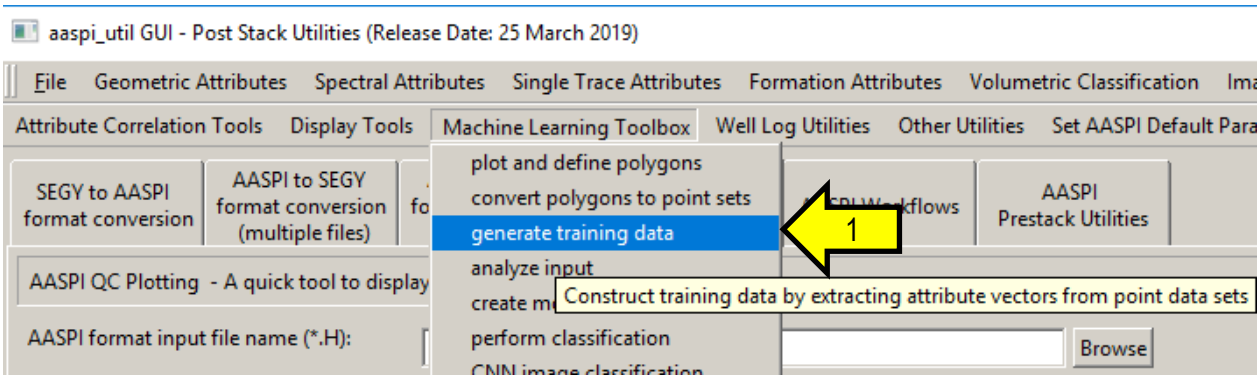
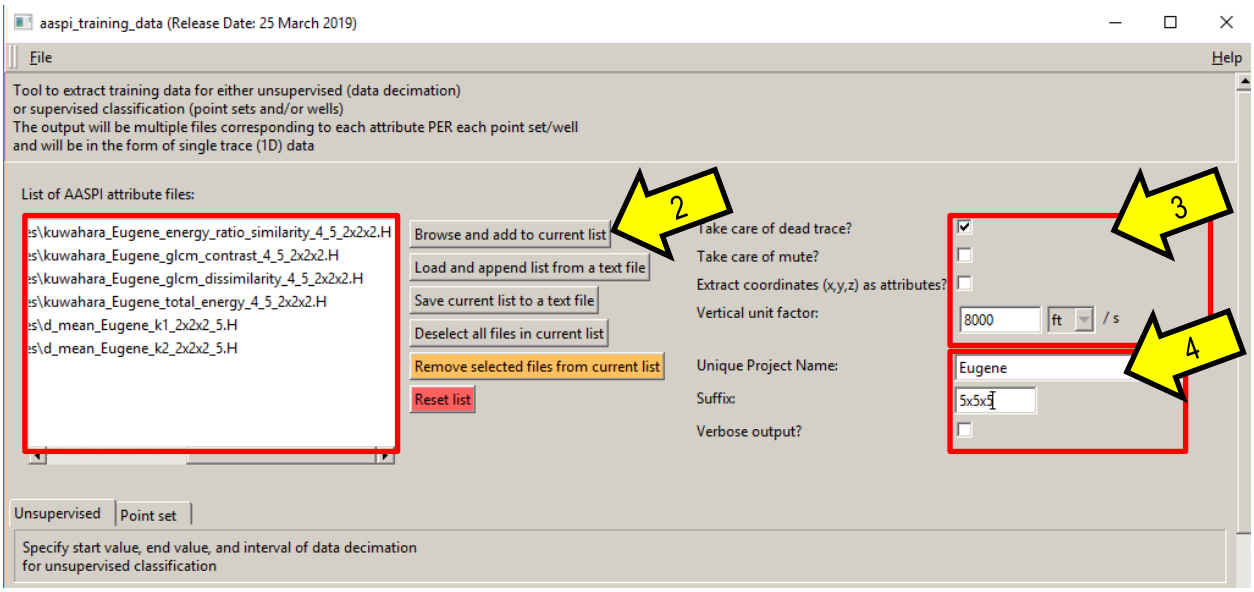
MACHINE LEARNING TOOLBOX: GENERATING TRAINING DATA VECTORS FROM POINT SETS – PROGRAM aaspi_training_data

