

# Downloading OBS Data from IRIS and Processing

Chao An (anchao@sjtu.edu.cn, April 2024)

This document describes a procedure to download OBS data from IRIS and preprocess the data. Python scripts and SAC software are used. All the Python scripts are provided in folder *DataProcessingFiles*. There are approximately 5 steps: data downloading, resampling, synchronizing, merging, and removing instrumental responses. For each step, copy everything from *DataProcessingFiles* to the working folder, modify the parameters in the corresponding python script, and run it. Below is a detailed instruction of the procedure. Note that python modulus "requests" is required. In this document, folders are marked in blue and executable programs are in red.

## 1 Download OBS data and instrumental responses from IRIS

Create a folder with name *OBSData*. Run `1_irisDataDownloadLongNoise.py` or `1_irisDataDownloadLongNoise_AutoRestart.py`. The latter simply restarts the former if an error occurs, so you don't have to manually restart the downloading process.

Parameters in script `1_irisDataDownloadLongNoise.py` are shown below, and explained in the following table.

```
26 GetStationListFromWeb = True
27
28 StartDate = [2018, 1, 1, 0, 0, 0]
29 EndDate   = [2019, 12, 31, 0, 0, 0]
30 Network   = 'X0'
31
32 ChannelsInOrder = ['HH1,HH2,HHZ,HDH', 'BH1,BH2,BHZ,BDH', 'LH1,LH2,LHZ,LDH']
33 DecimateData = True
34 fSampleRate = 5 # if channels are H channels, samplerate = fSampleRate
35 Exceptions = [['LT',25]]
36
37
38 segment_length_seconds = 5*24*60*60
39 Endian = 'little'
```

<i>GetStationListFromWeb</i>	True if downloading station and channel list from IRIS False if reading stations and channels from <i>AllStations.txt</i>
<i>StartDate, EndDate, Network</i>	Download data of this network code during this period

### ChannelsInOrder

For each station, check data availability of channel group in order. Download data of the first available channel group

### DecimateData

Download data at lower sampling rate for H\* channels

### fSampleRate

For H\* channels, download data at fSampleRate Hz

### Exceptions

Stations starting with Exceptions[0], sampling is Exceptions[1]

### segment\_length\_seconds

< 7 days. IRIS doesn't allow downloading longer data.

### Endian

= 'little' for MAC/Linux/WIN systems

Input Files : NONE.

Output Files : *AllStationsInfo.txt* , *AllStations.txt* , *NoDataFiles.txt* , *Finish.txt*

