Volumetric principle component analysis for 3D SEISMIC FACIES ANALYSIS – PROGRAM pca3d

Overview

Principal component analysis (PCA) is widely used to reduce the redundancy and excess dimensionality of the input attribute data. Such reduction is based on the assumption that most of the signals are preserved in the first few principle components (eigenvectors) and the last eigenvectors are mainly noise or redundant data that are already captured in the first few eigenvectors. While many workers use PCA to reduce redundant attributes into "meta attributes" to simplify the computation, in AASPI, we use PCA as the first iteration of the SOM and GTM algorithms, but a standalone module for volumetrically PCA analysis was not provided in previous releases. **pca3d** is developed for users who want to analyze data before feeding them into more complicated "black-box" algorithms, and for users who want to identify the most representative features in multiple seismic attributes. The first principal component (or first eigenvector) is a vector in *N*-dimensional attribute space that least-squares fits the data. If we then project the data onto this vector and subtract it, the second principal component (eigenvector) is the vector in (*N-1*)-dimensional space that least-squares fits the data residual that was not represented by the first eigenvector. This process continues for all *N*-dimensions resulting in *N* principal components.

Theory

Given a suite of *N* attributes, the covariance matrix is defined as

$$C_{mn} = \frac{1}{J}\sum_{j=1}^{J}\left(a_{jm}(t_j, x_j, y_j) - \mu_m\right)\left(a_{jn}(t_j, x_j, y_j) - \mu_n\right),$$

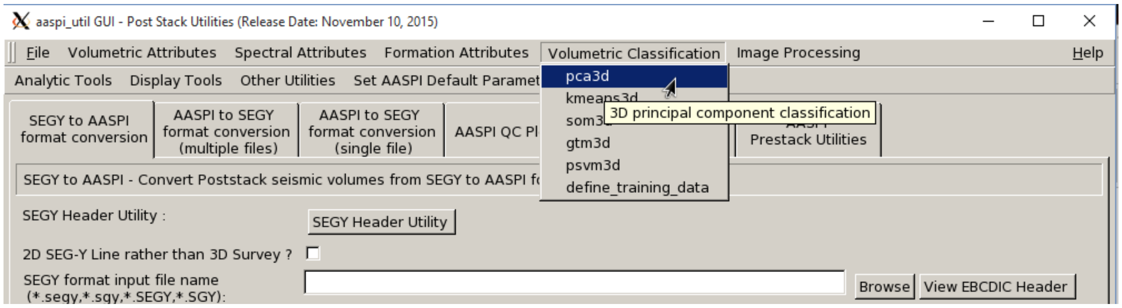
where a_{jm} and a_{jn} are the m^{th} and n^{th} attributes, J is the total number of data vectors, and where

$$\mu_n = \frac{1}{J}\sum_{j=1}^{J} a_{jn}(t_j, x_j, y_j),$$

is the mean of the n^{th} attribute. If we compute the eigenvalues, λ_i, and eigenvectors, $\mathbf{v}_i$, of the real, symmetric covariance matrix, **C**, the i^{th} principal component at data vector j is defined as

