The Ambient Noise Cross-Correlation Data Product: USARRAY TA Vertical Components West of 105°W (ANCC)

V 1.0-0

User's Manual

The User's Manual Prepared by:
Mikhail Barmine
University of Colorado at Boulder
Send comments, question, suggestions to: ***barmin@colorado.edu***
Funding and Support Provided By:
Incorporated Research Institutions for Seismology (IRIS)

Contents

Introduction

This manual describes a data product consisting of vertical component ambient noise cross-correlations and associated metadata for USArray Transportable Array (TA) stations in the western US. Cross-correlations are presented between the vertical components of all _US_TA array virtual network stations west of 105°W that were recorded from Jan l 2005 through Dec 3l 2010. This region encompasses the tectonically deformed part of the western US with 622 TA stations operating in this timeframe (Fig. 1) and produces 171,120 cross-correlation waveforms. Traditional ambient noise data processing procedures (Bensen at al., 2007; Lin et al., 2008) were applied, but small refinements to the method (Lin ct al.. 2011b) were applied to facilitate the potential interpretation of the amplitudes in the cross-correlations. The algorithm was performed with recent software available from IRIS, including **rdseed 5.1** and **SAC 104.1**. In addition, prior to producing the data product data processing programs were refined and accelerated in order to speed and unify the computations. Although the data processing procedures were developed and described in earlier studies that were applied to subsets of the data contained in this product (a number of these papers are in the reference list), all data processing was performed from scratch in order to insure the uniformity and accuracy of the product.

The resulting data product cannot be subsumed naturally within SEED; thus, we have defined a database format to encompass the waveforms and metadata. This manual describes the metadata and database structure.

The data product consists of six components:

- **Data Product Descriptor table** - file `COR_2005.1_2010.1.descriptor`
- **Cross-Correlation Descriptor table** - file `COR_2005.1_2010.12.ccwfdisc`
- **Site Descriptor table** - file `COR_2005.1_2010.1.site`
- **Shortcut Site table** - file `Station.lst`
- **Cross-correlation waveform files** - directory `COR_2005.1_2010.12`
- **This User's Manual** - file `CCManual.pdf`

1. Ambient Noise Data Processing

Bensen et al. (2007) and Lin et al. (2008) described what we refer to as the traditional method of ambient noise data processing and this data product uses these procedures. The method breaks into four principal stages: (1) single station data preparation, (2) cross-correlation and temporal stacking, (3) measurement of dispersion curves (performed with frequency–time analysis for both group and phase speeds) and (4) quality control, including error analysis and selection of the acceptable surface wave dispersion measurements. The procedures have been applied to broad-band seismic data around the world and have been shown to produce robust, largely unbiased measurements of Rayleigh and Love wave phase velocities (e.g., Shapiro et al. 2005; Moschetti et al. 2007; Bensen et al. 2008; Lin et al. 2009; Lin and Ritzwoller 2011a; amongst many others). Simulations and, most importantly, comparisons with earthquake data also establish the reliability of the dispersion measurements and maps (e.g., Lin et al. 2008; Yang and Ritzwoller 2008; Ritzwoller et al. 2011). They have been used to produce 3D models of the crust and uppermost mantle for isotropic shear velocity, radial anisotropy, and azimuthal anisotropy (e.g., Yang et al. 2008; Bensen et al. 2009; Moschetti et al. 2010a, 2010b; Lin et al. 2011a).

The ambient noise cross-correlations or Empirical Green's functions, which define the substance of the data product, arise entirely from the first two stages of this procedure. The principal step in single-station data preparation is 'temporal normalization' which is designed to ameliorate the contamination of the ambient noise signals by earthquakes, instrumental irregularities, and non-stationary noise sources near to stations (such as passing storms and high local sea heights). This has been done by applying 'running absolute-mean' normalization, which we prefer to sign-bit normalization because it allows for tuning to regional earthquake conditions. In particular, for each day of data we filter the data set between 15 and 50 sec period, compute the absolute mean of the data in a running 80 sec long time window, and normalize the central data point of the unfiltered data by the reciprocal of this mean. In addition, we will apply a smooth spectral whitening procedure prior to cross-correlation in order to minimize contamination from spatially localized microseisms and to broaden the measurement band.

Useful quality control metrics that exist after the second stage of data processing include signal-to-noise ratio (SNR) and the inter-station distance. We do not include estimates of SNR in the data product, but Bensen et al. (2007) describe a useful way to measure SNR. We find that inter-station distances greater than 2 – 3 wavelengths are needed to produce reliable dispersion measurements and this distance is included in the metadata for each cross-correlation.

Data gaps introduce the need to define an "effective" time series length. Lin et al. (2011b) show that the square of the rms amplitude of trailing noise provides an accurate proxy for time series length. Here, however, we compute the time series length in hours and include it in the metadata. Cross-correlations are not normalized by this length but if ambient noise amplitudes across inter-stations pairs are to be compared, dividing by this quantity (tsnorm) is recommended.