SAC files, example:

```
[ca298@localhost ~]$ ls LT01*
LT01_XO_HDH.2018.05.20.23.50.00-2018.05.26.00.10.00.SAC
LT01_XO_HDH.2018.05.25.23.50.00-2018.05.31.00.10.00.SAC
LT01_XO_HDH.2018.05.30.23.50.00-2018.06.05.00.10.00.SAC
LT01_XO_HDH.2018.06.04.23.50.00-2018.06.10.00.10.00.SAC
LT01_XO_HDH.2018.06.09.23.50.00-2018.06.15.00.10.00.SAC
LT01_XO_HDH.2018.06.14.23.50.00-2018.06.20.00.10.00.SAC
LT01_XO_HDH.2018.06.19.23.50.00-2018.06.25.00.10.00.SAC
LT01_XO_HDH.2018.06.24.23.50.00-2018.06.30.00.10.00.SAC
LT01_XO_HDH.2018.06.29.23.50.00-2018.07.05.00.10.00.SAC
LT01_XO_HDH.2018.07.04.23.50.00-2018.07.10.00.10.00.SAC
LT01_XO_HDH.2018.07.09.23.50.00-2018.07.15.00.10.00.SAC
LT01_XO_HDH.2018.07.14.23.50.00-2018.07.20.00.10.00.SAC
LT01_XO_HDH.2018.07.19.23.50.00-2018.07.25.00.10.00.SAC
LT01_XO_HDH.2018.07.24.23.50.00-2018.07.30.00.10.00.SAC
LT01_XO_HDH.2018.07.29.23.50.00-2018.08.04.00.10.00.SAC
LT01_XO_HDH.2018.08.03.23.50.00-2018.08.09.00.10.00.SAC
LT01_XO_HDH.2018.08.08.23.50.00-2018.08.14.00.10.00.SAC
LT01_XO_HDH.2018.08.13.23.50.00-2018.08.19.00.10.00.SAC
LT01_XO_HDH.2018.08.18.23.50.00-2018.08.24.00.10.00.SAC
LT01_XO_HDH.2018.08.23.23.50.00-2018.08.29.00.10.00.SAC
LT01_XO_HDH.2018.08.28.23.50.00-2018.09.03.00.10.00.SAC
LT01_XO_HDH.2018.09.02.23.50.00-2018.09.08.00.10.00.SAC
LT01_XO_HDH.2018.09.07.23.50.00-2018.09.13.00.10.00.SAC
LT01_XO_HDH.2018.09.12.23.50.00-2018.09.18.00.10.00.SAC
LT01_XO_HDH.2018.09.17.23.50.00-2018.09.23.00.10.00.SAC
LT01_XO_HDH.2018.09.22.23.50.00-2018.09.28.00.10.00.SAC
LT01_XO_HDH.2018.09.27.23.50.00-2018.10.03.00.10.00.SAC
LT01_XO_HDH.2018.10.02.23.50.00-2018.10.08.00.10.00.SAC
LT01_XO_HDH.2018.10.07.23.50.00-2018.10.13.00.10.00.SAC
LT01_XO_HDH.2018.10.12.23.50.00-2018.10.18.00.10.00.SAC
LT01_XO_HDH.2018.10.17.23.50.00-2018.10.23.00.10.00.SAC
LT01_XO_HH1.2018.11.16.23.50.00-2018.11.22.00.10.00.SAC
LT01_XO_HH1.2018.11.21.23.50.00-2018.11.27.00.10.00.SAC
LT01_XO_HH1.2018.11.26.23.50.00-2018.12.02.00.10.00.SAC
LT01_XO_HH1.2018.12.01.23.50.00-2018.12.07.00.10.00.SAC
LT01_XO_HH1.2018.12.06.23.50.00-2018.12.12.00.10.00.SAC
LT01_XO_HH1.2018.12.11.23.50.00-2018.12.17.00.10.00.SAC
LT01_XO_HH1.2018.12.16.23.50.00-2018.12.22.00.10.00.SAC
LT01_XO_HH1.2018.12.21.23.50.00-2018.12.27.00.10.00.SAC
LT01_XO_HH1.2018.12.26.23.50.00-2019.01.01.00.10.00.SAC
LT01_XO_HH1.2018.12.31.23.50.00-2019.01.06.00.10.00.SAC
LT01_XO_HH1.2019.01.05.23.50.00-2019.01.11.00.10.00.SAC
LT01_XO_HH1.2019.01.10.23.50.00-2019.01.16.00.10.00.SAC
LT01_XO_HH1.2019.01.15.23.50.00-2019.01.21.00.10.00.SAC
LT01_XO_HH1.2019.01.20.23.50.00-2019.01.26.00.10.00.SAC
LT01_XO_HH1.2019.01.25.23.50.00-2019.01.31.00.10.00.SAC
LT01_XO_HH1.2019.01.30.23.50.00-2019.02.05.00.10.00.SAC
LT01_XO_HH1.2019.02.04.23.50.00-2019.02.10.00.10.00.SAC
LT01_XO_HH1.2019.02.09.23.50.00-2019.02.15.00.10.00.SAC
LT01_XO_HH1.2019.02.14.23.50.00-2019.02.20.00.10.00.SAC
LT01_XO_HH1.2019.02.19.23.50.00-2019.02.25.00.10.00.SAC
LT01_XO_HH1.2019.02.24.23.50.00-2019.03.02.00.10.00.SAC
