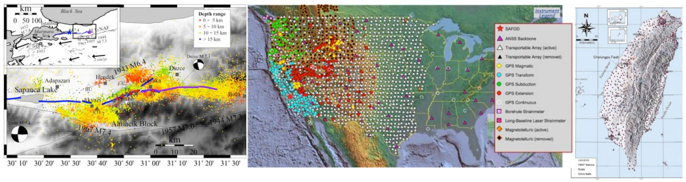
EAS 4803/8803 - Obs. Seismology
Lec#08: Data management and basic processing techniques

- Dr. Zhigang Peng, Spring 2011



Last Time

- Structure for the SEED instrument response
- Compute gain and phase from poles and zeros
- How to remove instrument response
- Homework 1 (how to play with FFT in Matlab)

SEED Manual: http://www.iris.edu/manuals/SEEDManual_V2.4.pdf

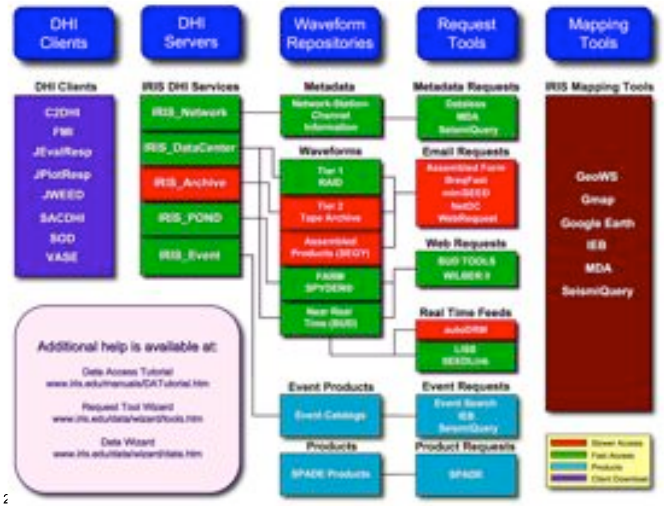
This Time

- Request data from the data center
- Data management and basic data processing tools
- Introduction to precision and accuracy
- Waveform stacking
- Array analysis

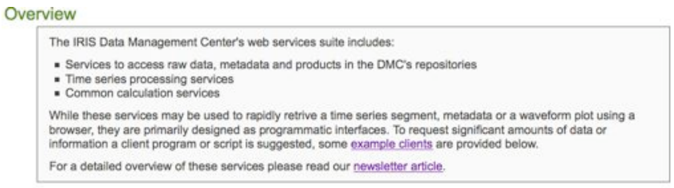
Downloading data from the data center

- Data center: http://www.iris.edu/, http://www.data.scec.org, http://www.ncedc.org, http://www.hinet.bosai.go.jp/, http://www.fnet.bosai.go.jp/freesia/index.html, http://www.kyoshin.bosai.go.jp/, http://www.kik.bosai.go.jp/kik/index_en.shtml
- Data access method: BREQ_FAST, NetDC, Wilber II, stp, EVT_FAST, etc.
- IRIS Data Management Center Data Access Tutorial: http://www.iris.edu/manuals/DATutorial.htm

---



Web Services at IRIS DMC http://www.iris.edu/ws/



Overview

The IRIS Data Management Center's web services suite includes:

- Services to access raw data, metadata and products in the DMC's repositories
- Time series processing services
- Common calculation services

While these services may be used to rapidly retrive a time series segment, metadata or a waveform plot using a browser, they are primarily designed as programmatic interfaces. To request significant amounts of data or information a client program or script is suggested, some example clients are provided below.

For a detailed overview of these services please read our newsletter article.