2. Data Product Descriptor Table

The Data Product Descriptor table is an ASCII file that provides information that is common across the entire data product. This includes information about spatial and temporal boundaries of the data product, the data processing procedures, the software used, and the data processing parameters.

1. Database name: **"`COR_2005.1-2010.12`"**
2. Version: **1.0**
3. Release: **0**
4. Comments: ***"Cross-correlation between Z-component of all _US-TA array west of 105°W from Jan 1 2005 to Dec 31 2010"***
5. Boundaries of region investigation: **31°N, 50°N, 105°W, 125°W**
6. Total stacking period: **2005/001 – 2010/365**
7. Total number of cross-correlations in database**: 171120**
8. Total number of seismic stations: **622**
9. Cross-Correlation Descriptor table file name: **`COR_2005.1_2010.12.ccwfdisc`**
10. Site Descriptor table file name: **`COR_2005.1_2010.12.site`**
11. Shortcut Site table file name: **`Station.lst`**
12. Data processing procedure list:
    - 12.1. Data preprocessing: **yes**
    - 12.2. Instrument response: **yes**
    - 12.3. Temporal normalization: **yes**
    - 12.4. One-bit normalization: **no**
    - 12.5. Spectral whitening: **yes**
    - 12.6. Applying notch correction for 26 sec period: **no**

- 12.7. Time duration normalization: **no**
- 12.8. Signal to noise ratio (SNR) measured: **no**

13. Auxiliary IRIS software:
    - 13.1. **evalresp, v3.3.3** (imbedded in **sac**)
    - 13.2. **rdseed, v5.1**
    - 13.3. **sac, v101.4**
14. Data processing procedures parameters:
    - 14.1. Data preprocessing parameters:
        - 14.1.1. Maximum allowed data gap per day: **10%**
    - 14.2. Instrument response parameters:
        - 12.2.1 Apply band pass filter with corner periods: **170.0, 150.0, 5.0, 4.0** (in sec)
    - 14.3. Temporal normalization parameters:
        - 14.3.1. Filter corner periods: **50.0, 15.0** (in sec)
        - 14.3.2. Sliding time window length (in sec): **80**
        - 14.3.3. Stacking window length: **83000**
    - 14.4. Spectral whitening parameters:
        - 14.4.1. Filter corner periods: **170.0, 150.0, 5.0, 4.0** (in sec)
    - 14.5. Cross-correlation parameters:
        - 14.5.1. Sampling rate: **1.0** (in sec)
        - 14.5.2. Number of samples in cross-correlation: **7201**
        - 14.5.3. Channels list: **LHZ-LHZ**
        - 14.5.4. Cross-correlations lag length: **3600**
        - 14.5.5. Samples type: ***single float***

Some comments on this information follow.

Item 12 relates to data pre-processing prior to cross-correlation. Instrument responses have been removed from both stations, temporal normalization (not one-bit normalization) and spectral whitening have been applied, a notch filter has not been applied around the 26 sec microseism, the time series length correction has not been applied, and SNR has not been measured.

Item 14.1 states that whole days of data are dropped if data gaps are larger than 10% on that day.

Item 14.3 states that the temporal normalization has been computed using data filtered between 15 and 50 sec period in order to filter out earthquakes in a sliding window of length 80 sec. The reciprocal of the absolute mean of the data in that window weights the central data point of the unfiltered data.

Item 14.4 states that smooth spectral whitening occurred in a band between 5 and 170 sec period.

3. Cross-Correlation Descriptor Table

Table 1. Cross-Correlation Descriptor table

Relation:			**ccwfdisc**		
Description:			cross-correlation header and descriptive information		
Attribute name	Field no	Storage type	External format	Character position	Attribute description
sta1	1	c6	a6	1-6	station 1
net1	2	c8	a8	8-15	network for station 1
sta2	3	c6	a6	17-22	station 2
net2	4	c8	a8	24-31	network for station 2
chan1	5	c8	a8	33-40	channel name for station 1
chan2	6	c8	a8	42-49	channel name for station 2
time	7	f8	f12.3	51-62	Minimum lag time of the cross-correlation
wfid	8	i4	i8	64-71	waveform id
endtime	9	f8	f12.3	73-84	time+(nsamp-1)/samprate
nsamp	10	i4	i8	86-93	number of samples in the cross-correlation
samprate	11	f4	f11.7	95-105	sampling rate in samples/sec
snrn	12	f4	f16.6	107-122	negative lag signal-to-noise ratio
snrp	13	f4	f16.6	124-139	positive lag signal-to-noise ratio
sdate	14	i4	i8	141-148	Julian date of beginning of data
edate	15	i4	i8	150-157	Julian date of end of data
stdays	16	i4	i6	159-164	number of days in cross-correlation
range	17	f4	f10.3	166-175	distance between stations in km
tsnorm	18	f8	f14.4	177-190	time series length or amplitude normalization constant (hours)
datatype	19	c2	a2	192-193	numeric storage type
dir	20	c32	a32	195-226	directory of cross-correlation
dfile	21	c32	a32	228-259	data file name
foff	22	i4	i10	261-270	byte offset (after header)
lddate	23	date	a17	272-288	load date

