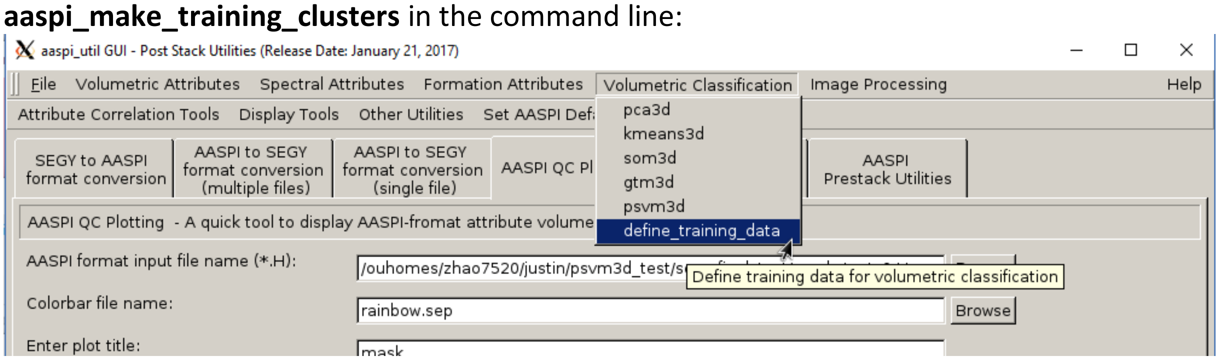
GENERATING TRAINING DATASET FOR SUPERVISED SEISMIC FACIES CLASSIFICATION –PROGRAM make_training_clusters

Contents

Overview

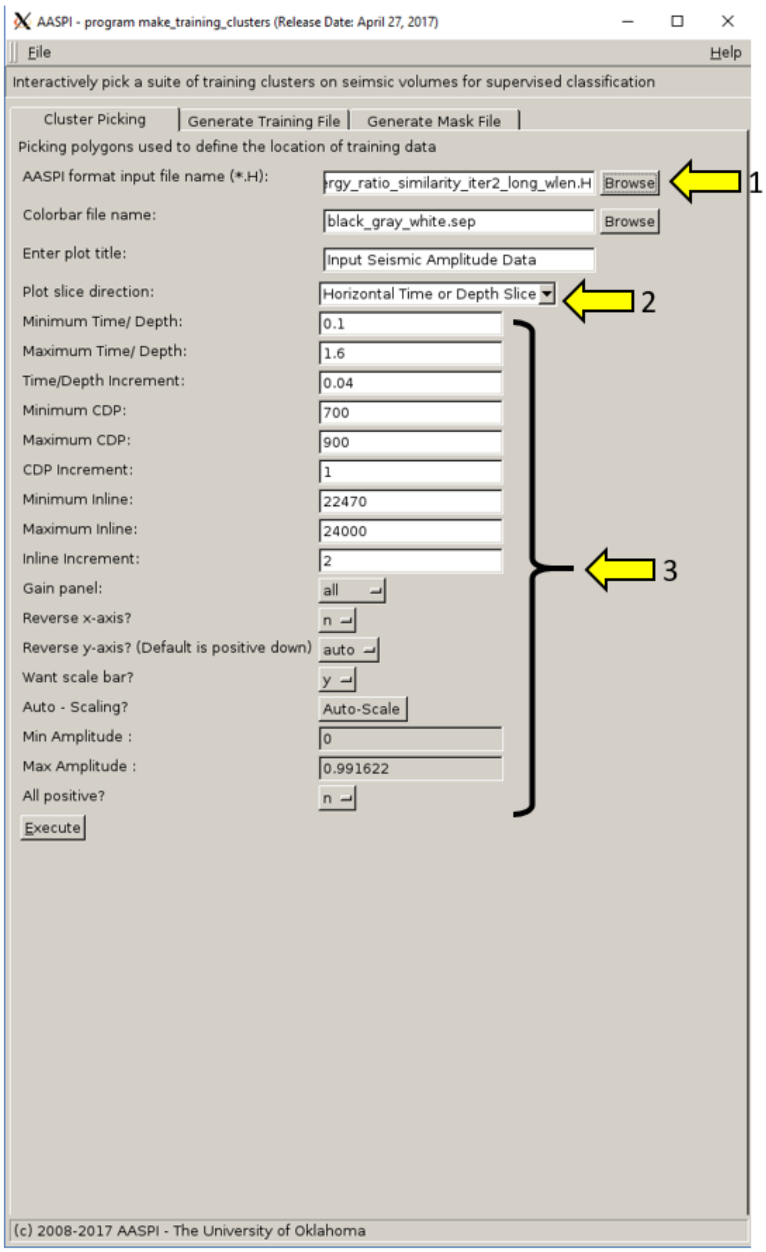
With the great improvement in computation power, 3D multiattribute seismic facies analysis and classification have been routinely performed in interpretation workstation software. With a limit amount of training data, the computer can mimic the behavior of a human expert by providing interpretation, consistent with a given model. In the AASPI package, there are several seismic facies analysis utilities, many of which perform supervised classification that need training (supervision) data. Namely, **som3d** and **gtm3d** are unsupervised seismic facies analysis unities, with the capability of adding posteriori supervision, and **psvm3d** is a supervised seismic facies analysis utility which must be used with training (supervision) data. Program **make_training_clusters** is a utility to build; 1). A training data file by manually picking the location of training data in a 3D seismic volume and labeling with facies index, then extracting seismic attributes at such picked locations to generate a training file that can be further used in the previously mentioned programs; and 2). A mask file from manually picked locations to define a subset of input attributes for building the model in unsupervised seismic facies analysis.