LT01_XO_HH1.2019.03.01.23.50.00-2019.03.07.00.10.00.SAC
LT01_XO_HH1.2019.03.06.23.50.00-2019.03.12.00.10.00.SAC
LT01_XO_HH1.2019.03.11.23.50.00-2019.03.17.00.10.00.SAC
LT01_XO_HH1.2019.03.16.23.50.00-2019.03.22.00.10.00.SAC
LT01_XO_HH1.2019.03.21.23.50.00-2019.03.27.00.10.00.SAC
LT01_XO_HH1.2019.03.26.23.50.00-2019.04.01.00.10.00.SAC
LT01_XO_HH1.2019.03.31.23.50.00-2019.04.06.00.10.00.SAC
LT01_XO_HH1.2019.04.05.23.50.00-2019.04.11.00.10.00.SAC
LT01_XO_HH1.2019.04.10.23.50.00-2019.04.16.00.10.00.SAC
LT01_XO_HH1.2019.04.15.23.50.00-2019.04.21.00.10.00.SAC
LT01_XO_HH2.2019.04.05.23.50.00-2019.04.11.00.10.00.SAC
LT01_XO_HH2.2019.04.10.23.50.00-2019.04.16.00.10.00.SAC
LT01_XO_HH2.2019.04.15.23.50.00-2019.04.21.00.10.00.SAC
LT01_XO_HH2.2019.04.20.23.50.00-2019.04.26.00.10.00.SAC
LT01_XO_HH2.2019.04.25.23.50.00-2019.05.01.00.10.00.SAC
LT01_XO_HH2.2019.04.30.23.50.00-2019.05.06.00.10.00.SAC
LT01_XO_HH2.2019.05.05.23.50.00-2019.05.11.00.10.00.SAC
LT01_XO_HH2.2019.05.10.23.50.00-2019.05.16.00.10.00.SAC
LT01_XO_HH2.2019.05.15.23.50.00-2019.05.21.00.10.00.SAC
LT01_XO_HH2.2019.05.20.23.50.00-2019.05.26.00.10.00.SAC
LT01_XO_HH2.2019.05.25.23.50.00-2019.05.31.00.10.00.SAC
LT01_XO_HH2.2019.05.30.23.50.00-2019.06.05.00.10.00.SAC
LT01_XO_HH2.2019.06.04.23.50.00-2019.06.10.00.10.00.SAC
LT01_XO_HH2.2019.06.09.23.50.00-2019.06.15.00.10.00.SAC
LT01_XO_HH2.2019.06.14.23.50.00-2019.06.20.00.10.00.SAC
LT01_XO_HH2.2019.06.19.23.50.00-2019.06.25.00.10.00.SAC
LT01_XO_HH2.2019.06.24.23.50.00-2019.06.30.00.10.00.SAC
LT01_XO_HH2.2019.06.29.23.50.00-2019.07.05.00.10.00.SAC
LT01_XO_HH2.2019.07.04.23.50.00-2019.07.10.00.10.00.SAC
LT01_XO_HH2.2019.07.09.23.50.00-2019.07.15.00.10.00.SAC
LT01_XO_HH2.2019.07.14.23.50.00-2019.07.20.00.10.00.SAC
LT01_XO_HH2.2019.07.19.23.50.00-2019.07.25.00.10.00.SAC
LT01_XO_HH2.2019.07.24.23.50.00-2019.07.30.00.10.00.SAC
LT01_XO_HH2.2019.07.29.23.50.00-2019.08.04.00.10.00.SAC
LT01_XO_HH2.2019.08.03.23.50.00-2019.08.09.00.10.00.SAC
LT01_XO_HH2.2019.08.08.23.50.00-2019.08.14.00.10.00.SAC
LT01_XO_HH2.2019.08.13.23.50.00-2019.08.19.00.10.00.SAC
LT01_XO_HH2.2019.08.18.23.50.00-2019.08.24.00.10.00.SAC
LT01_XO_HH2.2019.08.23.23.50.00-2019.08.29.00.10.00.SAC
LT01_XO_HH2.2019.08.28.23.50.00-2019.09.03.00.10.00.SAC
LT01_XO_HH2.2019.09.02.23.50.00-2019.09.08.00.10.00.SAC
LT01_XO_HH2.2019.09.07.23.50.00-2019.09.13.00.10.00.SAC
LT01_XO_HH2.2019.09.12.23.50.00-2019.09.18.00.10.00.SAC
LT01_XO_HH2.2019.09.17.23.50.00-2019.09.23.00.10.00.SAC
LT01_XO_HH2.2019.09.22.23.50.00-2019.09.28.00.10.00.SAC
LT01_XO_HH2.2019.09.27.23.50.00-2019.10.03.00.10.00.SAC
LT01_XO_HH2.2019.10.02.23.50.00-2019.10.08.00.10.00.SAC
LT01_XO_HH2.2019.10.07.23.50.00-2019.10.13.00.10.00.SAC
LT01_XO_HH2.2019.10.12.23.50.00-2019.10.18.00.10.00.SAC
LT01_XO_HH2.2019.10.17.23.50.00-2019.10.23.00.10.00.SAC
LT01_XO_HH2.2019.10.22.23.50.00-2019.10.28.00.10.00.SAC
LT01_XO_HH2.2019.10.27.23.50.00-2019.11.01.00.10.00.SAC
```