This Cross-Correlation Descriptor table is a large formatted ASCII flat file that provides metadata unique to each cross-correlation waveform and points to the relative location of the cross-correlation waveform on a disk. Each line of the file includes 23 fields that describe the network, station and channel names for both stations in the cross-correlation, a description of the cross correlation time-series itself (e.g., length, sampling rate, start time and end time, etc.), number of days included in the cross-correlation, start and end date of the raw seismograms that were cross-correlated, and so on. The Cross-Correlation Descriptor table points to the directory and file name (fields *dir* and *dfile*) of the unique cross-correlation waveform file. For more detailed information about the fields refer to **Appendix A**.

In this data product, SNR was not calculated for either correlation lag so that the values in the database are set to default values (-1.0 here, see Appendix). The time series amplitude normalization constant (tsnorm) has been calculated, however. This is equal to the number of hours of the ambient noise waveforms involved in stacking process. For the amplitudes of cross-correlations to be compared to one another, the amplitude of the cross-correlation should be divided by this value.

An **Example 1** of the 5 lines taken from the Cross-Correlation Descriptor table **`COR_2005.1_2010.12.ccwfdisc`** are shown below:

Example 1

```
MONP   AZ      OMM    NN      LHZ      LHZ        -3600.000      69
3600.000    7201   1.0000000      -1.000000      -1.000000  2005256
2005284    22    574.584      507.2222 f4 COR_2005.1_2010.12/AZ.MONP
COR_AZ.MONP_NN.OMM.SAC                   632 1334617915.00000

MONP   AZ      WCN    NN      LHZ      LHZ        -3600.000      70
3600.000    7201   1.0000000      -1.000000      -1.000000  2006063
2007296   566    771.762    13049.4444 f4 COR_2005.1_2010.12/AZ.MONP
COR_AZ.MONP_NN.WCN.SAC                   632 1334617915.00000

MONP   AZ      109C   TA      LHZ      LHZ        -3600.000      71
3600.000    7201   1.0000000      -1.000000      -1.000000  2005001
2007296   948     63.871    21856.6667 f4 COR_2005.1_2010.12/AZ.MONP
COR_AZ.MONP_TA.109C.SAC                  632 1334617915.00000

MONP   AZ      112A   TA      LHZ      LHZ        -3600.000      72
3600.000    7201   1.0000000      -1.000000      -1.000000  2007127
2007296   162    177.184     3735.0000 f4 COR_2005.1_2010.12/AZ.MONP
COR_AZ.MONP_TA.112A.SAC                  632 1334617915.00000

MONP   AZ      113A   TA      LHZ      LHZ        -3600.000      73
3600.000    7201   1.0000000      -1.000000      -1.000000  2007129
2007296   160    249.043     3688.8889 f4 COR_2005.1_2010.12/AZ.MONP
COR_AZ.MONP_TA.113A.SAC                  632 1334617915.0000
```

4. Site Descriptor Tables

The area of investigation includes 629 stations of eight seismic networks (Fig. 1): TA, US, IU, CI, AZ, BK, NN and UU forming the _US-TA virtual network. However, data processing was performed for only 622 stations because of the inaccessibility of the following stations from the IRIS DMC data archive at the time of data order (Nov, Dec 2011): DNR, EDW, GOR, MPI, SDR, WER (CI network), and SRU (UU network). This inaccessibility derived from three reasons: stations were offline between the relevant parts of 2005-2010, data were not available from the archive, or only triggered data were available from the archive.

Table 2. Site Descriptor table.

Relation:		**site**			
Description:		station location information			
Attribute name	Field no	Storage type	External format	Character position	Attribute description
vnet	1	c10	a10	1-10	virtual network identifier
net	2	c8	a8	12-19	network identifier
sta	3	c6	a6	21-26	station identifier
ondate	4	c10	a10	28-37	start date
ontime	5	c8	c8	39-46	start time
offdate	6	c10	c10	48-57	off date
offtime	7	c8	c8	59-66	off time
lat	8	f4	f11.6	68-78	latitude
lon	9	f4	f11.6	80-90	longitude
elev	10	f4	f9.1	92-100	elevation
lddate	11	c17	a17	102-118	load date
descr	12	c160	c160	120-279	station description

The Site Descriptor table is an ASCII file with the name **`COR_2005.1_2010.12.site`** which provides information about the network and the location of seismic stations. Each line of the Site file describes a single station location. The single line format is described in **Table 2,** field attributes are described in **Appendix B**. For the region of investigation the Site Descriptor table includes 629 stations. For computational purposes it is convenient to shortcut the site table by using a file that includes only the following fields: *net*, *sta*, *lon* and *lat*, which is created on a monthly basis from the downloaded SEED volumes. We also deliver this shortcut Site table that has been merged over all months (622 stations) as a file with the name `Station.lst`. All 622 stations that have cross-correlation waveforms are shown in **Figure 1**.