Details can be found at the following newsletter article:
http://www.iris.edu/news/newsletter/vol12no3/web_services.htm
Access of waveforms: http://www.iris.edu/ws/timeseries/
IRIS Timeseries Webservice URL Builder:
http://www.iris.edu/ws/timeseries/builder

Data Management

- Increasing volume of seismic data
  - New project (EarthScope) -> many stations
  - Cheap disk -> continuous recording
- Increasing demand for data mining before publishing scientific papers
  - Few people can publish a nice paper based on only 1 or a few seismograms
  - More data, better statistics, higher confidence

Management Tools

- Antelope/ Datascope
  - The Antelope Relational Database System
  - Nice way to organize seismic data, pick phases, locate events
  - Relatively hard to conduct scientific research (*IMHO*)

Yet another simple way

- Store your data in SAC format and organize it
- Seismic Analysis Tools (SAC)
  - http://www.iris.edu/manuals/sac/
  - http://geophysics.eas.gatech.edu/classes/SAC/

Data Organization Tips

- Follow the same rules
- Put most updated information in SAC header
- Keep the original data intact
- Use shell script to interact with the SAC data
- Plotting tools: SAC/Matlab/GMT
- Backup, backup, backup!

Rules for organizing the data

- Most seismic data are event based (one event recorded by many stations)
- Put all waveforms in the same directory
- The directory is named after the event origin time, or something that is unique and easy to use.
- For example, an event occurred on Wed Mar 7 15:17:27 EST 2007, we name the directory as 2007066151727, or 20070307151727.
- 066 is julian day (or the number of days since January 1st in that year)

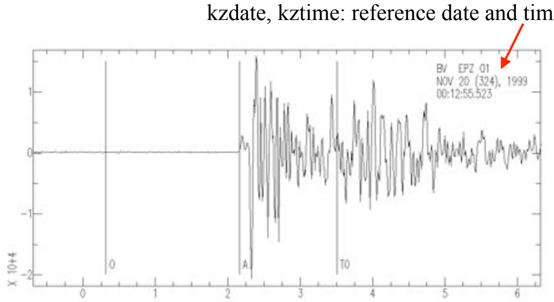
Rules for organizing the data

- Waveform name extracted from SEED volume:
- 2006.288.17.15.16.1474.PR.CDVI..BHN.R.SAC
- My waveform naming convention: NET.STN.COMP.SAC (e.g., BP.MMNB.DP1.SAC).
- You can come up with your own rules, but it must be unique and consistent.

Put most updated information in the SAC header

- Time: origin time (o), P and S arrival (either from existing catalog and phase picks, or auto/hand picker)
- Event location: evla, evlo, evdp, (mag, kevnm)
- Station location: stla, stlo, stel, (kstnm)
- Channel information: cmpaz, cmpinc (kcmpnm)
- Synchronize the time so that the origin time o start from time 0 s, or a common time.

Time in SAC header



kzdate, kztime: reference date and time

Putting time information into the SAC header

- Information in the catalogue
  - Event_ID Longitude Latitude Depth Mag Date Time
  - 1999324001255 31.0033 40.7993 8.15 2.31 11/20/1999 0:12:55.84
- Reference time in the SAC header
  - Wf year jday hh mm sec msec
  - BV.z 1999 324 0 12 55 523
- Your origin time is: 55.84 – 55.523 = 0.317 s
- Convert everything in epoch time (sec since 1970/01/01): /usr/local/geophysics/bin/epoch
- My own code: /usr/local/geophysics/bin/gsact
- Usage: gsact year month day hour min sec minsec f sac_files ...
- : Calculate the SAC origin and arrival time relative to kztime
- based on catalog and arrivals

Obtaining SAC header information

- To list the SAC header information, you can open the data in SAC, and use lh (listhdr) command.
- saclst
- Usage: saclst header_lists f file_lists
- saclst evla evlo stla stlo f BV.?

  # list the SAC header evla evlo stla stlo

• BV.e	40.7993	31.0033	40.7552	31.0149
• BV.n	40.7993	31.0033	40.7552	31.0149
• BV.z	40.7993	31.0033	40.7552	31.0149

Keep the original data intact

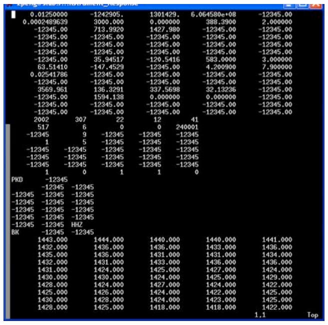
- Once you finish organizing the data, keep it in a safe place.
- Backup your data, or at least your scripts frequently.
- When you use a subset of data, or apply some procedure (resampling, filtering), do not overwrite the original data.

What else?