$$p_{ji} = \sum_{n=1}^{N} a_{jn}(t_j, x_j, y_j)\, v_{ni},$$

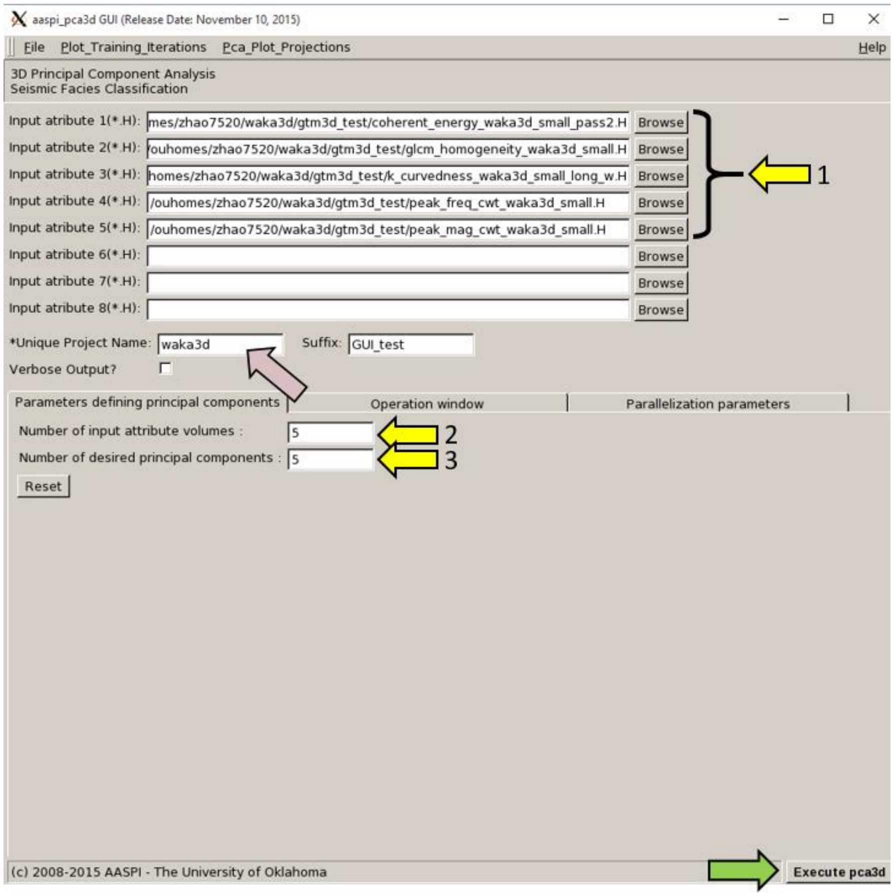
where v_{ni} indicates the n^{th} attribute component of the i^{th} eigenvector.



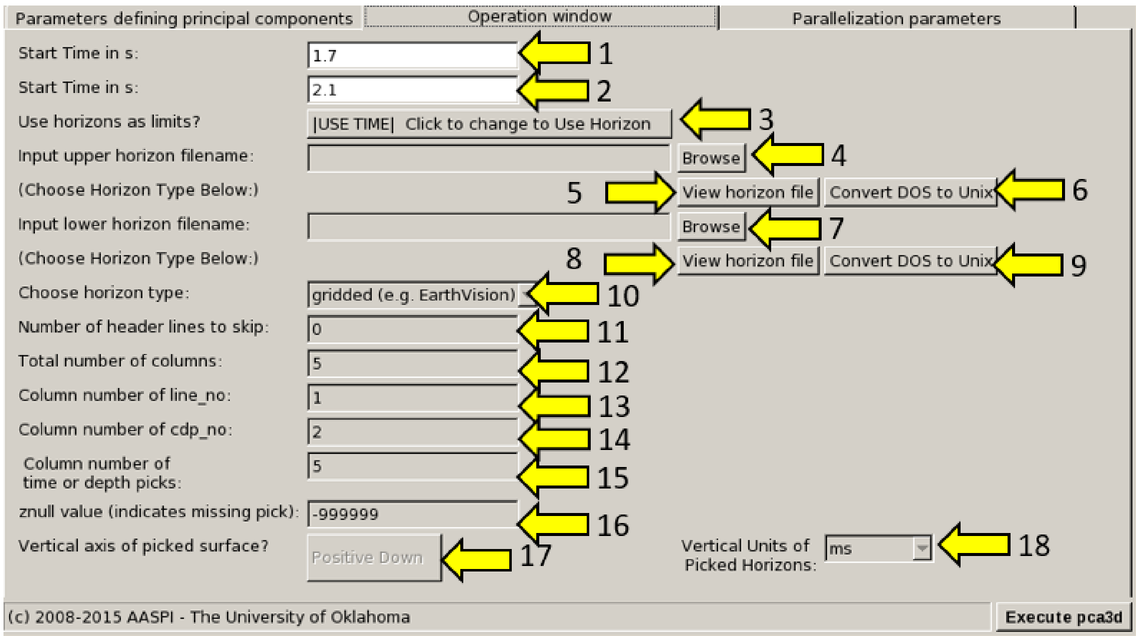
This Program **pca3d** is launched from the *Volumetric Classification* in the main **aaspi_util** GUI