An example of several lines of the site table is given below in **Example 2** and the shortcut site table is exemplified in **Example 3**.

Example 2. The 5 first lines from the file **`COR_2005.1_2010.12.site`**

```
_US-TA     AZ        MONP   1998-02-02 00:00:00 2007-10-24 23:59:59   32.892700
-116.422500      1920.0 1334230168.000000 Monument   Peak, Mt. Laguna, CA, USA   Anza
Real-Time Broadband Network
_US-TA     AZ        MONP2  2007-10-24 00:00:00 2599-12-31 23:59:59   32.892000
-116.422300     1000.0 1334230168.000000 Monument   Peak TA Vault, Mt. Laguna, CA, USA
Anza Real-Time Broadband Network
_US-TA     AZ        PFO    1982-10-01 00:00:00 2599-12-31 23:59:59   33.611700
-116.459400    1259.0 1334230168.000000 Pinyon  Flats Observatory, CA, USA  Anza Real-
Time Broadband Network
_US-TA     BK        BDM    1998-11-12 00:00:00 2599-12-31 23:59:59   37.953970
-121.865540      219.8 1334230168.000000 Black   Diamond Mines Park, Antioch, CA, USA
Berkeley Digital Seismic Network  (BDSN)
_US-TA     BK        CMB    1996-09-25 19:19:00 2599-12-31 23:59:59   38.034550
-120.386510     697.0 1334230168.000000 Columbia  College, Columbia, CA, USA  Berkeley
Digital Seismic Network (BDSN)
```

Example 3. The first five lines from the shortcut Site table **`Satation.lst`**

```
AZ   MONP    -116.422500  32.892700
AZ   MONP2   -116.422300  32.892000
AZ   PFO     -116.459400  33.611700
BK   BDM     -121.865500  37.954000
BK   CMB     -120.386500  38.034600
```

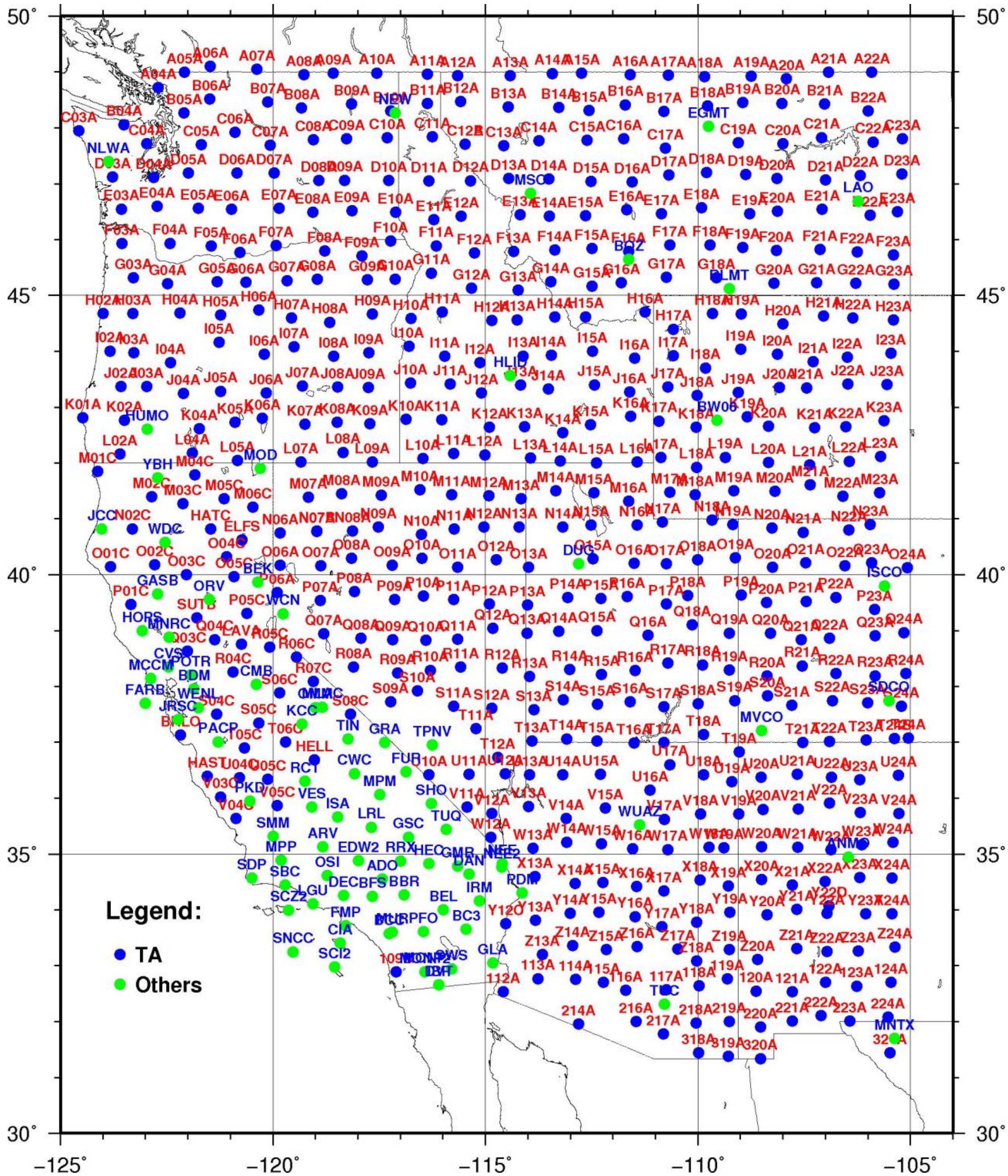

Figure 1. Map of the _US-TA array virtual seismic network (622 stations).