The program can be invoked from **aaspi_util** under *Volumetric Classification* menu or by typing **aaspi_make_training_clusters** in the command line:

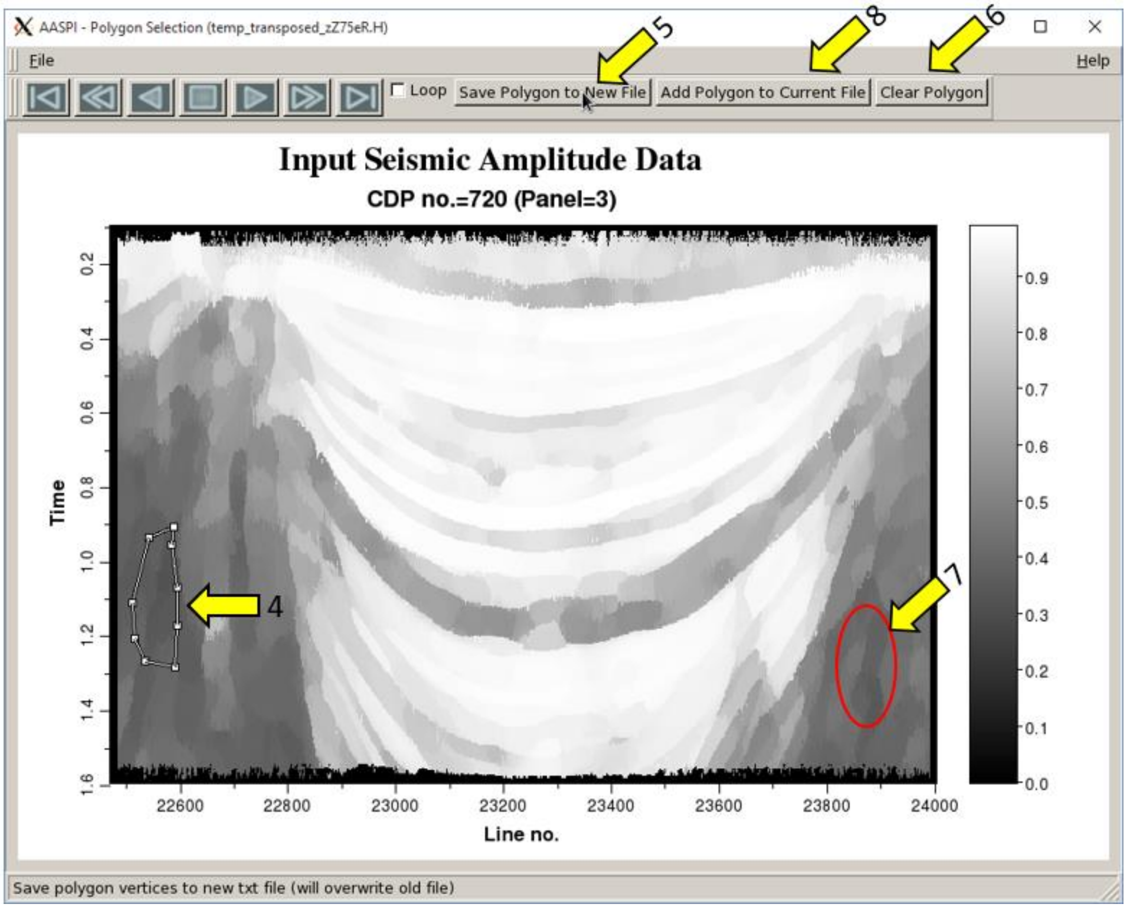


Defining the location of training data

After invoking program **make_training_clusters**, there are three tabs: *Cluster Picking*, *Generate Training File*, and *Generate Mask File*. In the *Cluster Picking* tab, the user will manually pick some polygons on a seismic volume, and such picks will be saved into a user defined number of *Polygon* files, with each *Polygon* file representing one facies (class). The picking process can be performed on time slices, or inline/crossline sections. In this document we illustrate by picking on crossline sections.



Firstly, **(1)** a seismic (attribute) volume is selected, **(2)** then select which slice to pick on, and **(3)** define the plotting parameters similar to the **QCplot** utility. After selecting *Execute*, a new window will be displayed:

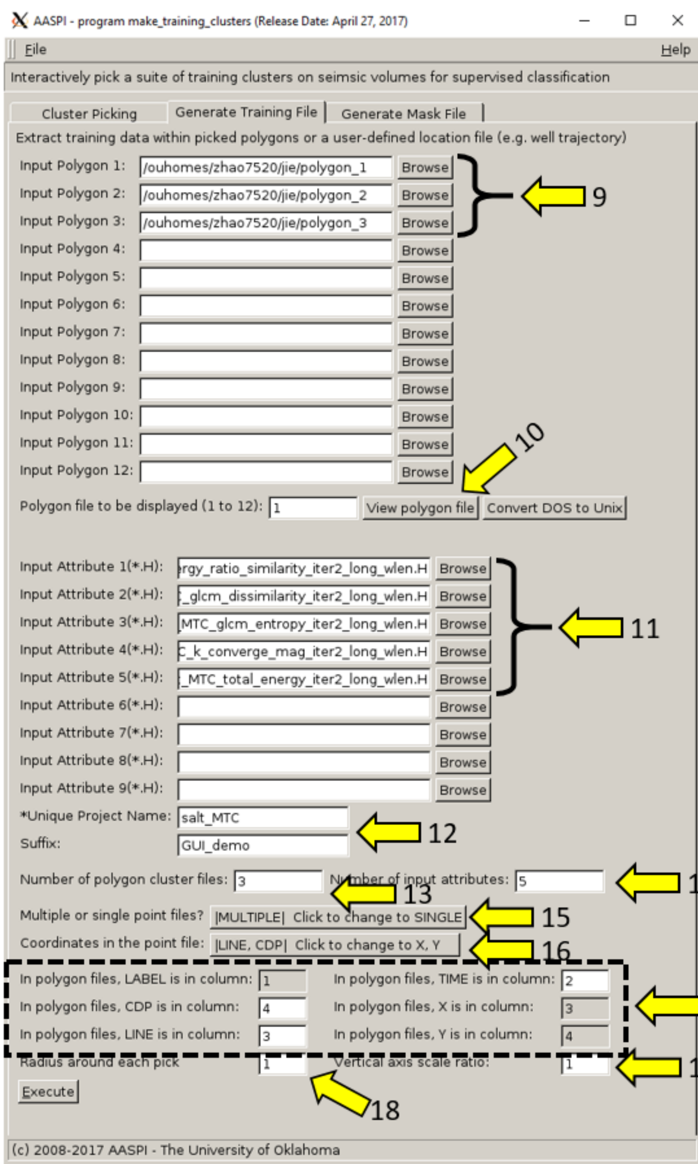


In this specific demonstration the goal is to classify among salt, mass transport complex (MTC), and sediments. The attribute shown above is Kuwahara filtered similarity − assume the user firstly wants to pick some training data to represent salt. The user draws a polygon (**4**) and define all samples within this polygon to be salt, then hits (**5**) to save the time, crossline, inline coordinates to a file named as *polygon*. If the user wants to pick more training samples for salt, one can hit (**6**) to clean the previous pick, and pick at another location that best represent salt (**7**), then hit (**8**) to save it to the previous *polygon* file. If more picks are needed, the user can repeat steps (**6**) – (**8**) on the same crossline, or navigate to other crosslines. Once finished picking for salt, the file *polygon* needs to be renamed (as *polygon_1* or *polygon_salt*) to avoid overwriting the new facies picks on the previous picked facies.

Repeat the same process to pick MTC and sediments, and now the user has 3 files (namely polygon_1, polygon_2, and polygon_3) representing 3 facies. With these three files, now the user is able to extract seismic attributes at such picked locations for each facies.

Extracting seismic attributes and generating the training file

In the second tab, *Generate Training File*, the user will take the previously generated *polygon* files as inputs, and extract seismic attributes at locations defined in the *polygon* files. Please note, the seismic attributes should be in the exact same order as they would appear in the input list of the classification program (**som3d**, **gtm3d**, or **psvm3d**) to be used.



