Tools and topics in seismic waveform cross-correlation

Michael West
GI Seismology lab
March 27, 2008
Data load $\approx 800$ hrs / person / day

Volcanoes, earthquakes, landslides, glaciers, cultural, electronic noise, etc.

Pattern recognition
Correlation toolbox and techniques
Applications
1) volcanic swarms (Augustine 2006)
2) differential traveltimes (Nikolski swarm)
3) coda interferometry (Bezymianny)
Matlab object with 7 fields
Object approach motivated by C. Reyes
Rewritten spring 2006 to include \textit{WAVEFORM}

\begin{verbatim}
>> c = correlation(...)
WAVEFORMS: 20x1 vector
TRIG: 20x1 vector
CORR: 20x20 square matrix
LAG: 20x20 square matrix
STAT: 20x1 matrix
LINK: 19x1 matrix
CLUST: 20x1 vector
\end{verbatim}
Based on waveform
Antelope, Winston, SAC files

*Any waveform object*

```plaintext
>> correlation('AUE','EHZ',trig,-5,10)

... where `trig` is a vector of "trigger" times

trig = [ 12-Jan-2006 06:23:31
         12-Jan-2006 06:24:36
         12-Jan-2006 06:25:59
         ... ];
```
All correlation objections have *waveforms* and *triggers*.

```matlab
>> c = correlation(...)
WAVEFORMS: 20x1 vector
TRIG: 20x1 vector
CORR: 0x0 square matrix
LAG: 0x0 square matrix
STAT: 0x0 matrix
LINK: 0x0 matrix
CLUST: 0x0 vector
```
Cross-correlation
Frequency domain multiplication
Optimized for large datasets
Not built-in Matlab routines
Normalized to $[-1 \ 1]$
Correlation ($\approx 0.9$)

Lag time (time correction)

$$\text{c} = \text{correlation}(...)$$

WAVEFORMS: 20x1 vector
TRIG: 20x1 vector
CORR: 20x20 square matrix
LAG: 20x20 square matrix
STAT: 0x0 matrix
LINK: 0x0 matrix
CLUST: 0x0 vector
All waves correlated against all waves

\[ \text{plot}(\text{c, 'corr'}) \]
Align traces based on optimum correlation

```matlab
>> c = adjusttrig(c)
>> plot(c,'shaded')
```

Demo dataset: 100 traces
So called “hierarchical cluster tree”
Events clustered by correlation value

```python
>> c = linkage(c)
```
How to link waveforms?
Numerous methods
Two most useful are average and single linkage
Country club analogy

“average” linkage

“single” linkage

\[
C_{p,q} = \frac{1}{n_p n_q} \sum_{i=1}^{n_p} \sum_{j=1}^{n_q} C_{pi,qj}
\]
Linkage info is stored in LINK field

```matlab
>> c = correlation(...)
WAVEFORMS: 20x1 vector
TRIG: 20x1 vector
CORR: 20x20 square matrix
LAG: 20x20 square matrix
STAT: 0x0 matrix
LINK: 19x3 matrix
CLUST: 0x0 vector
```
Easy to call because info is in the object

```python
>> c = linkage(c)
>> plot(c,'dendrogram')
```
Cut the “branches” to form clusters
Each cluster gets unique id number

```python
>> c = cluster(c, 0.8)
```
Cluster info is stored in LINK field

```matlab
>> c = correlation(...)

WAVEFORMS: 20x1 vector
TRIG: 20x1 vector
CORR: 20x20 square matrix
LAG: 20x20 square matrix
STAT: 0x0 matrix
LINK: 19x3 matrix
CLUST: 20x1 vector
```
### Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>butter</td>
<td>butterworth filter</td>
</tr>
<tr>
<td>cat</td>
<td>concatenate correlation objects together</td>
</tr>
<tr>
<td>cluster</td>
<td>create clusters of events</td>
</tr>
<tr>
<td>correlation</td>
<td>create a correlation object (constructor)</td>
</tr>
<tr>
<td>crop</td>
<td>adjust start and end times</td>
</tr>
<tr>
<td>display</td>
<td>show correlation object fields</td>
</tr>
<tr>
<td>find</td>
<td>extract individual clusters</td>
</tr>
<tr>
<td>get</td>
<td>get fields from correlation object</td>
</tr>
<tr>
<td>getstat</td>
<td>Vandecar and Crosson (BSSA 1990) statistics</td>
</tr>
<tr>
<td>interferogram</td>
<td>create interferogram of correlation</td>
</tr>
<tr>
<td>linkage</td>
<td>create hierarchical cluster tree</td>
</tr>
<tr>
<td>minus</td>
<td>subtract one trace from the others</td>
</tr>
<tr>
<td>norm</td>
<td>normalize trace amplitudes</td>
</tr>
<tr>
<td>set</td>
<td>set object fields</td>
</tr>
<tr>
<td>sort</td>
<td>sort traces based on various parameters</td>
</tr>
<tr>
<td>stack</td>
<td>stack traces</td>
</tr>
<tr>
<td>subset</td>
<td>create a new object from a subset of traces</td>
</tr>
<tr>
<td>xcorr</td>
<td>cross correlate traces against one another</td>
</tr>
<tr>
<td>check</td>
<td>check consistency of object fields</td>
</tr>
<tr>
<td>verify</td>
<td>verify the consistency of the object</td>
</tr>
</tbody>
</table>