5. Cross-Correlation Waveform files

The final element of the product is the Cross-Correlation Waveform files. Perhaps the simplest storage method would be to keep all files in a single directory. But the maximum number of files in a single directory may vary between different operating systems. For example, for RHEL 5 it is approximately 400000 files. Another issue is losing efficiency of data retrieval that leads to longer processing times.

Due to these reasons we prefer a bi-level or indexed file system to store the waveform data. The Cross-Correlation file name is of the form **`COR_`*`net1.sta1`*`_`*`net2.sta2`*`.SAC`**, where ***`net1,sta1`*** are the network and station name for the first station and ***`net2,sta2`*** are the network and station name for the second station in the cross-correlation. For example, **`COR_TA.F05A_TA.X15A.SAC`** is the name of the file containing the cross-correlation between TA stations F05A and X15A**.** The waveform for this example is shown in **Figure 2**. All cross-correlations in the product range between times of plus or minus 3600 seconds.

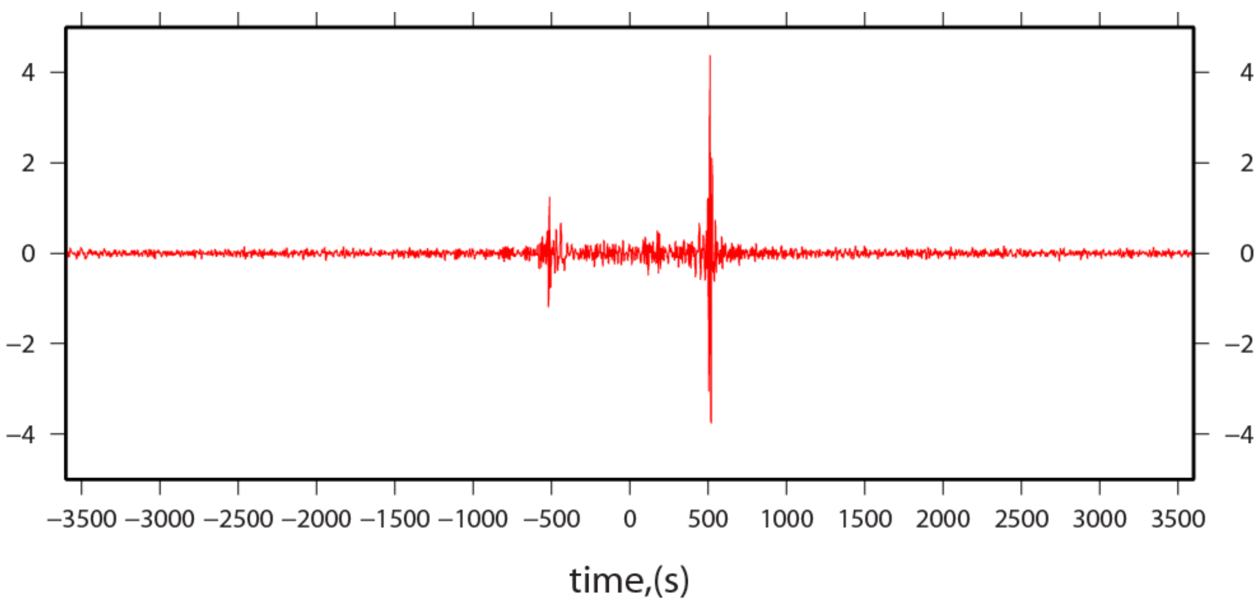

Figure 2. SAC file, cross-correlation function: **`TA.F05A-TA.X15A`**

In each case, ***`net1.sta1`*** alphabetically precedes ***`net2.sta2`***. The order the network-station names in the file name is important because it determines the meaning of the positive and negative lags in the cross-correlations. The positive lag corresponds to waves that travel from

net1.sta1 to ***net2.sta2***. The negative lag corresponds to waves that travel from ***net2.sta2*** to ***net1.sta1***.

