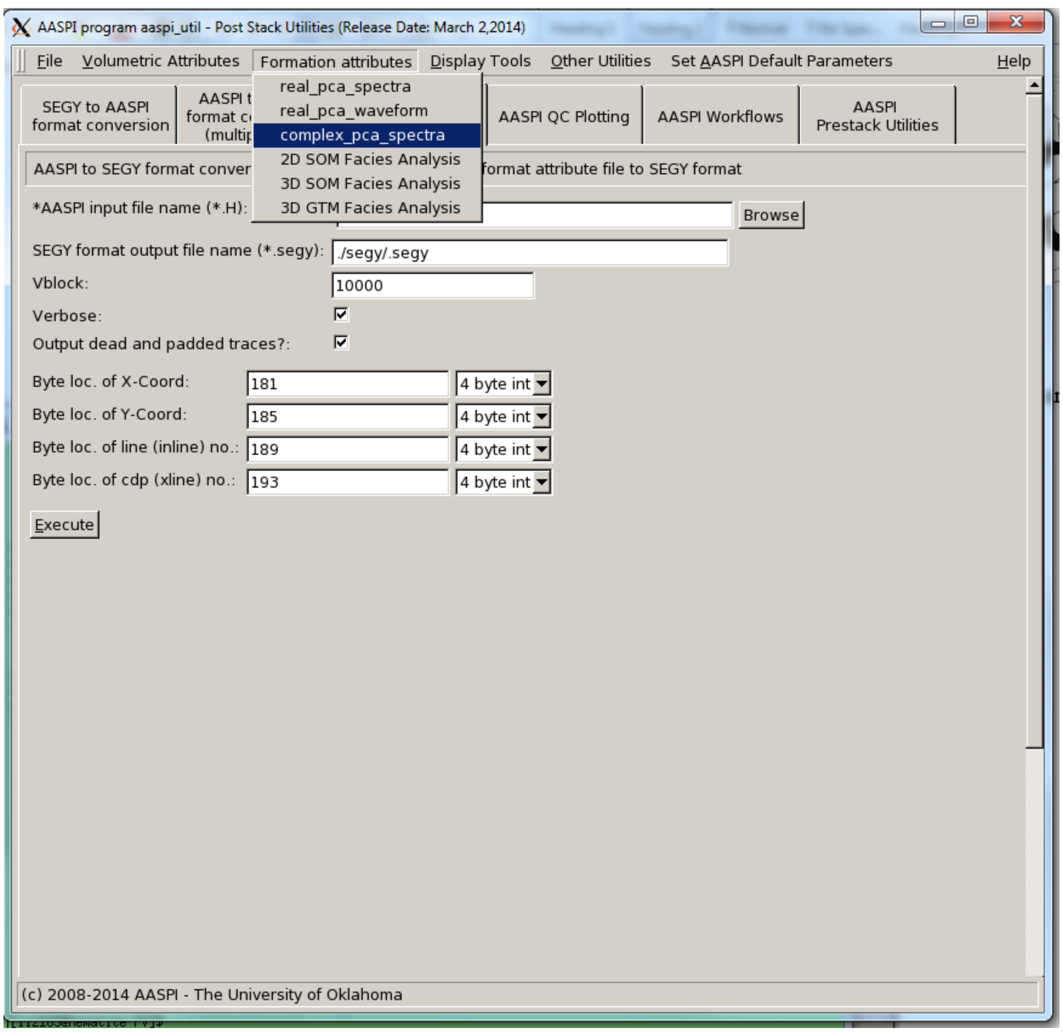
# COMPLEX PRINCIPAL COMPONENT SPECTRA EXTRACTION

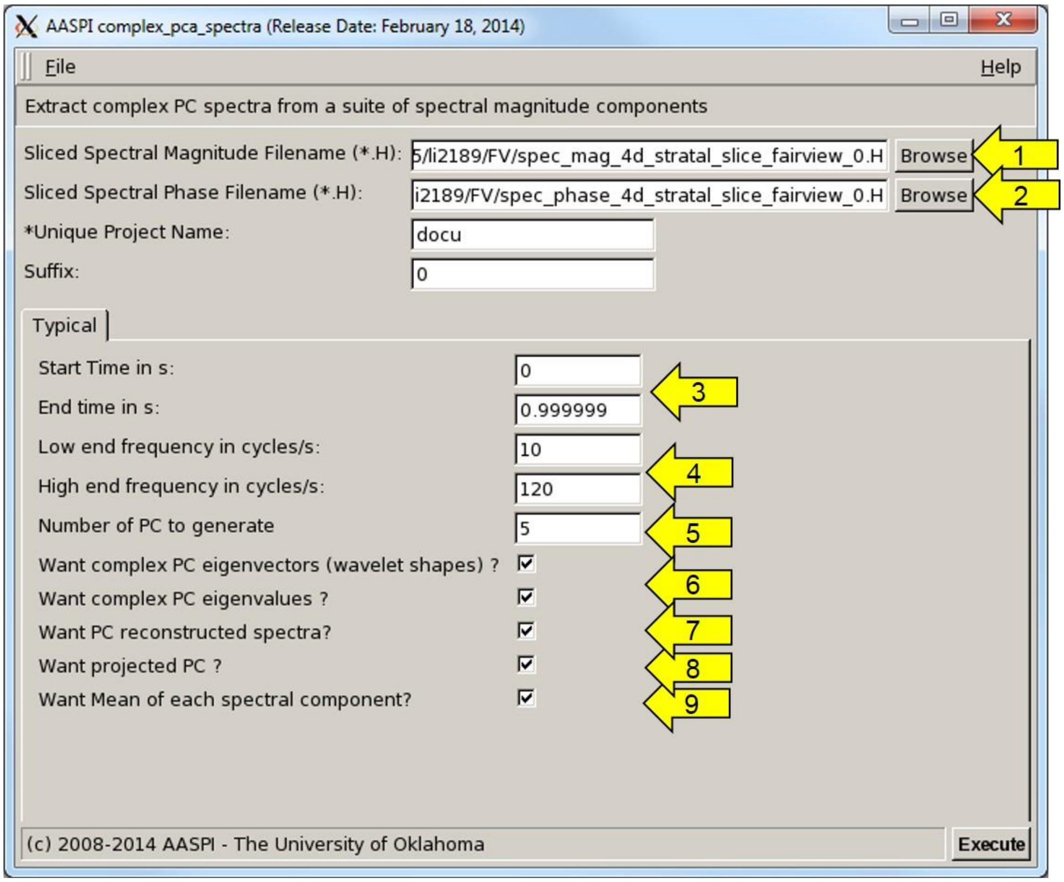
# – PROGRAM complex_pca_spectra

## Computing principal components

To begin, click the Formation attributes tab in the AASPI-UTIL window and select program **complex_pca_spectra**:



Program **complex_pca_spectra** performs complex stratal slice filter by Principal Component Analysis (PCA). The following window appears:




First, enter the (1) name of the *Spectral magnitude* and (2) name *spectral phase stratal slice input files(*.H)* file you wish to filter, generated by **Complex_stratal_slice** as well as a *Unique Project Name* and *Suffix* as you have done for other AASPI programs. (3) and (4) are the properties of the input complex stratal slices, and would be filled automatically. Here, you can change the lowest and highest frequency (in Hz) to be analyzed.

## Principal Component Analysis

Seismic data are so noisy that we need to condition date before using Q estimation methods. Here we filter the stratal spectra slices by using Principal Component Analysis (PCA), an orthogonal linear transformation that estimates the correlation between measurements. There are a number of methods to implement the PCA algorithm. We employed the one described by Guo et al. (2009). Given the spectra of a data volume, each frequency component represents a measurement. The measurements are normalized to zero-mean, then input to the PCA algorithm. If the spectra along a given horizon or stratal slice have the same shape, they will be completely defined by the first eigenspectrum. Eigenspectra that are associated with second, third, fourth and so on, largest eigenvalues reveal variations in the spectra that are critical to Q estimation algorithm.

Principal Component Analysis is a linear orthogonal transformation, $P=[p_1, p_2, p_3, \cdots, p_m]^T$ , that transforms a set of correlated observations, $X=[x_1, x_2, x_3, \cdots, x_n]^T$ where the vector $x_{\mathrm{j}}$ is the jth observation, to a new domain such that the transformed data, $\mathrm{Y}=[y_1, y_2, y_3, \cdots, y_n]$, consist of only uncorrelated events. The relationship between X and Y is represented by:

$$PX = Y \, , \tag{1}$$

The set of all vectors $p_j$ makes up the principal components of X. Each vector $y_j$ is the projection $x_j$ onto all the principal components $p_k$.

