. The first of these blockettes is the volume index control headers. These blockettes provide initial information about the volume as well as the time spans (Fig. 2). For example, blockette 11 describes the location in the dataless volume of the start of the metadata for a given station (blockette 11, field 5). There will also be abbreviation dictionary control headers, which describe various abbreviations in later blockettes.

The station control headers follow the volume index control headers. For each station in a dataless volume, there is usually at least one station identifier blockette (blockette 50). This blockette contains information about the physical location of the station (fields 4 through 6) as well as other information about the station. For each of the channels at the station, there is a corresponding blockette 52 (Ahern *et al.*, 2009, pp. 67) called a channel identifier blockette. This blockette gives information about a physical sensor. It is important to check the location code field (field 3) in this blockette because many stations will contain multiple channel-identifier blockettes.

A Pitfall

The physical location of a given station can be different for different location codes, though generally by much less than 1 km. For example, when a sensor is buried at depth, it will have a different elevation (blockette 52, field 12) than a sensor with a different location code that is installed at the surface (blockette 50, field 6). For a surface instrument blockette 52, field 13 will be set to zero. Another thing to be careful about is the elevation (blockette 52, field 12) and the local depth (blockette 52, field 13). The local depth corresponds to the overburden of the instrument relative to the elevation. The elevation corresponds to the geodetic elevation of the instrument and is described in the channel blockette (blockette 52).

CORRECTING THE RESPONSE

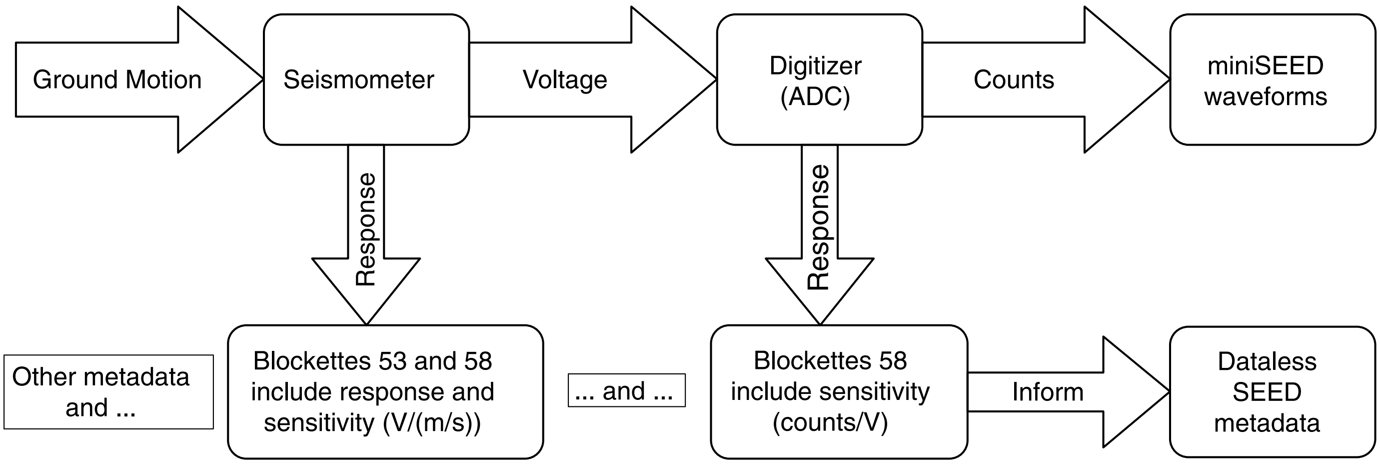
The most seismic data are collected from analog sensors that output a voltage. This voltage is then fed into a digitizer that converts the voltage to counts, then into miniSEED files or a proprietary binary format. We refer the reader to Bormann (2012) or Havskov and Alguacil (2006) for detailed treatments of seismic instrumentation. To convert seismic data to ground-motion units such as velocity or acceleration, the user must correct the data for the response of the digitizer and the response of the sensor, which are described as analog and digital filters. We do not delve into the details of describing the instrument responses or their representations. Instead, we refer the reader to Steim (2015) for an overview on describing the seismic response and limit ourselves to how the response is described in the dataless SEED. Each necessary instrument correction is described by a different filter stage in the dataless SEED file, with the last stage (called 0 or 99) being the total sensitivity for all of the nonzero stages combined (Fig. 4). This final stage can be used as a check to verify the total sensitivity of all the other stages. In addition to simple response description, there will often be additional blockettes that describe the decimation of the data (blockette 57) as well as finite impulse response filter descriptions or infinite impulse response filter descriptions (blockette 54).