... plus WAVEFORM methods

---

*Not very many!*
plot(c,'style', ...)

'wig'  wiggle traces
'sha'  shaded traces
'ove'  overlain traces
'int'  interferogram
'cor'  correla. matrix
'lag'  lag matrix
'den'  dendrogram
'eve'  event rate plot

plotting

Maximum correlation coefficient

Lag time for maximum correlation

Inter-cluster correlation
Excellent for data mining swarms

Total number of waveforms = 3,514
get times from arrival database ...

```matlab
>> c = correlation('AU13','HHZ',trig,-1,6,'/home/admin/.../AU06_01')
>> c = butter(c,[0.5 20]);
>> c = xcorr(c);
>> plot(c,'corr');
```
Keep only repeating events

```matlab
>> c = linkage(c)
>> c = cluster(c, 0.8)
>> N = find(c, 'clust', 1:40)
>> c = subset(c, N)
>> plot(c, 'cor')
```
2006 eruption of Augustine volcano

Buurman and West, 2009?

Total number of waveforms = 376

DATE (AKD), 2006

Swarms

Precursory

Explosive phase

Continuous phase

0

0.1

0.2

0.3

0.4

0.5

0.6

0.7

0.8

0.9

1

CORRELATION VALUE

Jan. 10

Jan. 11

Jan. 12

Jan. 13

Jan. 14

Jan. 15

Jan. 16

Jan. 17

Jan. 18

Jan. 19

Jan. 20

Jan. 21

Jan. 22

Jan. 23

Jan. 24

Jan. 25

Jan. 26

Jan. 27

Jan. 28

Jan. 29

Jan. 30

Jan. 31

Feb. 01

Feb. 02

Feb. 03

Feb. 04

Feb. 05

Feb. 06

Feb. 07

Feb. 08

Feb. 09

Feb. 10

Feb. 11

Feb. 12

Feb. 13

Feb. 14

Feb. 15

Feb. 16

Feb. 17

Feb. 18

Feb. 19

Feb. 20

Feb. 21

Feb. 22

Feb. 23

Feb. 24

Feb. 25

Feb. 26

Feb. 27

Feb. 28
Get pick times from monthly catalog

```matlab
>> db = '/home/admin/databases/AEIC_CATALOG/2008_01/aeic_2008_01';
>> db = dbopen(dbname,'r');
>> db = dblookup(db,'','origin','','');
>> db = dbsubset(db,'deg2km(distance(lat,lon,52.2,-168.1))<35');
>> db1 = dblookup(db,'','event','','');
>> db = dbjoin(db,db1);
>> db = dbsubset(db,'oriid==prefor');
>> db = dbsort(db,'time');
>> db1 = dblookup(db,'','assoc','','');
>> db = dbjoin(db,db1);
>> db1 = dblookup(db,'','arrival','','');
>> db = dbjoin(db,db1);
>> db = dbsubset(db,'phase=="P"');
>> db = dbsubset(db,'chan=="BHZ"');
>> db = dbsubset(db,'sta=="NIKO"');
>> nrecords = dbquery(db,'dbRECORD_COUNT')
>> [epoch,arid,oriid] = dbgetv(db,'arrival.time','arid','oriid');
>> dbclose(db);
>> % convert epoch to matlab date
>> time = epoch2str(epoch,'%m %d %Y %H %M %S.%s');
>> time = datenum(time,'mm dd yyyy HH MM SS.FFF');
```
Get waveforms

```matlab
>> !cat /iwrn/op/db/archive/archive_2008/archive_2008_01_??.wfdisc >
   archive_2008_01.wfdisc
>> !ln -s /iwrn/op/db/archive/2008 2008
>> c = correlation('NIKO','BHZ',time,-3,5,'archive_2008_01');
>> c = butter(c,[1 10]);
>> plot(c)
```
Cross correlate pick window

```matlab
>> c = xcorr(c,[-.5 1]);
>> plot(c,'corr')
```
Extract lag time of well-correlated events