To calculate P, we first compute the covariance matrix of X, $C_{\mathrm{X}}$:

$$C_{\mathrm{X}} = \frac{1}{n} XX^T \, , \tag{2}$$

Where the element $C_{\mathrm{X}}(i, j)$ is the cross-correlation between two observation $x_i$ and $x_j$.

The covariance matrix in the transformed domain, $C_{\mathrm{Y}}$, is related to $C_{\mathrm{X}}$ by:

$$C_{\mathrm{Y}} = PC_{\mathrm{X}}P^T \, , \tag{3}$$

Where, $C_{\mathrm{Y}}$ is a diagonal matrix since Y consists of uncorrelated events.

$C_{\mathrm{X}}$ is a symmetric positive definite matrix. Therefore, its eigenvector matrix V is an orthonormal matrix and satisfies $V^T = V^{-1}$. It can be verified that $P=V$ is a solution to equation (3).

By picking $P=V$, P can be rewritten as the sum of weighted outer-products of the principal components (eigenvalue decomposition):

$$P = \sum_{j=1}^{m} e_j p_j p_j^T , \qquad (4)$$

Where $e_j$ is the eigenvalue associated with the eigenvector $p_j$.

Substituting equation (4) into equation (1) gives:

$$Y = \left[ \sum_{j=1}^{m} e_j p_j p_j^T \right] X , \qquad (5)$$

Without loss of generality, suppose the eigenvalues $e_j$ 's are sorted in descending order $e_1 > e_2 > \cdots > e_m > 0$. Then define $Y_k$ to be the partial sum of $Y$ :

$$Y_k = \left[ \sum_{j=1}^{k} e_j p_j p_j^T \right] X , \qquad (6)$$

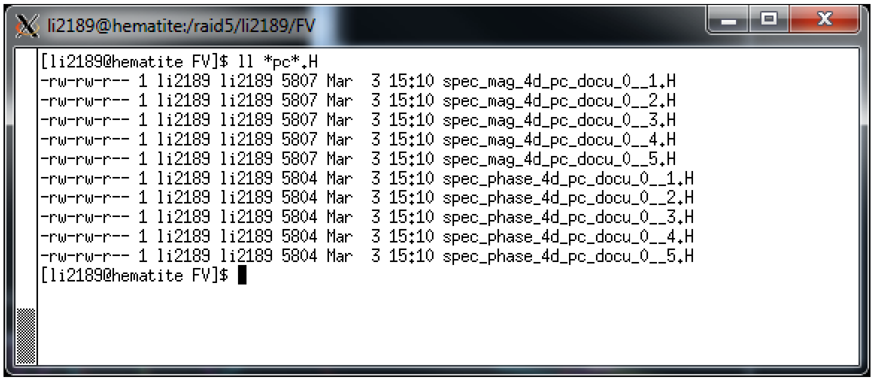
In practice, assuming that the large variance is associated with important structures and small variance is associated with noise, limiting the number of terms in the sum is a filter that enhances the SNR.

The PCA can be extended to apply to complex data. The mathematics is the same as for real data, except the covariance matrix of X, $C_X$, is a self-adjoint matrix: $C_X = C_X^H$.