Digitizer Sensitivity

The first step in correcting your data to ground motion is to make a correction that accounts for the digitizer sensitivity, which is a conversion from counts to volts (Fig. 4). This sensitivity is usually described using a channel sensitivity in blockette 58 and is analogous to what was used to describe the seismometer's sensitivity. Although analog-to-digital converters often have other response elements, such as decimation filters, these are typically expressed only as a corner frequency, often left implicit.

Instrument Response

Once our data have been corrected to account for the sensitivity of the digitizer, our data will be in volts. The conversion from an output voltage of a seismometer to ground-motion units requires knowing the seismometer's sensitivity at a given



▲ **Figure 4.** Schematic of the stages from ground motion to miniSEED (top row). The seismometer records ground motion and outputs a voltage as described in blockettes 53 and 58 (bottom row). The voltage is then digitized into counts as described in an additional blockette 58 and finally converted into miniSEED.

frequency as well as the shape of its response function (Fig. 4). Because most seismic instruments have an essentially flat response to a particular ground-motion unit over some frequency band, that sensitivity is often stated there. For example, the Güralp CMG-3T is a flat-to-velocity sensor from 50 Hz to 120 s period. Hence it is advantageous to state the midband sensitivity at a frequency, in which the sensor is flat to velocity, such as a 10–20 s period, which is well away from the Nyquist frequency as well as the long-period corner (120 s period). We further restrict our example to the vertical component of the secondary sensor (location code 10) at IU station TUC (Tucson, Arizona) was a Güralp CMG-3T from 28 May 2009 to 21 December 2010, and it had a sensitivity that was $\sim 20{,}000\ \mathrm{V/(m/s)}$, which is at 50 s period (Fig. 5). Upon dividing our voltage time-series data by the instrument sensitivity, we obtain data that are in ground-motion units, in this case meters per second, but is bandlimited (because we have not yet accounted for the response of the instrument, only the midband sensitivity). Figure 5 shows the response as a function of frequency as well as the blockettes used to obtain the response curve. The sensitivity value is put in blockette 58, field 4 (sensitivity) and field 5 (corresponding frequency).

Along with the midband sensitivity, it is important to correct for any instrument roll-off in sensitivity at high or low frequencies (for which velocity sensors and some others are not flat). This response element is generally described in blockette 53 with a pole and zero model. The type of pole and zero model used is described in field 3 of this blockette, and is called the transfer function type. For example, Laplace pole–zeros are type A and described in radians per second, whereas type B is Fourier poles and zeros in hertz (types A and B differ from one another by a factor of 2π). Detailed descriptions of these two pole–zero models as well as examples are described in appendix C of Ahern *et al.* (2009). Further background on signal processing can be found in Scherbaum (2007). The pole and zero model gets normalized by A_0 (field 7) so that the pole and zero model has unity gain at the normalization frequency given in field 8 (Fig. 5). This avoids confusion when combining the sensitivity of the sensor with the response model because the response model in conjunction with A_0 will always be unity at the normalization frequency.

