Segy Clusterer …. a rapid volume scanner

The Segy Clusterer scans a 3D Segy volume for values that are similar in touching Inline-Xline-Z locations. For projects worked at Quantum Earth, this is a normal first step. It quickly gives a set of starting horizons for further work that are guaranteed to have passed a statistical test.

It is an agglomerative or aggregative cluster constructor. If 2 Inline-Xline-Z locations both pass the amplitude cut, they are placed in the same cluster if they are adjacent. If they are not adjacent, they go into different clusters. The 2 different clusters may ultimately merge together, if there is a path of touching values out-of-the-plane, for instance.

There is no new science here. It is based on standard multi-variate statistics. For example, a good reference is: Multivariate data analysis, 6th Ed., Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., Pearson-Prentice Hall, 2006, ISBN 0-13-032929-0, see chapter 8 called "Cluster Analysis". Since there is no new science here, Quantum Earth's only contribution is in generating a program that works with Segy data and outputs clusters that can be loaded to workstations for further work.

Output … Horizons, summary files and clusters, of course

The Segy Clusterer takes Segy as input and outputs a whole several file types:

- **Horizon files** containing clusters ready to load to a workstation using canned formats
- **Image files** containing pictures of mapped clusters
- **Summary statistics**, as Csv files
- **Cluster Files** as text files which contain sets of clusters

host of text files, mostly as Csv files. At Quantum Earth, a will output a Segy File with the same header byte locations used on input. It will have the same size as the input file and will contain zeros (0.0) at locations which are not outliers. Each outlier value is the distance from the background cluster, with positive numbers in quadrants 1 and 4 (positive X-Axis) and negative values in quadrants 2 and 3 (negative X-axis). The distances are scaled to be in units of standard deviation.

Horizon file output

One of the most useful things to do with clusters is to write them back to a workstation as horizons. These cluster based horizons then form the seed points for further work. At Quantum Earth it is normal to include the number of acres in the horizon name so that it is easy to see which are big when scanning down a list of horizon names in the workstation.

Before being written out as a horizon, each cluster is smashed down to its medial plane. The output amplitude and Z value at each medial Inline-Xline location is the weighted mean (normally) of all the cluster values in the Inline-Xline bin that are part of that cluster. The clusters are then written to text files in a canned format like "Geoframe Card Image 7," which seems to work well for both SMT and Geoframe.

Cluster editor

Since many clusters may be produced, the Segy Clusterer includes an editor that can help select which clusters are worth further work. You can sort and select based on acreage, mean amplitude, Z-range, etc.

After sorting, deleting or selecting, a new cluster file can be written that only contains the ones you are interested in pursuing.

Inside the editor, individual clusters or groups can also be mapped for quick QC and quick seismic lines can be shown through the centroid of clusters, again as a QC aid.

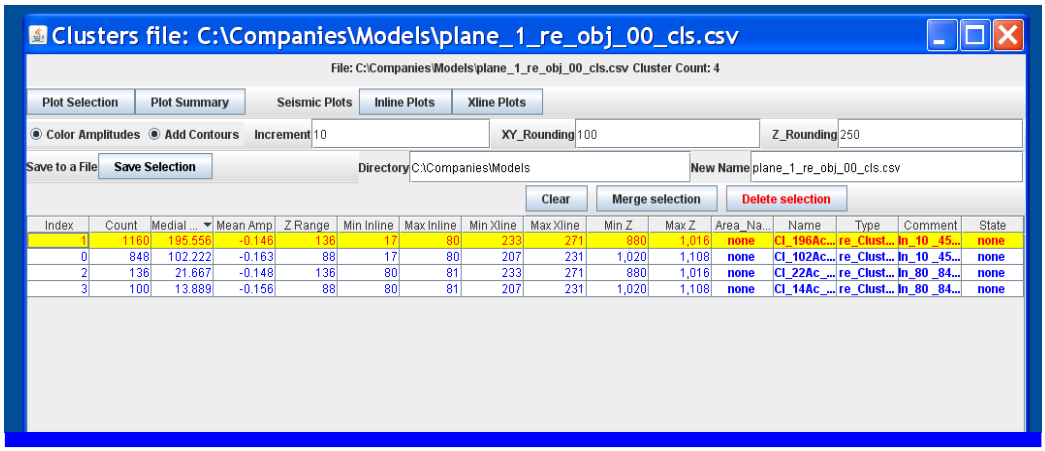

Figure 1: A screen shot from the cluster editor showing 3 clusters ready for plotting, saving or whatever.