- It's time to 'mess around' with your organized data
- Always use shell script to automate the daunting task, and keep a record of your script (parameters)
- Keep in mind that you can do many things with the same data
- Sometimes other people may also use your organized data

Other useful SAC tools (mostly written by myself or other people)

- Convert SAC from binary to ASCII format
  - SAC> r BK.PKD.HHZ.SAC
  - SAC> w alpha temp.dat
- Using command:
  - sacdump BK.PKD.HHZ.SAC > tmp.dat
  - sacdump_slice 700 1200 BK.PKD.HHZ.SAC > tmp1.dat



```
      0.01250000      -1242905.       1301429.   6.064580e+08      -12345.00
    0.0002489639       3000.000       0.000000       388.3900       2.000000
       -12345.00       713.9920       1427.980      -12345.00      -12345.00
       -12345.00      -12345.00      -12345.00      -12345.00      -12345.00
       -12345.00      -12345.00      -12345.00      -12345.00      -12345.00
       -12345.00      -12345.00      -12345.00      -12345.00      -12345.00
       -12345.00       35.94517      -120.5416       583.0000       3.000000
        63.51410      -147.4529      -12345.00       4.200900       7.900000
      0.02541786      -12345.00      -12345.00      -12345.00      -12345.00
       -12345.00      -12345.00      -12345.00      -12345.00      -12345.00
        3569.961       136.3291       337.5698       32.13236      -12345.00
       -12345.00       1594.138       0.000000       0.000000      -12345.00
       -12345.00      -12345.00      -12345.00      -12345.00      -12345.00
       -12345.00      -12345.00      -12345.00      -12345.00      -12345.00
       2002         307          22          12          41
        517           6           0           0      240001
     -12345           9      -12345      -12345      -12345
          1           5      -12345      -12345      -12345
     -12345      -12345      -12345      -12345      -12345
     -12345      -12345      -12345      -12345      -12345
     -12345      -12345      -12345      -12345      -12345
          1           0           1           1           0
PKD          -12345
        -12345  -12345
-12345  -12345  -12345
-12345  -12345  -12345
-12345  -12345  -12345
-12345  -12345  -12345
-12345  -12345  HHZ
BK      -12345  -12345
        1443.000       1444.000       1440.000       1440.000       1441.000
        1432.000       1436.000       1436.000       1433.000       1436.000
        1435.000       1436.000       1431.000       1431.000       1434.000
        1432.000       1431.000       1433.000       1436.000       1436.000
        1431.000       1424.000       1425.000       1427.000       1424.000
        1430.000       1430.000       1424.000       1425.000       1429.000
        1428.000       1424.000       1427.000       1424.000       1422.000
        1425.000       1426.000       1424.000       1422.000       1425.000
        1430.000       1428.000       1424.000       1423.000       1425.000
        1428.000       1425.000       1418.000       1418.000       1422.000
                                                                 1,1          Top
```

Convert ASCII data into SAC file

- Command:
- > col2sac
- Usage: stdin | col2sac sac_file_name delta t0
- Example:
- gawk '{print $2}' tmp1.dat |\
- col2sac tmp1.sac 0.0125 700

Introduction to Stacking

- Seismology use seismic data to estimate quantities related to the Earth structure and seismic source.
- Ideally these estimates are both *accurate* and *precise*.
  - *Accuracy* measures the deviation of the estimate from its true value.
  - *Precision* measures the repeatability of individual estimates.

Chap. 6.5 of the Stein book

Accuracy vs. Precision

- *Accuracy* depends on systematic errors that bias groups of estimates.
- *Precision* depends on random errors that affect individual estimates.
- Estimates can be precise but inaccurate, or accurate but imprecise.
- Can you think of any example in seismology?

High accuracy, but low precision

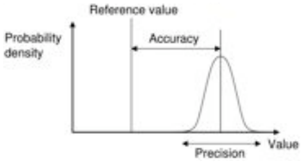
High precision, but low accuracy

Example

- An estimate of an earthquake's location depends on the quality of the travel time data and the accuracy of the velocity model.
- High-quality travel time data with an incorrect velocity model, can yield location that is precise (small uncertainty), but inaccurate in that the resulting location is not where earthquake occurred.
- Conversely, an accurate velocity model and poor travel time data give "relatively" accurate and imprecise location.