Computing pca3d module

Setting the parameters defining principle components is the first step of analysis. Use the browser on the first eight lines to choose the input seismic data file (*Arrow 1*). It is not mandatory to take in eight inputs. The number of inputs can vary from two – eight. The input attributes that one considers for facies analysis will vary according to the specific applications. For identifying the depositional facies variation the volumetric attributes such as dip magnitude, coherency, GLCM attributes, spectral magnitude, coherent energy can be considered as input. For characterizing geo-mechanical variation in shale plays one should consider different volumes that helps in identifying the rock physics such as inversion volumes, lambda-rho, mu-rho, intercept or gradient AVO volumes, etc. Specify the number of input attributes in the field labeled “Number of attributes to use” (*Arrow 2*). This value will be updated automatically when a file is selected. Do not forget to give a “Unique Project Name”. A z-score algorithm is used to normalize the input files. Then specify the number of desired principle components to output (*Arrow 3*). The number of principle components cannot go beyond the number of input attributes.



Then a user need to define the operation window in the **Operation Window** tab shown below. A user can either use a fixed time window, or a window defined by two horizons.



Horizon definition

The horizon definition panel will look the same for almost all AASPI GUIs:

1. Start time (upper boundary) of the analysis window.
2. End time (lower boundary of the analysis window.
3. Toggle that allows one to do the analysis between the top and bottom time slices described in 1 and 2 above, or alternatively between two imported horizons. If *USE HORIZON* is selected, all horizon related options will be enabled. If the horizons extend beyond the window limits defined in 1 and 2, the analysis window will be clipped.
4. Browse button to select the name of the upper (shallower) horizon.
5. Button that displays the horizon contents (see Figure 1).
6. Button to convert horizons from Windows to Linux format. If the files are generated from Windows based software (e.g. Petrel), they will have the annoying carriage return (^M) at the end of each line (Shown in Figure 1). Use these two buttons to delete those carriage returns. Note: This function depends on your Linux environment. If you do not have the program **dos2unix** it may not work. In these situations, the files may have been automatically converted to Linux and thus be properly read in.
7. Browse button to select the name of the lower (deeper) horizon.
8. Button that displays the horizon contents (see Figure 1).
9. Button to convert horizons from Windows to Linux format. (see 6 above).
10. Toggle that selects the horizon format. Currently *gridded* (e.g. EarthVision in Petrel) and *interpolated* (ASCII free format, e.g. SeisX) formats are supported. The gridded horizon are nodes of B-splines used in mapping and have no direct correlation to the seismic data survey. For example, gridded horizons may be computed simply from well tops. The x and y locations are aligned along north and east axes. In contrast interpolated horizons have are defined by *line_no, cdp_no (crossline_no)* and *time* triplets for each trace location. Examples of both format are shown in Figure 1. If *interpolated* is selected, the user needs to manually define each column in the file.
11. Number of header lines to skip in the *interpolated* horizon files.
12. Total number of columns in the *interpolated* horizon files.
13. Enter the column number containing the *line_no (inline_no)* of the interpolated data triplet.
14. Enter the column number containing the *cdp_no (crossline_no)* of the interpolated data triplet.
15. Enter the column number containing the *time* or *depth* value of the interpolated data triplet.
16. *Znull* value (indicate missing picks) in the horizon files.
17. Toggle to choose between positive down and negative down for the horizon files (e.g. Petrel uses negative down).
18. Choose the vertical units used to define the horizon files (either *s*, ms, *kft, ft, km,* or *m*).