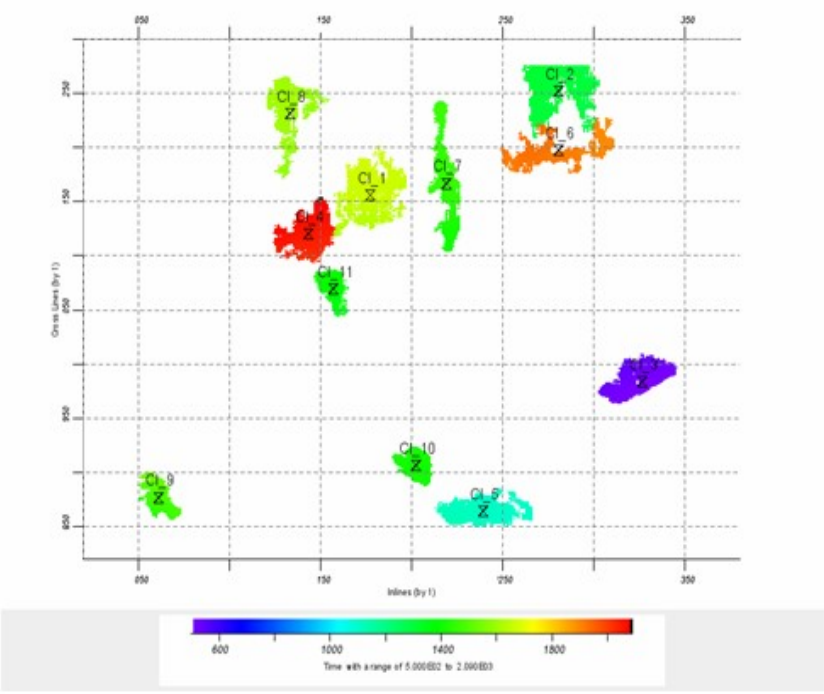

Figure 2: This is a summary map made using the cluster editor for 11 clusters that were extracted from 500 – 2000 MSec. The area is about 1 OCS block, colors are time in Msec, with purple being shallow. The indices like "Cl_9" are tied to the spreadsheet shown in Figure 2.