In order to reduce the number of files in the data product, we create under the root data directory subdirectories with all possible values of the first station identifier (***net1.sta1***) extracted from all file names. After that we place each file into the subdirectory with the first station name of that file name. For the previous example we will have the following path to file:

....../COR_2005.1_2010.12/TA.F05A/COR_TA.F05A_TA.X15A.SAC

The directory **COR_2005.1_2010.12** is the root data directory that coincides with the database name. Note, there is no cross-correlation between stations **TA.F05A** and **TA.X15A** under the directory for network-station **TA.X15A**. Therefore, if users wish to find all cross-correlations between **TA** station **X15A** and all other stations, they will have to look in all of the station sub-directories. This is one of the uses of the cross-correlation descriptor table. It provides tabular look-up for the locations of all cross-correlation files.

Each cross-correlation data file is a SAC binary file that includes in the header the most relevant information from the Cross-Correlation Descriptor table. Stripping the first 632 bytes from the beginning of each file would allow treating this file as a regular binary file with floating point single precision values.

An example of a SAC header for the cross-correlation file **TA.F05A-TA.X15A** is listed below. There are some changes in the SAC header values compared to a standard SAC header interpretation, particularly motivated by the need to store information for two stations rather than an event and a station. Note, that the **KZDATE** and **KZTIME** fields in the SAC header describe a fictitious reference time point 0 hours of Jan 1 2000 for all cross-correlations. This is done to be in compliance with the SAC timing policy.

A printout of an example SAC header is presented below. Again it is for the station pair **F05A-X15A**.

```
 FILE: COR_TA.F05A_TA.X15A.SAC - 1
-----------------------------
       NPTS = 7201                # number of samples
          B = -3.600000e+03       # start time
          E = 3.600000e+03        # end time
     IFTYPE = TIME SERIES FILE
      LEVEN = TRUE
      DELTA = 1.000000e+00        # time step
       IDEP = UNKNOWN             # unknown type of dependent variable
     DEPMIN = -5.994123e+00
     DEPMAX = 6.324740e+00
     DEPMEN = 1.468014e-05
     KZDATE = JAN 01 (001), 2000  # date of zero time shift
     KZTIME = 00:00:00.000        # time of zero time shift
      KSTNM = X15A                # second station name
      CMPAZ = 0.000000e+00
     CMPINC = 0.000000e+00
       STLA = 3.448660e+01        # second station latitude
       STLO = -1.122368e+02       # second station longitude
       STEL = 4.548000e+02
       STDP = 0.000000e+00
      KEVNM = F05A                # first station name
       EVLA = 4.588390e+01        # first station latitude
       EVLO = -1.214588e+02       # first station longitude
       DIST = 1.487096e+03        # distance between stations, km
         AZ = 1.450983e+02        # 1st station to 2nd station azimuth (°).
        BAZ = 3.310629e+02        # back azimuth (°)
      GCARC = 1.337737e+01
     LOVROK = TRUE
      USER0 = 6.260000e+02        # number of stacking days
      USER1 = 1.443278e+04        # time normalization constant, (hours)
     KUSER1 = TA                  # network for the second station
     KUSER2 = LHZ                 # second station channel name
      NVHDR = 6
      SCALE = 1.000000e+00
      NORID = 0
      NEVID = 0
     LPSPOL = FALSE
     LCALDA = TRUE
     KCMPNM = LHZ                 # first station channel name
     KNETWK = TA                  # network for the first station
     KT0 = 2006082                # stacking begin YYYY/MM
     KT1 = 2008025                # stacking end YYYY/MM
```

Appendix A
Cross-Correlation Descriptor table field attribute description

---

Name: *sta1*

Description: Station code, or name for the first station of the cross-correlation waveform. Station belongs to *net1*. Generally only three or four characters are used.

NA Value: A valid entry is required.

Range: Any uppercase string of up to 6 characters.

Name: *net1*

Description: Unique network identifier. This character string is the name of seismic network containing station *sta1*. Usually this is a string of two bytes. One example is TA.

NA Value: A valid entry is required.

Range: Any uppercase string of up to 8 characters.

Name: *sta2*

Description: Station code, or name for the second station of the cross-correlation waveform . Station belongs to *net2*. Generally only three or four characters are used.

NA Value: A valid entry is required.

Range: Any uppercase string of up to 6 characters.

Name: *net2*

Description: Unique network identifier. This character string is the name of seismic network containing station *sta2*. Usually this is a string of two bytes. One example is TA.

NA Value: A valid entry is required.

Range: Any uppercase string of up to 8 characters.

Name: *chan1*

Description: Channel identifier. This is usually a three-character code, which, taken together with *sta1*, *net1*, uniquely identifies the source of seismic data, including the geographic location and spatial orientation. One example is BHZ.

NA Value: A valid entry is required.

Range: Any uppercase string of up to 8 characters.

Name: *chan2*

Description: Channel identifier. This is usually a three-character code, which, taken together with *sta2*, *net2*, uniquely identifies the source of seismic data, including the geographic location and spatial orientation. One example is BHZ.

NA Value: A valid entry is required.

Range: Any uppercase string of up to 8 characters.

Name: *time*

Description: The minimal value of cross-correlation argument. Refer to the first sample of cross-correlation waveform. The double precision floating number allows 12 decimal digits.
NA Value: A valid entry is required.

Units: seconds.