Contents

Running aaspi_training_data

For unsupervised classification, the first step is to generate a subset of input attributes via decimation (i.e. skipping while reading data at a specific interval, such as 5 inlines x 5 crosslines x 5 vertical samples). For supervised classification, after you convert polygon to point sets, the next step is to extract input attributes at those points to generate training data. Go to *Machine Learning Toolbox* menu → **generate training data** (1).





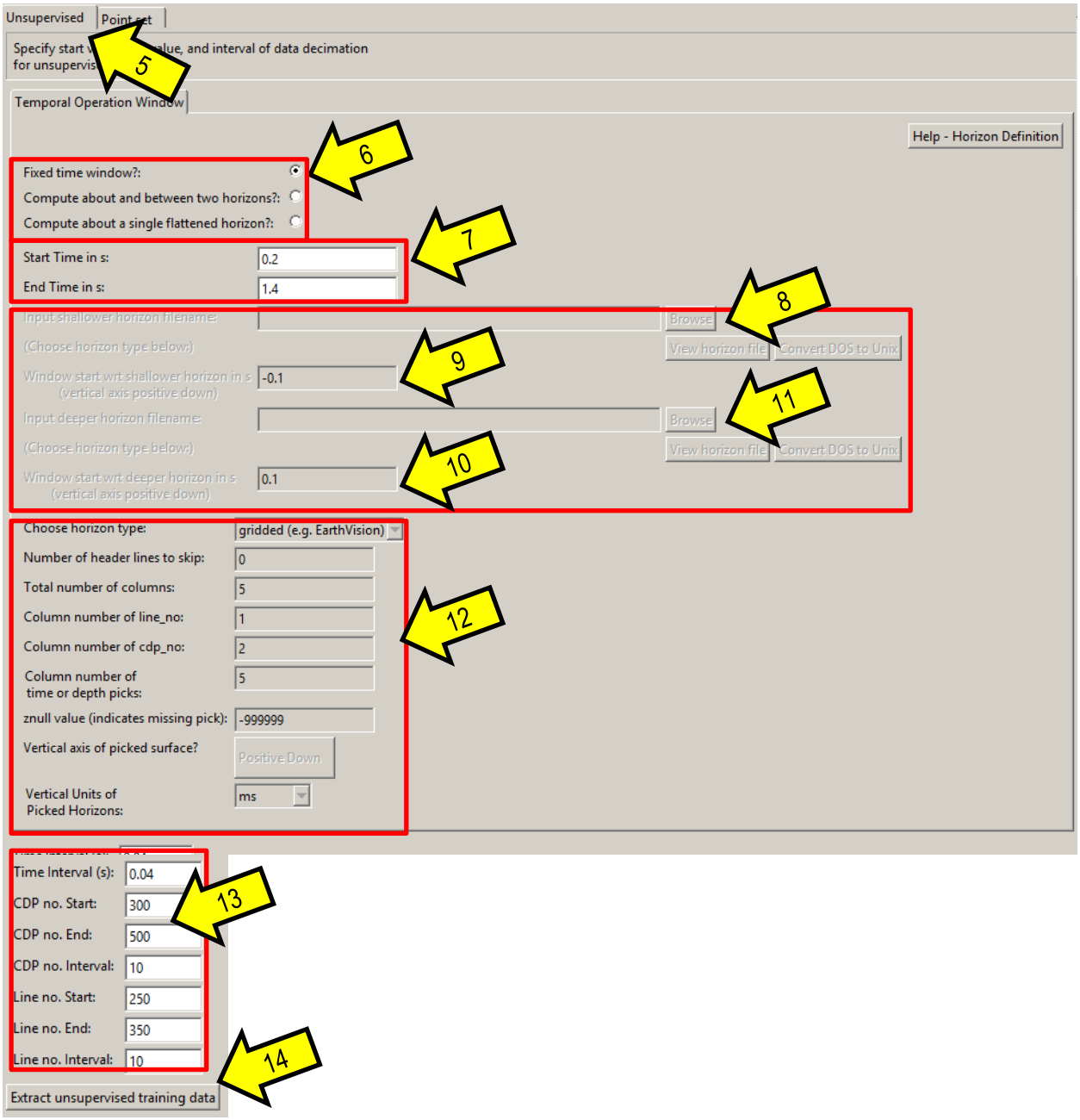
Click "Browse and add to current list" (2) to select input attributes for training data extraction. These attributes should be preconditioned via kuwahara filtering in order to reduce noise and give a more realistic, geological classification, unless you aim to perform fine-scaled reservoir characterization using inversion products (density, impedance, v_p/v_s ratio, etc…) as input attributes.
Choose whether you want dead traces and mute zones to be ignored in extraction and whether to extract the coordinates of data points as spatial attributes (3). If the data is in the time domain, specify a constant velocity to convert time to depth. Give a unique project name and suffix (4).