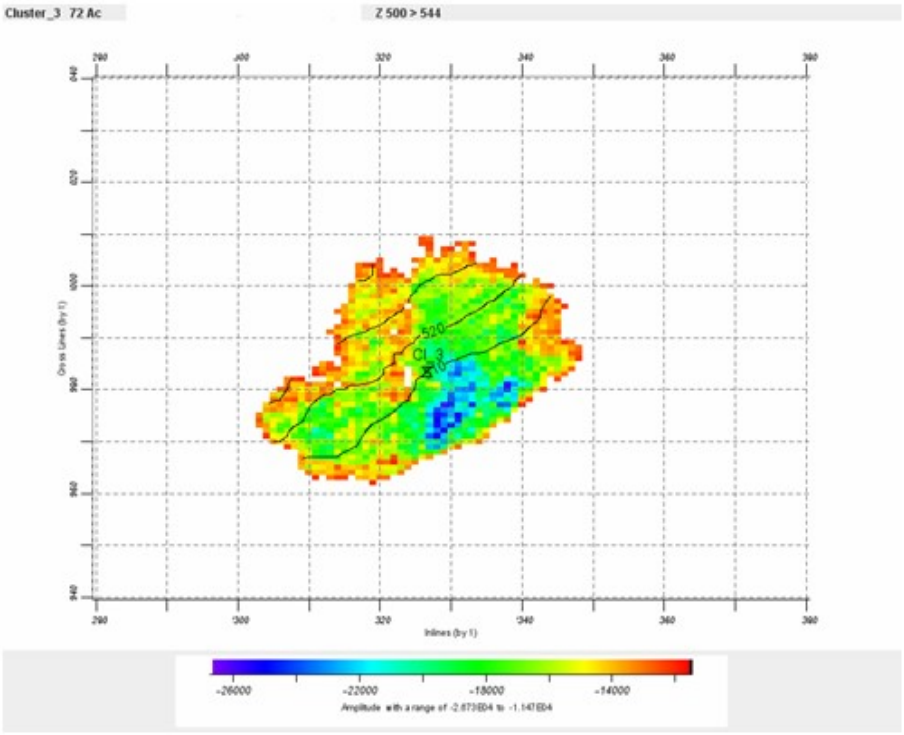

Figure 3: This map was made using the cluster editor for a single Cluster. The colors are negative amplitudes cut at -4 StDev. Contours are 10 Msec. This is a fairly standard QC plot that is made before deciding if a cluster should be loaded to a workstation.

Index	Inline	Xline	Time or Z	Area: Acres	Mean Amp	Min Inline	Max Inline	Min Xline	Max Xline	Min Z	Max Z	Point Count
1	177	156	1656	101	-15890	156	197	118	201	1624	1672	1774
2	281	251	1299	92	-21850	261	303	208	275	1284	1312	1621
3	326	984	515	72	-16340	303	348	-38	9	500	544	1268
4	143	120	2052	66	-14220	124	158	93	154	2006	2084	1164
5	239	864	1110	65	-18590	213	266	850	884	1096	1122	1147
6	281	197	1913	65	-14350	249	311	176	227	1884	1930	1140
7	219	167	1439	63	-21020	209	227	104	242	1392	1462	1104
8	133	231	1599	61	-16540	121	155	172	265	1584	1608	1079
9	61	877	1484	31	-23090	50	73	856	899	1428	1540	540
10	202	906	1404	28	-24520	190	211	887	923	1398	1414	499
11	157	69	1380	27	-22820	147	165	43	87	1340	1404	480

Figure 4: This is a summary spreadsheet sorted by decreasing acres. This is used to help QC loaded horizons or manage a set of clusters.

Dynamic statistics for input specification

Before starting a run, the Segy Clusterer has a facility generating statistics on the fly. This is done by randomly sampling the Inline, Xline and Z range that is planned to be scanned and then computing statistics for the random samples. Statistics are computed for each Z value (time step), so that data cuts can very with depth during the run. By doing statistics on the fly, you do not need to know the specific data values in your

dataset in order to cluster it. For example, if you want to capture all negative values that are 3 standard deviations from the mean, this is easily done.

Segy Header Scan required

As of 4/2009, the program requires that the input dataset be Segy Header Scanned. The .lns file produced by the header scan is used both for sampling for dynamic statistics and for the usage of multiple processors. If this requirement gets to be a show stopper, please call or contact Quantum Earth. This restriction has not been relaxed at this time because there has been no request to do so, and it would require some re-thinking.

Expected run times should be minutes to hours

As of 4/2009, the program is executing on a line-group basis using multiple processors. For a 1 Gig dataset, run time should be on the order of minutes (not hours), depending on the processor speeds and how severe the data cut is. This process runs longer if there are more clusters. A harder data cut that makes fewer clusters will run much faster. The program is concurrent, so that if you have more processors, it will run faster. Basically, it should run fast enough so that for a region of interest, you can get a set of clusters in 15-20 minutes.

Java required

This program is written in Java (Sun Trademark now part of Oracle). It needs version 1.6.0.1 or later to run. Please download the JRE from Sun's website if you need to. As a Java program, it has been run for several years on Windows, Linux and Solaris 10.

Last modified 4/22/2009