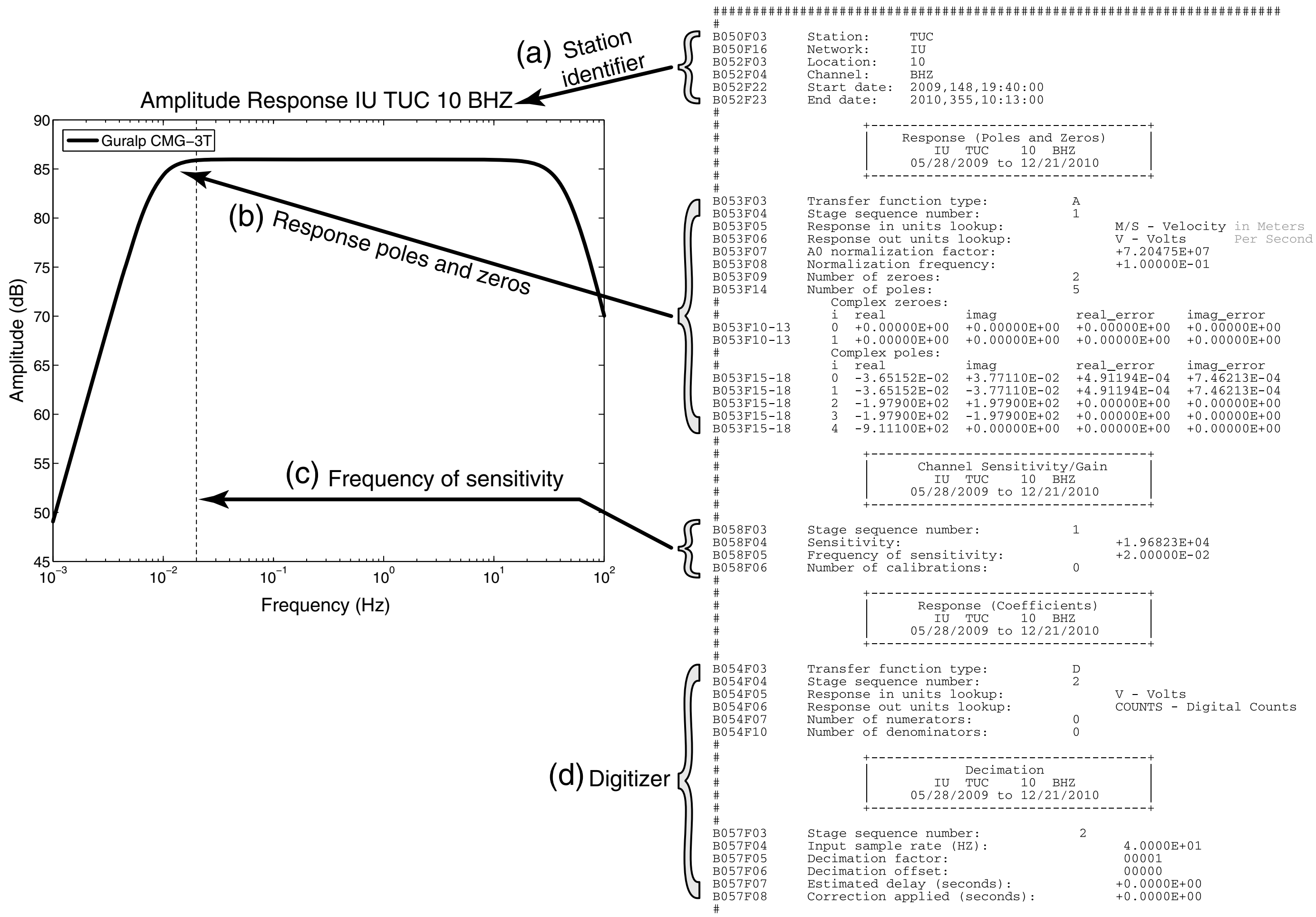
When describing the instrument sensitivity and the response normalization A_0, it is valuable to do it at a frequency at which both the instrument and the digitizer are flat. This avoids the complication of having to check normalization values at different frequencies in the cascade.

ADDITIONAL COMMONLY USED BLOCKETTES

In addition to the blockettes, we have described that there are other possible approaches to describing metadata. For example, instead of using pole and zero response models it is possible to list the instrument response using a frequency, amplitude, and phase listing (blockette 55). It is also possible to describe polynomial-shaped responses using blockette 62. Finally, there are a number of additional blockettes that can be used to describe various calibration events (blockettes 300 through 395). A detailed list of the possible blockettes is contained in Figure 2, and detailed descriptions of these blockettes are contained in Ahern *et al.* (2009).

SUMMARY

Seismic data are widely stored and exchanged in SEED format. In current usage, these data are commonly divided into a file of slowly changing metadata, called a dataless SEED file, whereas time series are in miniSEED files each containing limited time intervals of waveforms. Although the SEED format *in toto* is complex, can be difficult to understand, and contains anachronisms from its birth in the time of reel-to-reel digital tape drives, it remains the dominant format in which earthquake seismic data are exchanged. By having a flexible data



▲ **Figure 5.** The response of Incorporated Research Institutions for Seismology (IRIS), and Data Management System (IRIS)/U.S. Geological Survey (USGS; network code IU) station TUC (Tucson, Arizona) for the secondary sensor's (location code 10) broadband high-gain vertical channel (channel code BHZ). The upper-left panel shows the response of the instrument as derived from the metadata (right panel). Mark (a) represents the channel identifier along with the time span is the content of blockette 50, (b) represents the shape of the response is the content of blockette 53, (c) represents the frequency of the sensitivity described in blockette 58, and (d) the sensitivity of the digitizer. The right side of this figure is the metadata description of the response with the left-most alphanumeric strings corresponding to the blockette and the field. From top to bottom, the blockettes that are contained in this example are 50, 52, 53, 58, 54, and 57.