After the downloading process is finished (or during the process), you can run script [1\\_irisDataDownloadLongNoise\\_CheckSACFiles.py](#) to check if the downloaded SAC files are good. Due to internet problems, some files might be broken. This script uses SAC software and attempts to read the SAC files. If errors occurs, it means these SAC files are broken, and they are removed. Rerun script [1\\_irisDataDownloadLongNoise\\_AutoRestart.py](#) to redownload the missing data.

Create a folder with name `instrument-responses`. Similar to downloading OBS data, run [1\\_irisRESPDownloadLongNoise.py](#) to download RESP files from IRIS. Example of output:

```
[ca298@localhost instrument-responses]$ ls
1_irisRESPDownloadLongNoise.py  RESP.LA39.XO.HDH  RESP.LT03.XO.HDH  RESP.LT16.XO.HDH  RESP.WD55.XO.HDH  RESP.WD66.XO.HDH
AllStationsInfo.txt             RESP.LA39.XO.HH1  RESP.LT03.XO.HH1  RESP.LT16.XO.HH1  RESP.WD55.XO.HH1  RESP.WD66.XO.HH1
AllStations.txt                 RESP.LA39.XO.HH2  RESP.LT03.XO.HH2  RESP.LT16.XO.HH2  RESP.WD55.XO.HH2  RESP.WD66.XO.HH2
requests                         RESP.LA39.XO.HHZ  RESP.LT03.XO.HHZ  RESP.LT16.XO.HHZ  RESP.WD55.XO.HHZ  RESP.WD66.XO.HHZ
RESP.LA21.XO.HDH                 RESP.LD35.XO.HDH  RESP.LT04.XO.HDH  RESP.LT20.XO.HDH  RESP.WD56.XO.HDH  RESP.WD67.XO.HDH
RESP.LA21.XO.HH1                 RESP.LD35.XO.HH1  RESP.LT04.XO.HH1  RESP.LT20.XO.HH1  RESP.WD56.XO.HH1  RESP.WD67.XO.HH1
RESP.LA21.XO.HH2                 RESP.LD35.XO.HH2  RESP.LT04.XO.HH2  RESP.LT20.XO.HH2  RESP.WD56.XO.HH2  RESP.WD67.XO.HH2
RESP.LA21.XO.HHZ                 RESP.LD35.XO.HHZ  RESP.LT04.XO.HHZ  RESP.LT20.XO.HHZ  RESP.WD56.XO.HHZ  RESP.WD67.XO.HHZ
RESP.LA23.XO.HDH                 RESP.LD36.XO.HDH  RESP.LT05.XO.HDH  RESP.WD46.XO.HDH  RESP.WD57.XO.HDH  RESP.WD68.XO.HDH
```