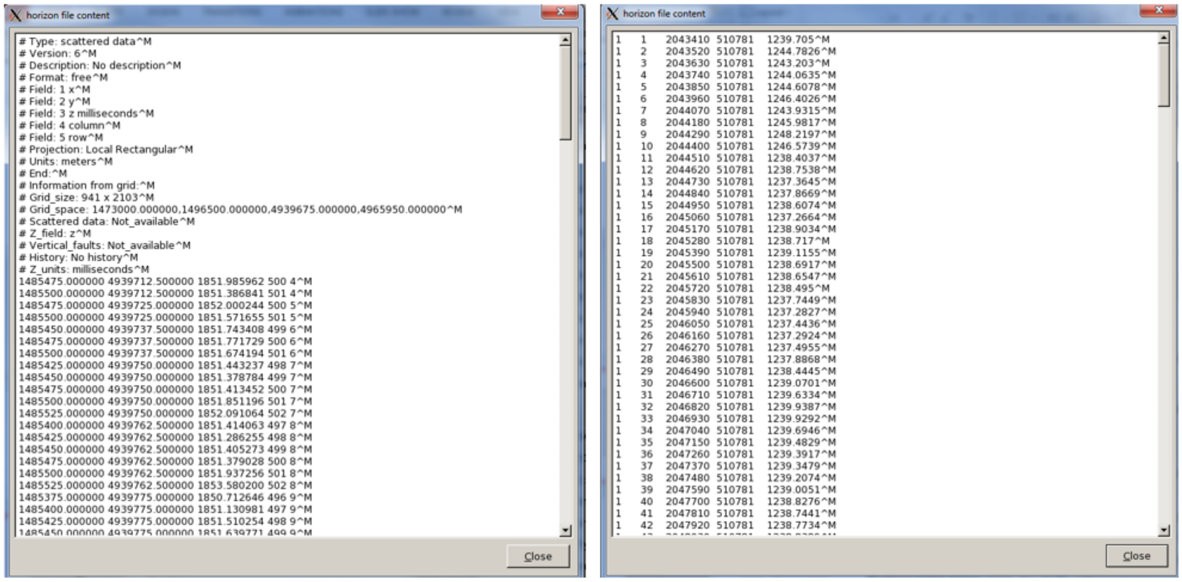
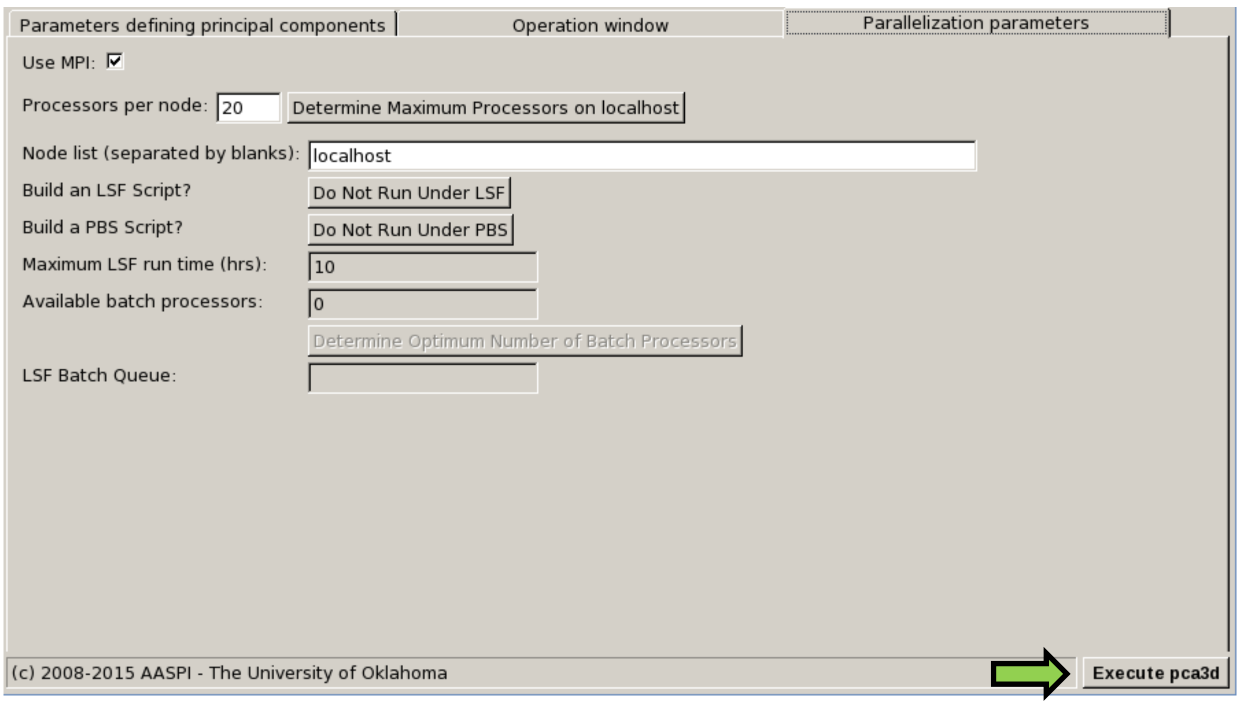