Unsupervised Training Data Extraction

For unsupervised classification, click on *Unsupervised* tab (5) to define the boundary limits and decimation intervals of vertical axis, crossline, and inline. Vertical axis boundaries can be defined in 3 ways (6):

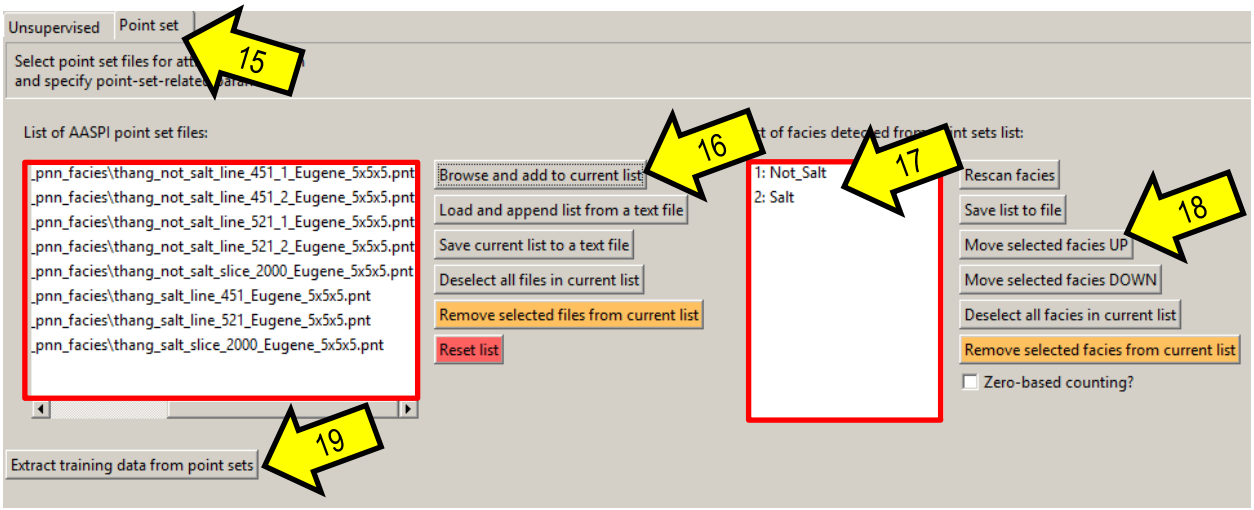
1. Fixed window: specify start and end values for vertical axis (7).
2. About a horizon: browse the first horizon (8) and specify the window size above (9) and below (10) that horizon.
3. Between two horizons: browse upper (8) and lower (11) horizons, then specify the window size above the upper horizon (9) and below the lower horizon (10).

AASPI support EarthVision grid format horizon file type. If your horizon format is different from the standard EarthVision grid, you need to modify the value in box (12).
Decimation intervals, crossline boundaries, and inline boundaries can be defined in box (13).
After all parameters are set, click (14) to extract training data for unsupervised classification.