**Resampling, synchronizing and merging the data are optional, and the order of doing these operations can be adjusted. Sometimes resampling the data before merging causes spikes between two merged files. So you may try synchronizing and merging the data first and then resampling.**

## 2 Resample OBS data

Create a folder with name `OBSData_Resampled`. Run `2_resampleNoiseData.py` to resample the OBS data. Running the script without any arguments will resample all the SAC files. If you want to only resample certain stations, provide station names as arguments , e.g.,

```
./2_resampleNoiseData.py LT01 LT20
```

Parameters in script `2_resampleNoiseData.py` are shown below, and explained in the following table. Note that the decimation factors may be different for different channels, and must be reviewed manually. For example, the pressure channel may have a different original sampling rate than the 12Z channels, and the same sampling rate is required for all channels after resampling. The script will pause and wait for confirmation to continue.

```

8 From = './OBSData'
9 To   = '.'
10
11 TargetDT = 0.2
12 nDecimateDefault = [[2,2],[2,4],[2,4],[2,4]] # decimate 2 to 7, can have different number of channels
13 Exceptions = [['XXX', [1], [1], [1], [1]], \
14               ['XXX', [1], [1], [1], [1]], \
15               ['YYY', [2, 4], [2, 5], [2, 5], [2, 5]], \
16               ['XXX', [1], [1], [1], [1]]]

```

<i>From</i>	Folder of original OBS data
<i>To</i>	Folder of OBS data after resampling
<i>TargetDT</i>	Expected dt after resampling
<i>nDecimateDefault</i>	Factors for downsampling data. For each channel, the decimation factor is 2*2, 2*4, 2*4, 2*4, respectively. The number of channels is not necessarily equal to 4.
<i>Exceptions</i>	Stations starting with Exceptions[i][0], factors are Exceptions[i][1:]

Input Files : OBS data downloaded from IRIS (SAC files).

NOTE: script identifies SAC files, stations and channels according to file names.

Output Files : resampled OBS data (SAC files of same names).

### 3 Synchronize OBS data

Create a folder with name `OBSData_Resampled_Synced`. Run `3_syncNoiseData.py` to synchronize the OBS data between different channels . Running the script without arguments will process all the SAC files. Again, you can use station names as arguments to only process specified stations, e.g.,

`./3_syncNoiseData.py LT01 LT20`

Parameters in script `3_syncNoiseData.py` are shown below, and explained in the following table.

```
8 From = './OBSData_Resampled'
9 To = '.'
10
11 StrictFileTimeLengthCheck = True
12
13 AllowedChannelsInOrder = [['BH1', 'BH2', 'BHZ', 'BDH'], ['LH1', 'LH2', 'LHZ', 'LDH'], ['HH1', 'HH2', 'HHZ', 'HDH'], \
14     ['BHE', 'BHN', 'BHZ', 'BDH'], ['LHE', 'LHN', 'LHZ', 'LDH'], ['HHE', 'HNN', 'HHZ', 'HDH'], \
15     ['BH1', 'BH2', 'BHZ', 'HYD'], \
16     ['SHX', 'SHY', 'SHZ', 'HYD'], \
17     ['BHE', 'BHN', 'BHZ', 'BHH']]
```

<i>From</i>	Folder of OBS data after removing instrumental response
<i>To</i>	Folder of OBS data after synchronizing
<i>StrictFileTimeLengthCheck</i>	If = True, script will exit if paired SAC files have different start or end time
<i>AllowedChannelsInOrder</i>	For each station, check if data exists for these channel groups in order. Synchronize data of the first available channel group

Input Files : OBS data (SAC files).

NOTE: script identifies SAC files, stations and channels according to file names.