Improving accuracy and precision

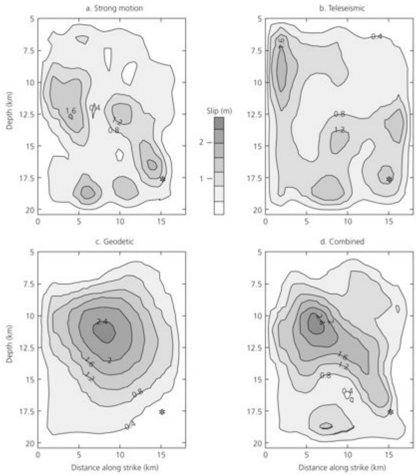
- Accuracy can be improved by using different measuring tools, ideally calibrated against each other.
- Precision can be improved by making multiple measurements, ideally by different people.



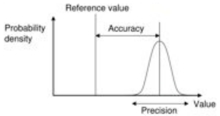
Complications

- For example, an earthquake is (in most cases) a non-repeatable experiment, so we cannot make additional measurements.
- Estimating depth from travel times and waveform modeling are only partially independent – both can be biased similarly by incorrect assumptions about near source mechanisms.
- A further complication is that different methods can measure related but not identical entities. For example, finite source modeling from near-field strong-motion recordings, teleseismic waveforms, and geodetic measurements often differ with each other.
- Can you provide other examples?

Figure 4.5-10: Slip inversions for the 1994 Northridge earthquake using different data sets.



Systematic error



- Most discussions focus on random errors because they are easy to estimate from the scatter of measurements.
- It appears that assessments of the formal or random uncertainty often significantly underestimate the systematic error, so the overall uncertainty is dominated by the unrecognized systematic error and thus larger than expected.
- Measurements of a quantity often remains stable for a while, then suddenly change by much more than the previously assumed uncertainty.
- Systematic biases are difficult to detect, but sometimes are identified from discrepancy between different approaches.

Random Error

- We estimate a quantity x from multiple measurements, x_i (due to noise and limitations of the measurements).
- With enough measurements, a pattern generally emerges in which the values xi are distributed around x'.
- If we neglect system errors of measurements, we can estimate x from the measured value x_i, and associated uncertainties.

Gaussian Distribution

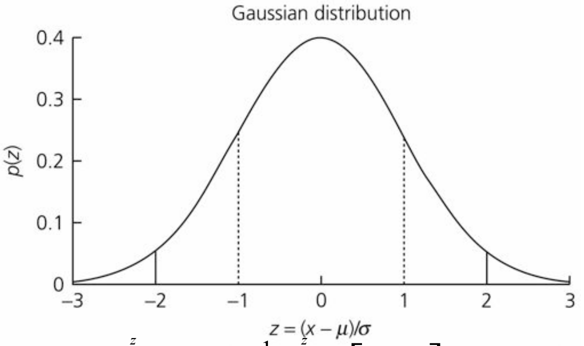
$$p(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right]$$

Two variable: the mean μ, and the standard deviation σ.

$$z = (x - \mu)/\sigma$$

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-z^2/2\right]$$

Figure 6.5-1: Gaussian distribution.



$$A(z) = \int_{-z}^{z} p(y)dy = \frac{1}{\sqrt{2\pi}} \int_{-z}^{z} \exp\left[-y^2/2\right]dy$$

For $z = 1$, $A(z) = 0.68$; $z = 2$, $A(z) = 0.95$; $z = 3$, $A(z) = 0.997$.

Reducing errors by stacking

- One of the most useful methods for improving measurements from seismological data: stacking
- Stacking: taking multiple measurements and averaging them.
  - By averaging measurements such as travel times from different seismograms.
  - By adding many seismograms and then estimating parameters.
- Stacking will have two effects:
  - It improves precision by reducing the effects of random noise in the data.
  - If the data are averaged in special ways, the precision, and perhaps accuracy, can be improved by suppressing some features in the data while enhancing other desired features.

This Time

- Request data from the data center
- Data management and basic data processing tools
- Introduction to precision and accuracy
- Waveform stacking
- Array analysis

Next Time

- Data management and basic data processing tools
- Waveform stacking
- Array analysis