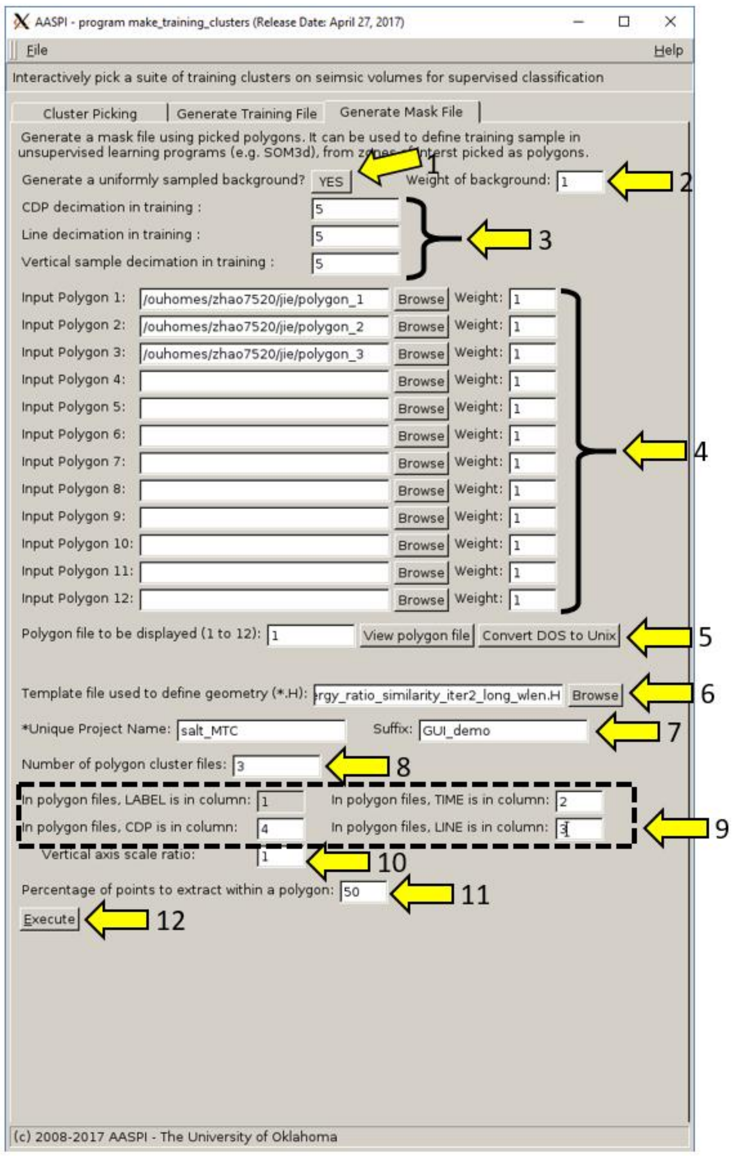
To generate a training file, first the user put the previously generated *polygon* files into the polygon list (**9**), and the user can view the file content of a *polygon* file using (**10**). A *polygon* file consists of 4 columns, and if the user selects *MULTIPLE* (**15**), three of the columns are the 3D coordinates, and the other column is not used. If *SINGLE* is selected in (**15**), the user needs to define the column for labels in (**17**). Toggle (**16**) is used to change the coordinate system in the polygon files between inline/crossline and X/Y. The order of the three coordinates may change depending on the slice used in the previous picking procedure (e.g. time, crossline, or inline), therefore the user needs to define these three dimensions in (**17**). After filling in the seismic attribute to be extracted (**11**), and defining (**13**) the number of facies (polygon files) and (**14**) attributes, the user also has an option to choose the radius of attribute extraction (**18**). A radius of 0 means no averaging will be performed, and a radius of 2 means the extracted attribute is averaged using a 5 by 5 by 5 window centered at the picked location. If the distance from a picked location to the data volume boundary is less than the defined radius, such picks will be ignored and stored in file **dropped_points.txt** for inspection. Finally, (**19**) *Vertical axis scale ratio* is used to change the scale ratio in case the vertical scale of the seismic data and polygon files do not match. After selecting *Execute*, an ASCII text format training file named **polygon_cluster_out_${proj name}_${suffix}.txt** will be generated, as shown below:

0.431	7.309	3.523	0.006	75728.698	1.000
0.424	7.322	3.523	0.006	75308.586	1.000
0.414	7.339	3.524	0.006	74392.820	1.000
0.412	7.383	3.525	0.006	73875.500	1.000
0.404	7.419	3.526	0.006	73608.906	1.000
0.398	7.465	3.526	0.006	73924.820	1.000
0.393	7.479	3.526	0.006	73352.062	1.000
0.393	7.497	3.526	0.006	71537.875	1.000
0.392	7.504	3.526	0.006	69722.719	1.000
0.390	7.515	3.527	0.006	68450.719	1.000
0.388	7.533	3.527	0.006	67807.102	1.000
0.386	7.559	3.527	0.006	67028.219	1.000
0.385	7.591	3.527	0.006	64761.949	1.000
0.384	7.610	3.528	0.006	61623.984	1.000
0.383	7.624	3.529	0.006	57872.953	1.000
0.383	7.626	3.531	0.006	55744.004	1.000
0.382	7.621	3.531	0.006	54842.719	1.000
0.584	6.606	3.482	0.006	144698.844	2.000
0.584	6.579	3.482	0.006	136801.312	2.000
0.584	6.543	3.481	0.006	134782.016	2.000
0.584	6.493	3.482	0.006	134482.078	2.000
0.583	6.427	3.480	0.006	133406.156	2.000
0.582	6.341	3.479	0.006	132014.422	2.000
0.581	6.299	3.479	0.006	129129.156	2.000
0.579	6.308	3.482	0.006	127886.930	2.000
0.579	6.386	3.485	0.006	131013.109	2.000
0.577	6.334	3.486	0.006	128236.367	2.000
0.577	6.391	3.487	0.006	129981.641	2.000
0.576	6.288	3.487	0.006	130141.203	2.000
0.576	6.376	3.488	0.006	131133.906	2.000
0.576	6.203	3.489	0.006	135950.547	2.000
0.576	6.323	3.488	0.006	136549.953	2.000
0.576	6.182	3.491	0.006	139927.766	2.000
0.576	6.302	3.490	0.006	140553.656	2.000
0.575	6.219	3.492	0.006	146262.203	2.000