Output Files : OBS data synchronized (SAC files; Same t0, data length, dt for all channels).

Errors may appear when synchronizing the data due to quality of the original data. For instance, when processing the Alaska XO data, station LA29 only has 1,2,z channels, and the pressure channel is missing. In this case, it requires manual restart of the script. Basically it is needed to remove all the LA29 data in the *From* folder (i.e., `OBSData_Resampled`) and re-run the this script.

After synchronizing, every station paired channels, e.g., 1, 2, z and p. The number of channels is not necessarily equal to 4, depending on the prescribed parameter `AllowedChannelsInOrder`. Generally, the files of paired channels have the same reference time, beginning time, end time and sampling rate, although these headers may not be exactly identical due to computer storing precisions.

## 4 Merge OBS data

Merging data into long segments reduces the number of files and makes the data folder neat. However, long segments cause SAC warnings because the beginning and ending time might be too large. So this step can be skipped if the original data is already long enough for later analysis.

Create a folder with name `OBSData_Resampled_Synced_Merged`. Run `4_mergeNoiseData.py` to merge the OBS data into longer and fewer files. Running the script without arguments will merge all the SAC files. Again, you can use station names as arguments to only process specified stations, e.g.,

`./4_mergeNoiseData.py LT01 LT20`

Parameters in script `4_mergeNoiseData.py` are shown below, and explained in the following table.

```
7
8 From = '../OBSData_Resampled_Synced'
9 To   = '.'
10
11 OutputFileLength = 10 # days
12 OutputFileLength *= 24*60*60.0
13
```

<i>From</i>	Folder of resampled OBS data
<i>To</i>	Folder of OBS data after merging
<i>OutputFileLength</i>	Time length per file after merging

Input Files : OBS data (SAC files).

NOTE: script identifies SAC files, stations and channels according to file names.

Output Files : OBS data merged (fewer and longer SAC files).

## 5 Remove instrumental responses

Create a folder with name `NoiseData`. Run `5_removeInstrumentalResponse.py` to remove instrumental response. Running the script without arguments will process all the SAC files. Again, you can use station names as arguments to only process specified stations, e.g.,

`./5_removeInstrumentalResponse.py LT01 LT20`

Parameters in script [5\\_removeInstrumentalResponse.py](#) are shown below, and explained in the following table.

```

8 From = '../OBSData_Resampled_Synced_Merged'
9 To   = '.'
10
11 IR = '../instrument-responses'
12 PathIR = os.path.abspath(IR)
13
14 # para = 'none' # IDEP = DISPLACEMENT (NM)
15 # para = 'acc'  # IDEP = ACCELERATION (NM/SEC/SEC)
16 # para = 'vel'  # IDEP = VELOCITY (NM/SEC)
17 # para = 'none' # for pressure channel
18
19 para = 'acc'
20 freqlims = '0.001 0.002 10 20'

```

<i>From</i>	Folder of merged OBS data
<i>To</i>	Folder of OBS data after removing instrumental response
<i>IR</i>	Folder of instrumental response
<i>para</i>	Convert data to displacement, velocity or acceleration. <b>SAC command for 1,2,z channels:</b> transfer from evalresp fname * to <i>para</i> freqlimits <i>freqlims</i> OR transfer from polezero subtype * to <i>para</i> freqlimits <i>freqlims</i> <b>SAC command for pressure channels:</b> transfer from evalresp fname * to vel freqlimits <i>freqlims</i> OR transfer from polezero subtype * to none freqlimits <i>freqlims</i>
<i>freqlims</i>	Frequency limits for removing instrumental responses

Input Files : OBS data (SAC files).

NOTE: script identifies SAC files, stations and channels according to file names.

Output Files : OBS data with instrumental response removed (SAC files).

## 6 Rotate OBS data

There can be an extra step to rotate the data horizontally to the N-E direction. To do this, the angle between channel 1 and north must be given. This step is done using script [6\\_rotateNoiseData.py](#).