Range: *time* < 0

Name: *wfid*

Description: Cross-correlation identifier. The key field is unique identifier for each file line.
NA Value: : A valid entry is required.

Range: *wfid* > 0

Name: *endtime*

Description: The maximal value of cross-correlation argument. Refer to the last sample in the cross-correlation waveform. The double precision floating number allows 15 decimal digits.
NA Value: A valid entry is required.

Units: seconds.

Range: *endtime* = (*nsamp* – 1)/*samprate*

Name: *nsamp*

Description: Number of samples. This quantity is the number of samples in waveform segment**.**
NA Value: NOT ALLOWED. A valid entry is required.

Range: *nsamp* > 0

Name: *samprate*

Description: Sampling rate. The attribute is the sample rate in samples/second.
NA Value: NOT ALLOWED. A valid entry is required.

Units: 1/seconds

Range: *samprate* > 0

Name: *snrn, snrp*

Description: Signal to noise ratio. Not used yet, for further extant ions.
NA Value: -1.0

Range: N/A

Name: *sdate*

Description: Julian date. This attribute is the starting date of waveform stacking form of **yyyyddd**, Where **yyyy** is a year and **ddd** is the day of the year starting from 1.
NA Value: -1

Range: Any value.

Name: *edate*

Description: Similar to *sdate*. This attribute is the ending Julian date of cross-correlation stacking.
NA Value: -1
Range: *edate* > *sdate*

Name: *stdays*

Description: This attribute defines the total number of days in cross-correlation stacking.
NA Value: NOT ALLOWED. A valid entry is required.
Range: *stdays* ≥ 1

Name: *range*

Description: This attribute define the big arc distance between pair of stations *sta1* and *sta2* .
Units: kilometers
NA Value: NOT ALLOWED. A valid entry is required.
Range: *range* > 0

Name: *tsnorm*

Description: The total size of ambient noise waveforms involved in stacking process.
Units: hours
NA Value: NOT ALLOWED. A valid entry is required.
Range: *tsnorm* > 0

Name: *datatype*

Description: Numeric storage data. This character string specifies the format of binary waveform file. Currently only the data type t4 and f4 are used.
NA Value: NOT ALLOWED. A valid entry is required.
Range: Currently recognized types (lowercase) are:

legal datatype values		
Datatype value	Size (bytes)	description
t4	4	SUN IEEE single precision real, (**BIGENDIAN**)
f4	4	PC/DEC/VAX IEEE single precision real, (**LOWENDIAN**)

Name: *dir*

Description: Directory. This is a relative path from the current directory containing the **ccwfdisc** relation to Cross Correlation database subdirectory. Example:
`......./COR_2005.1-2010.12/TA.N12A`
NA Value: NOT ALLOWED. A valid entry is required.
Range: Any string up to 32 bytes.

Name: *dfile*

Description: Data file. This attribute is the file name of a binary waveform file. The full relative path to file is constructed as *dir/dfile* .

NA Value: NOT ALLOWED. A valid entry is required.

Range: Any string up to 32 bytes.

Name: *foff*

Description: File offset. This is the byte offset of binary waveform data within data file. See *dir* and *dfile*. In case binary SAC file *foff* is equal to **632**.

NA Value: NOT ALLOWED. A valid entry is required.

Range: *foff* ≥ 0

Name: *lddate*

Description: Load date. This attribute is the date and time the record was created.

NA Value: "_"

Range: Any string of up to 17 characters.

Appendix B
Site Descriptor table field attribute description

Name: *vnet*
Description: IRIS virtual network identifier. Virtual network is a common name for a set of networks joint together by some reason.
NA Value: A valid entry is required.
Range: Any uppercase string of up to 10 characters.

Name: *net1*
Description: Unique network identifier. This character string is the name of seismic network containing station *sta*. Usually, this is a string of two bytes. One example is TA.
NA Value: A valid entry is required.
Range: Any uppercase string of up to 8 characters.

Name: *sta*
Description: Station code, or name for the first station of cross-correlation waveform. Station belongs to *net1*. Generally only three or four characters are used.
NA Value: A valid entry is required.
Range: Any uppercase string of up to 6 characters.

Name: *ondate*
Description: Start date of operation. This attribute is form of **yyyy-mm-dd**, where **yyyy** is a year, **mm** a months, and **dd** day in month.
NA Value: A valid entry is required.
Range: String of 10 byte long.

Name: *ontime*
Description: Start time of operation. This attribute is form of **hh:mm:ss**, where **hh** is hours, **mm** minutes, and **ss** seconds.
NA Value: A valid entry is required.
Range: String of 8 byte long.

Name: *offdate*
Description: Stop date of operation. This attribute is form of **yyyy-mm-dd**, where **yyyy** is a year, **mm** a months, and **dd** day in month.
NA Value: A valid entry is required.
Range: String of 10 byte long.

Name: *offtime*

Description: Stop time of operation. This attribute is form of **hh:mm:ss**, where **hh** is hours, **mm** minutes, and **ss** seconds.