format to exchange seismic data, it has been possible for many different seismic data providers to share data (Lay, 2009) with minimal complications caused by conversion between different formats. Many recording instruments also have adopted SEED or at least miniSEED as sole or optional waveform format for recording and transmission. Dataless SEED is currently the most common way to share metadata, but there is evidence that purpose-designed Extensible Markup Language (XML) formats, using the FDSN Station XML schema (http://www.fdsn.org/xml/station/, last accessed August 2015) may soon dominate metadata exchange. Even so, miniSEED seems to be maintaining its place as the dominant medium for exchange of waveform data and even if it has some deficiencies, it is a tribute to its developers that it has been one of the main formats in earthquake seismology since its inception in 1987.

DATA AND RESOURCES

All data and metadata used are freely available from the Incorporated Research Institutions for Seismology (IRIS) Data Management System (DMS). Additional software for reading miniSEED and dataless SEED can also be found from the IRIS DMS.

ACKNOWLEDGMENTS

The authors wish to thank Robert Anthony, Alejandro Gonzales, Mike Hagerty, and Battalgazi Yildirim for various conversations over the years that lead us to formulating a basic summary of the Standard for the Exchange of Earthquake Data (SEED) format. We also wish to thank Bob Hutt and Tyler Storm for verifying various SEED conventions and suggesting the value of writing a quick guide to data in SEED format. We thank Tim Ahern for help with links related to the Federation of Digital Seismic Networks (FDSN). Finally, we thank Robert Anthony, Jill McCarthy, Derek Schutt, Janet Slate, Valerie Thomas, John West, and an anonymous reviewer for careful reviews that helped to improve the article. We also thank Editor Alan Kafka for additional suggestions and comments on the article that helped broaden the scope.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

The facilities of Incorporated Research Institutions for Seismology (IRIS) Data Services, and specifically the IRIS Data Management Center, were used for access to waveforms, related metadata, and/or derived products used in this study. IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience and EarthScope (SAGE) Proposal of the National Science Foundation under Cooperative Agreement EAR-1261681.

Global Seismographic Network (GSN) is a cooperative scientific facility operated jointly by the IRIS, the U.S. Geological Survey (USGS), and the National Science Foundation (NSF), under Cooperative Agreement EAR-1261681.

REFERENCES

Ahern, T., R. Casey, D. Barnes, R. Benson, T. Knight, and C. Trabant (2009). *SEED Reference Manual, version 2.4*, http://www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf (last accessed August 2015).

Bormann, P. (Editor) (2012). *New Manual of Seismological Observatory Practice (NMSOP-2), IASPEI*, GFZ German Research Centre for Geosciences, Potsdam, Germany, http://nmsop.gfz-potsdam.de (last accessed August 2015), doi: 10.2312/GFZ.NMSOP-2.

Havskov, J., and G. Alguacil (2006). *Instrumentation in Earthquake Seismology*, Springer, Dordrecht, the Netherlands, 358 pp.

Lay, T. (Editor) (2009). Seismological grand challenges in understanding earth's dynamic systems, *Report to the National Science Foundation*, IRIS Consortium, 76 pp., http://www.iris.edu/hq/lrsps/seis_plan_final.pdf (last accessed August 2015).

Murdock, J. N., and C. R. Hutt (1983). A new event detector designed for the Seismic Research Observatories, *U.S. Geol. Surv. Open-File Rept. 83-785*, 39 pp.

Scherbaum, F. (2007). *Of Poles and Zeros: Fundamentals of Digital Seismology*, Second Ed., Springer, Dordrecht, the Netherlands, 268 pp.

Steim, J. M. (2015). Theory and observations—instrumentation for global and regional seismology, in *Treatise on Geophysics*, Vol. 1, Elsevier, Amsterdam, The Netherlands.

Adam T. Ringler
U.S. Geological Survey
Albuquerque Seismological Laboratory
P.O. Box 82010
Albuquerque, New Mexico 87198 U.S.A.
aringler@usgs.gov

John R. Evans
U.S. Geological Survey
400 Natural Bridges Drive
Santa Cruz, California 95060 U.S.A.

Published Online 23 September 2015