The first few columns are extracted seismic attributes. In this example there are 5 attributes. The order of attributes in the input list is consistent with the order of columns. The last column is the facies label, which is consistent with the *polygon* file list, meaning samples from *polygon* file 1 (in this case, *polygon_1*) will have a facies label of 1, and samples from *polygon* file N will have a facies label of N. This training file can then be used as the training file in **psvm3d**, or the supervision file in **som3d** and **gtm3d**.

Generating a mask file for non-uniform training sample extraction in SOM

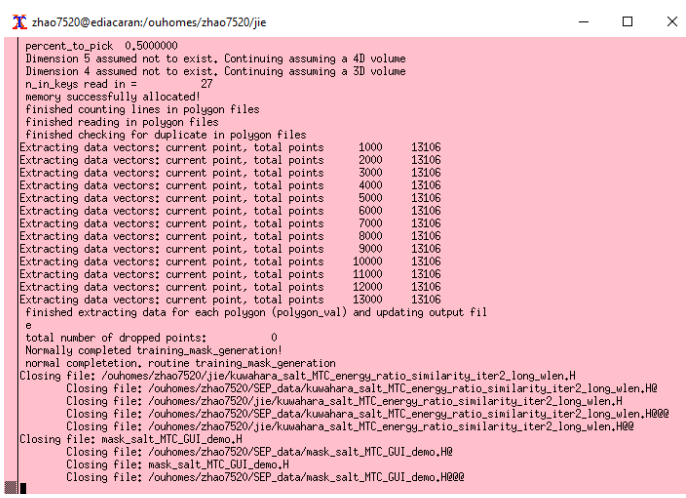
The third panel in this module is for generating an AASPI.H format mask file, which can be further used to define training sample locations in unsupervised seismic facies analysis modules (currently available in SOM). In AASPI unsupervised seismic facies analysis modules, by default a program extracts a subset of data uniformly to build the learning model, then applies this model to the whole data volume. By using a mask file, it substitutes the uniformly sampled subset by samples from user-picked, arbitrary shaped polygonal region(s). Such training subset forces the learning algorithm to focus more on regions of interest. The users also have the option to combine uniformly sampled points with points defined by picked polygons. The functionalities in this panel are as follows:



Arrow 1: “Yes” to generate a uniformly sampled mask background and “No” to mask only locations within picked polygons.

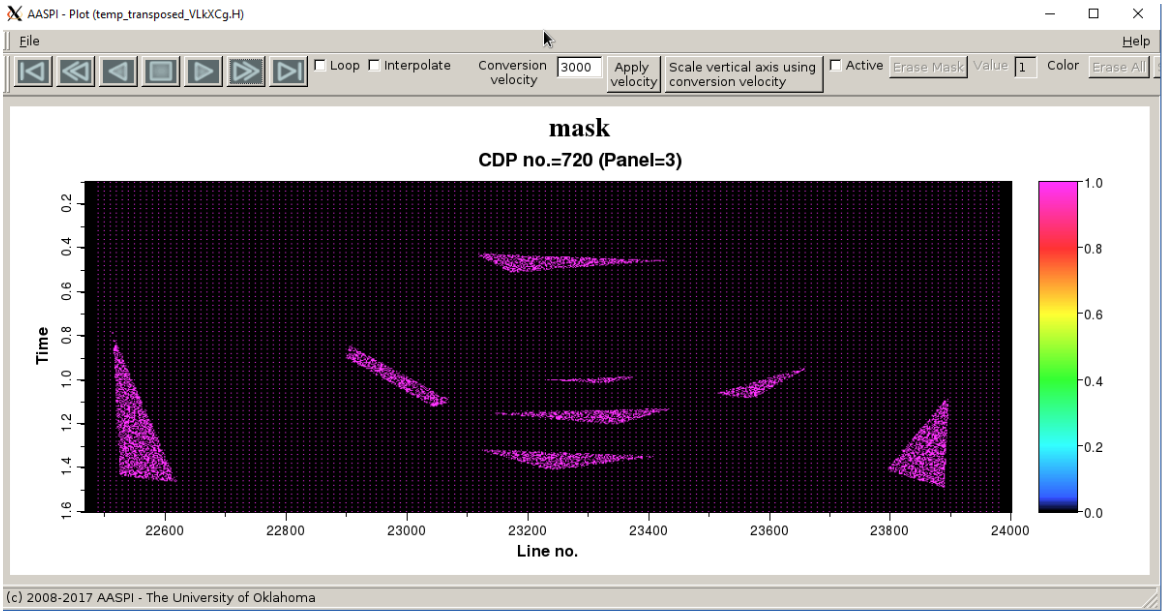
Arrow 2: weight for the background. Users may opt to use different weights for the uniform background and each picked polygon. A higher weight means the corresponding training samples have a higher impact on the subsequent learning model.