```matlab
corr = get(c,'CORR');
lag = get(c,'LAG');
corr = triu(corr) - diag(diag(corr));
[col,row] = ind2sub(size(corr),find(corr>0.8));
disp(' arid1 arid2 lag corr');
for n = 1:numel(col)
    txt = sprintf('%8.0f %8.0f %8.3f %8.3f', arid(col(n)), ...
                   arid(row(n)), lag(col(n),row(n)), corr(col(n),row(n)));
    disp(txt);
end
```

<table>
<thead>
<tr>
<th>arid1</th>
<th>arid2</th>
<th>lag</th>
<th>corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1834</td>
<td>1851</td>
<td>0.278</td>
<td>0.807</td>
</tr>
<tr>
<td>1851</td>
<td>2291</td>
<td>-0.287</td>
<td>0.940</td>
</tr>
<tr>
<td>1834</td>
<td>2434</td>
<td>0.060</td>
<td>0.976</td>
</tr>
<tr>
<td>1851</td>
<td>2434</td>
<td>-0.198</td>
<td>0.816</td>
</tr>
<tr>
<td>1834</td>
<td>4730</td>
<td>0.240</td>
<td>0.803</td>
</tr>
<tr>
<td>3430</td>
<td>6506</td>
<td>-0.000</td>
<td>0.815</td>
</tr>
<tr>
<td>13392</td>
<td>25084</td>
<td>-0.160</td>
<td>0.834</td>
</tr>
<tr>
<td>22891</td>
<td>25107</td>
<td>-0.055</td>
<td>0.891</td>
</tr>
<tr>
<td>13308</td>
<td>26002</td>
<td>-0.066</td>
<td>0.827</td>
</tr>
<tr>
<td>25084</td>
<td>26002</td>
<td>-0.046</td>
<td>0.811</td>
</tr>
</tbody>
</table>
How does a waveform evolve over time?

Clusters prior to Bezymianny eruption

Lifespan of 19 clusters (multiplets)

Cluster Number

December 2006

9 11 13 15 17 19 21 23 25 27 29 31

High Rockfall / Tremor signals

Courtesy of Wes Thelen, UW
469 nearly identical waveforms . . . . or are they?
Stack adjacent sets of 10 traces (noise reduction)

```matlab
>> for n = 1:10:469
    c = stack(c, n:n+9);
>> end
>> c = subset(c, 470:516);
```
Make interferogram

```matlab
>> c = interferogram(c)
>> plot(c,'inter');
```
http://giseis.alaska.edu/Seis/EQ/tools/matlab

cookbook
On website and command line
>> correlation('cookbook')

Demo dataset
100 Augustine traces. Used in cookbook.
>> c = correlation('demo')

Install and release notes - README.txt

#########################################################################
# Release notes v1.5  #
#########################################################################
March 2008
revised LINKAGE, CLUSTER and dendrogram plotting routines to minimize references to the obscure concept of dissimilarity. Since clusters in the correlation toolbox are almost always based on the idea of correlation, these functions were revised to operate directly on correlation values. This is a hack because the underlying Matlab routines deal with dissimilarity. The LINK field still stores values as dissimilarity. However the two main (only?) uses of the linkage field (to create clusters and to plot dendrogram) now work natively with correlation values.
....