NA Value: A valid entry is required.

Range: String of 8 byte long.

Name: *lat*

Description: Latitude. This attribute is the geographic latitude. Locations north of the equator have positive latitudes.

NA Value: A valid entry is required.

Range: $-90.0 \leq lat \leq 90.0$

Name: *lon*

Description: Longitude. This attribute is the geographic longitude in degrees. Longitudes are measured positive east of the Greenwich meridian.

NA Value: A valid entry is required.

Range: $-180.0 \leq lon \leq 180.0$

Name: *elev*

Description: Elevation. This attribute is the elevation if a seismic station relative to mean sea level.

NA Value: A valid entry is required.

Units: meters

Range: $-10.0 \leq elev \leq 10.0$

Name: *lddate*

Description: Load date. This attribute is the date and time the record was created.

NA Value: "_"

Range: Any string of up to 17 characters.

Name: *descr*

Description: Station description.

Range: Any string of up to 160 characters.

References

Bensen, G.D., M.H. Ritzwoller, M.P. Barmin, A.L. Levshin, F. Lin, M.P. Moschetti, N.M. Shapiro, and Y. Yang, Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, 169, 1239-1260, doi: 10.1111/j.1365-246X.2007.03374.x, 2007.

Bensen, G.D., M.H. Ritzwoller, and N.M. Shapiro, Broad-band ambient noise surface wave tomography across the United States, *J. Geophys. Res.*, 113, B05306, 21 pages, doi:10.1029/2007JB005248, 2008.

Bensen, G.D., M.H. Ritzwoller, and Y. Yang, A 3D shear velocity model of the crust and uppermost mantle beneath the United States from ambient seismic noise, *Geophys. J. Int.*, 177(3), 1177-1196, 2009.

Lin, F.C. and M.H. Ritzwoller, Helmholtz surface wave tomography for isotropic and azimuthally anisotropic structure, *Geophys. J. Int.*, 186, 1104-1120, doi:10.1111/j.1365-246X.2011.05070.x, 2011a.

Lin, F.C. and M.H. Ritzwoller, Apparent anisotropy in inhomogeneous isotropic media, *Geophys. J. Int.*, 186(3), 1205-1219, doi:10.1111/j.1365-246X.2011.05100.x, 2011b.

Lin, F., M.P. Moschetti, and M.H. Ritzwoller, Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps, *Geophys. J. Int.*, doi:10.1111/j1365-246X.2008.03720.x, 2008.

Lin, F.-C., M.H. Ritzwoller, and R. Snieder, Eikonal Tomography: Surface wave tomography by phase-front tracking across a regional broad-band seismic array, *Geophys. J. Int.*, 177(3), 1091-1110, 2009.

Lin, F.C., M.H. Ritzwoller, Y. Yang, M.P. Moschetti, and M.J. Fouch, Complex and variable crustal and uppermost mantle seismic anisotropy in the western United States, *Nature Geoscience*, Vol 4, Issue 1, 55-61, Jan 2011a.

Lin, F.C., M.H. Ritzwoller, and W. Shen, On the reliability of attenuation measurements from ambient noise cross-correlations, *Geophys. Res. Letts.*, 38, L11303, doi:10.1029/2011GL047366, 2011b.

Moschetti, M.P., M.H. Ritzwoller, and N.M. Shapiro, Surface wave tomography of the western United States from ambient seismic noise: Rayleigh wave group velocity maps, *Geochem., Geophys., Geosys.*, 8, Q08010, doi:10.1029/2007GC001655, 2007.

Moschetti, M.P., M.H. Ritzwoller, and F.C. Lin, Seismic evidence for widespread crustal deformation caused by extension in the western USA, *Nature*, 464, Number 7290, 885-889, 8 April 2010a.

Moschetti, M.P., M.H. Ritzwoller, F.C. Lin, and Y. Yang, Crustal shear velocity structure of the western US inferred from amient noise and earthquake data, *J. Geophys. Res.*, 115, B10306, doi:10.1029/2010JB007448, 2010b.

Ritzwoller, M.H., F.C. Lin, and W. Shen, Ambient noise tomography with a large seismic array, *Compte Rendus Geoscience*, 13 pages, doi:10.1016/j.crte.2011.03.007, 2011.

Shapiro, N.M. M. Campillo, L. Stehly, and M.H. Ritzwoller, High resolution surface wave tomography from ambient seismic noise, *Science*, 307(5715), 1615-1618, 11 March 2005.

Yang, Y. and M.H. Ritzwoller, The characteristics of ambient seismic noise as a source for surface wave tomography, *Geochem.*, *Geophys.*, *Geosys.*, 9(2), Q02008, 18 pages, doi:10.1029/2007GC001814, 2008.

Yang, Y., M.H. Ritzwoller, F.-C. Lin, M.P. Moschetti, and N.M. Shapiro, The structure of the crust and uppermost mantle beneath the western US revealed by ambient noise and earthquake tomography, *J. Geophys. Res.*, 113, B12310, 2008.