Supervised Training Data Extraction

For supervised classification, click on *Point set* tab (15). Browse to the AASPI-formatted point set files (16), which were converted from polygons in the previous step of AASPI machine learning workflow. Box (17) displays detected facies names from the browsed point sets. This list is NOT editable and is for informational purpose only. However, the order of facies names can be changed by "*Move selected facies UP/Down*" buttons (18)
After browsing point sets, click "*Extract training data from point sets*" (19).



Output file naming convention

Program **aaspi_training_data** will always generate the following output files:

Output file description	File name syntax
program log information	training_data_*unique_project_name_suffix*.log
program error/completion information	training_data_*unique_project_name_suffix*.err

where the values in red are defined by the program GUI. The errors we anticipated will be written to the **.err* file and be displayed in a pop-up window upon program termination. These errors, much of the input information, a description of intermediate variables, and any software trace-back errors will be contained in the **.log* file.

Results

Unsupervised training data files will have the following naming scheme:
training_data_unsupervised_<attribute_file_name>_<unique_project_name>_<suffix>.H
training_data_unsupervised_<attribute_file_name>_<unique_project_name>_<suffix>.H@

Supervised training data files will have the following naming scheme:
training_data_<point_set_file_name>_<attribute_file_name>.H
training_data_<point_set_file_name>_<attribute_file_name>.H@

Below is a sample of a .H unsupervised training data file:

training_data_unsupervised_amp_pos_amp_pos_0.H

```
training_data_unsupervised.exe:   Thang@   Mon May 13 01:33:54 2019
        in="training_data_unsupervised_amp_pos_amp_pos_0.H@"
        unit=""
        attribute_fn="amp_pos.H"
        limit_type=0
        upper_horizon_fn=""
        lower_horizon_fn=""
        horizon_type="gridded"
        t_start=0.2
        t_end=1.4
        relative_t_start=0
        relative_t_end=0
        skip1=0.04
        min2=300
        max2=500
        skip2=10
        min3=250
        max3=350
        skip3=10
        unique_project_name="amp_pos"
        suffix="0"
        n1=7161 o1=1 d1=1 label1="Sample"
```

The image below is an example of supervised training data file:

training_data_thang_not_salt_line_451_1_Eugene_5x5x5_d_mean_Eugene_k1_2x2x2_5.H

```
training_data_point_set.exe:   Thang@   Tue Mar 26 22:57:30 2019
        n1=1536 o1=1 d1=1 label1="Sample"
        in="training_data_thang_not_salt_line_451_1_Eugene_5x5x5_d_mean_Eugene_k1_2x2x2_5.H@"
        unit=""
        point_set_fn="thang_not_salt_line_451_1_Eugene_5x5x5.pnt"
        attribute_fn="d_mean_Eugene_k1_2x2x2_5.H"
        facies_list_fn="training_data_point_set_facies_list_Eugene_5x5x5.txt"
        facies_name="Not_Salt"
        facies_no=1
        zero_based="n"
        unique_project_name="Eugene"
        suffix="5x5x5"
```

In addition, a list of input attributes for training data extraction is saved to:
training_data_attribute_list_<unique_project_name>_<suffix>.txt

training_data_attribute_list_Eugene_5x5x5.txt

```
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\kuwahara_Eugene_energy_ratio_similarity_4_5_2x2x2.H
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\kuwahara_Eugene_glcm_contrast_4_5_2x2x2.H
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\kuwahara_Eugene_glcm_dissimilarity_4_5_2x2x2.H
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\kuwahara_Eugene_total_energy_4_5_2x2x2.H
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\d_mean_Eugene_k1_2x2x2_5.H
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\d_mean_Eugene_k2_2x2x2_5.H
```