Figure 1. (left) A gridded horizon file (EarthVision format). (right) An interpolated horizon file with five columns (ASCII free format).

After defining the parallelization parameters, press the *Execute pca3d*.

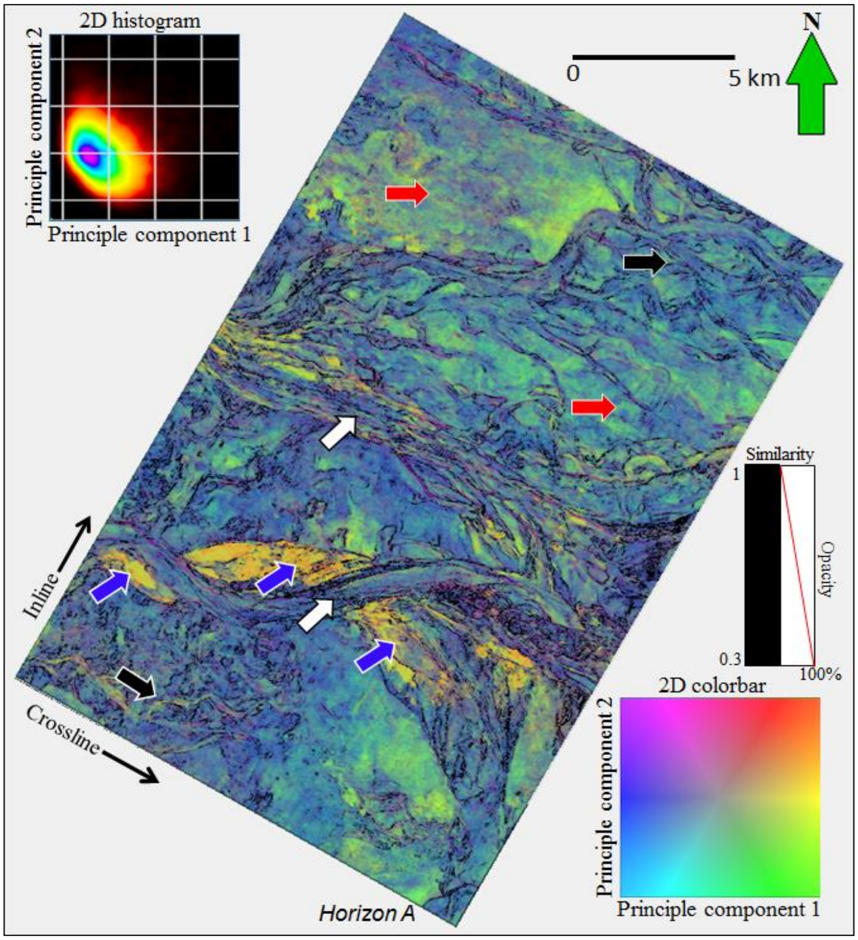


The generated principle component files are named as:
pc_${unique project name}_${suffix}__j.H

j is the principle component order.

Visualization of the result

To view the resulted principle components, a user can either use aaspi_plot to display each component individually, or use **aaspi_crossplot** to crossplot two components. One can also use crossplot tools in commercial interpretation packages. The figure below is a crossplot of first two principle components using the horizon probe in Petrel, along a horizon in a turbidite system in Canterbury basin, New Zealand. We interpret the white arrows as multistoried channels, black arrows as sinuous channel complexes, blue arrows as a sand filled channel, and red arrows as slope fans.



References

Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation, **3**, SAE29-SAE58.