Arrow 3: decimation rate for the uniformly sampled background. “5-5-5” means the background consist of one sample of every 5 line, 5 cdp, and 5 time samples, translating to a 1/125 decimation.
Arrow 4: Picked polygons and their corresponding weights. Similar to the previous panel, one polygon file may contain multiple picked 2D polygon regions that define the same facies or depositional environment, e.g. several cross sections from the same channel. A higher weight means the corresponding training samples have a higher impact on the subsequent learning model.
Arrow 5: view the content of a selected polygon file and remove the carriage returns “^M” if the file is generated from a window system. (“^M” will not affect running the program)
Arrow 6: select a template file to initialize the geometry of the mask file to be generated. Normally it is the file on which polygons are picked.
Arrow 7: project name and suffix.
Arrow 8: number of polygon files.
Arrow 9: define the polygon file format.
Arrow 10: vertical axis scaling of the polygon files. Normally if the polygon files are picked from a seismic volume, this scale is 1.
Arrow 11: the percentage of points to be masked within a polygon. Because of the spatial dependence of seismic data, we suggest to randomly select a smaller portion of points within a polygon instead of using all points, reducing the redundancy of training samples.
Arrow 12: execute the mask generation program.

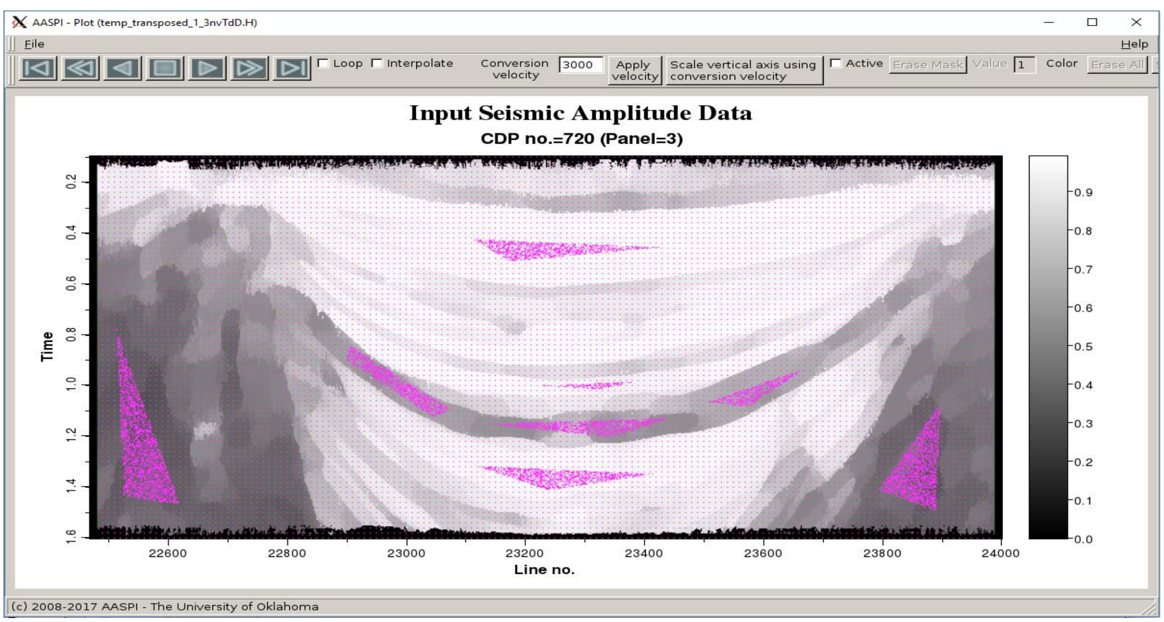


After executing the program, we have an AASPI.H format mask file named as **mask_$projectname_$suffix.H.**

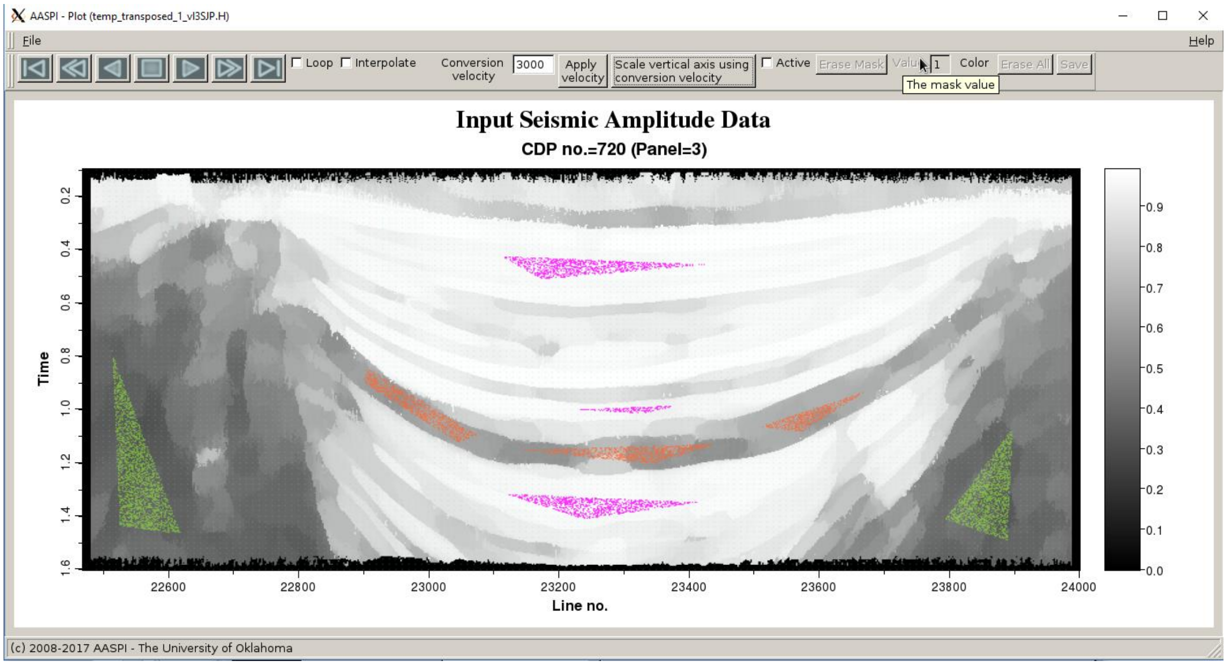
Using the parameters given above, we generate a mask file as a combination of uniform background and user-picked polygons like this (see next page):



By using AASPI **corender** program, we can also display the mask file on top of the template file. In this example, we picked three facies, the salt (left and right triangles), the mass transport complex (center darker region), and the sand/shale interbedded sediments (center lighter regions).



To emphasize picked polygons over the background, we set the weight of the three polygons to 2, 3, and 4 for pure illustration purpose:



Now each facies has a unique color (weight) that corresponds to the impact of such facies to the learning model. We will explore how to incorporate such user-picked mask file in self-organizing map (SOM) facies analysis in the **som3d** documentation.