For unsupervised classification, a list of all extracted training data files is provided in
training_data_unsupervised_extracted_list_<unique_project_name>_<suffix>.txt

training_data_unsupervised_extracted_list_amp_pos_0.txt

```
training_data_unsupervised_amp_pos_amp_pos_0.H
training_data_unsupervised_conf_amp_pos_0_amp_pos_0.H
training_data_unsupervised_crossline_dip_amp_pos_0_amp_pos_0.H
```

For supervised classification, a list of all extracted training data files is provided in
training_data_point_set_extracted_list_<unique_project_name>_<suffix>.txt
Point set list and detected facies list are also saved to text files:
training_data_point_set_list_<unique_project_name>_<suffix>.txt
training_data_point_set_facies_list_<unique_project_name>_<suffix>.txt

training_data_point_set_extracted_list_Eugene_5x5x5.txt

```
training_data_thang_not_salt_line_451_1_Eugene_5x5x5_kuwahara_Eugene_energy_ratio_similarity_4_5_2x2x2.H
training_data_thang_not_salt_line_451_1_Eugene_5x5x5_kuwahara_Eugene_glcm_contrast_4_5_2x2x2.H
training_data_thang_not_salt_line_451_1_Eugene_5x5x5_kuwahara_Eugene_glcm_dissimilarity_4_5_2x2x2.H
training_data_thang_not_salt_line_451_1_Eugene_5x5x5_kuwahara_Eugene_total_energy_4_5_2x2x2.H
training_data_thang_not_salt_line_451_1_Eugene_5x5x5_d_mean_Eugene_k1_2x2x2_5.H
training_data_thang_not_salt_line_451_1_Eugene_5x5x5_d_mean_Eugene_k2_2x2x2_5.H
training_data_thang_not_salt_line_451_2_Eugene_5x5x5_kuwahara_Eugene_energy_ratio_similarity_4_5_2x2x2.H
training_data_thang_not_salt_line_451_2_Eugene_5x5x5_kuwahara_Eugene_glcm_contrast_4_5_2x2x2.H
training_data_thang_not_salt_line_451_2_Eugene_5x5x5_kuwahara_Eugene_glcm_dissimilarity_4_5_2x2x2.H
training_data_thang_not_salt_line_451_2_Eugene_5x5x5_kuwahara_Eugene_total_energy_4_5_2x2x2.H
training_data_thang_not_salt_line_451_2_Eugene_5x5x5_d_mean_Eugene_k1_2x2x2_5.H
training_data_thang_not_salt_line_451_2_Eugene_5x5x5_d_mean_Eugene_k2_2x2x2_5.H
training_data_thang_not_salt_line_521_1_Eugene_5x5x5_kuwahara_Eugene_energy_ratio_similarity_4_5_2x2x2.H
training_data_thang_not_salt_line_521_1_Eugene_5x5x5_kuwahara_Eugene_glcm_contrast_4_5_2x2x2.H
training_data_thang_not_salt_line_521_1_Eugene_5x5x5_kuwahara_Eugene_glcm_dissimilarity_4_5_2x2x2.H
training_data_thang_not_salt_line_521_1_Eugene_5x5x5_kuwahara_Eugene_total_energy_4_5_2x2x2.H
```

training_data_point_set_list_Eugene_5x5x5.txt

```
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\thang_not_salt_line_451_1_Eugene_5x5x5.pnt
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\thang_not_salt_line_451_2_Eugene_5x5x5.pnt
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\thang_not_salt_line_521_1_Eugene_5x5x5.pnt
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\thang_not_salt_line_521_2_Eugene_5x5x5.pnt
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\thang_not_salt_slice_2000_Eugene_5x5x5.pnt
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\thang_salt_line_451_Eugene_5x5x5.pnt
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\thang_salt_line_521_Eugene_5x5x5.pnt
D:\AASPI_GIT.git\aaspi_testing_pnn_facies\thang_salt_slice_2000_Eugene_5x5x5.pnt
```

training_data_point_set_facies_list_Eugene_5x5x5.txt

```
Not_Salt
Salt
```

These training data files and text-based lists will be used to analyze input and create models for AASPI machine learning workflow in subsequent steps.