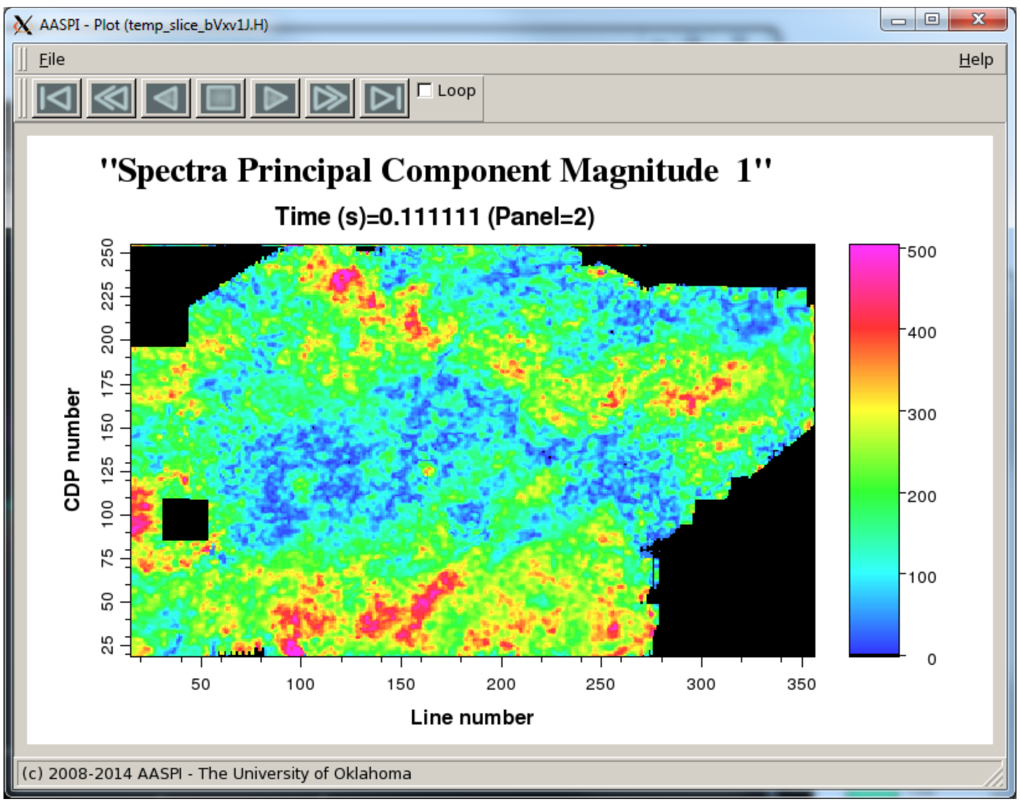
Number of principal components (5) is the number to generate, in this case is 5. Others are the boxes for output file options. Their names can stand for their meanings.

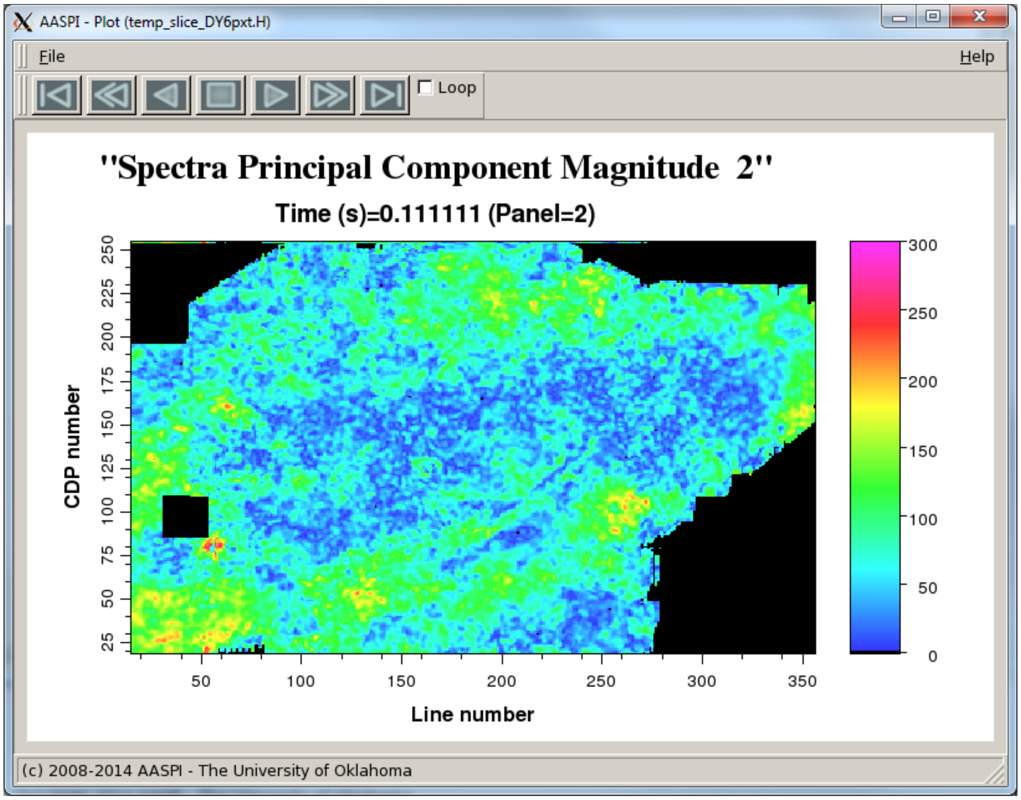
After executing the module, we can get filtered stratal slices generated by different principal components.

XX. Extract complex PCA spectra – Program **complex_pca_spectra**



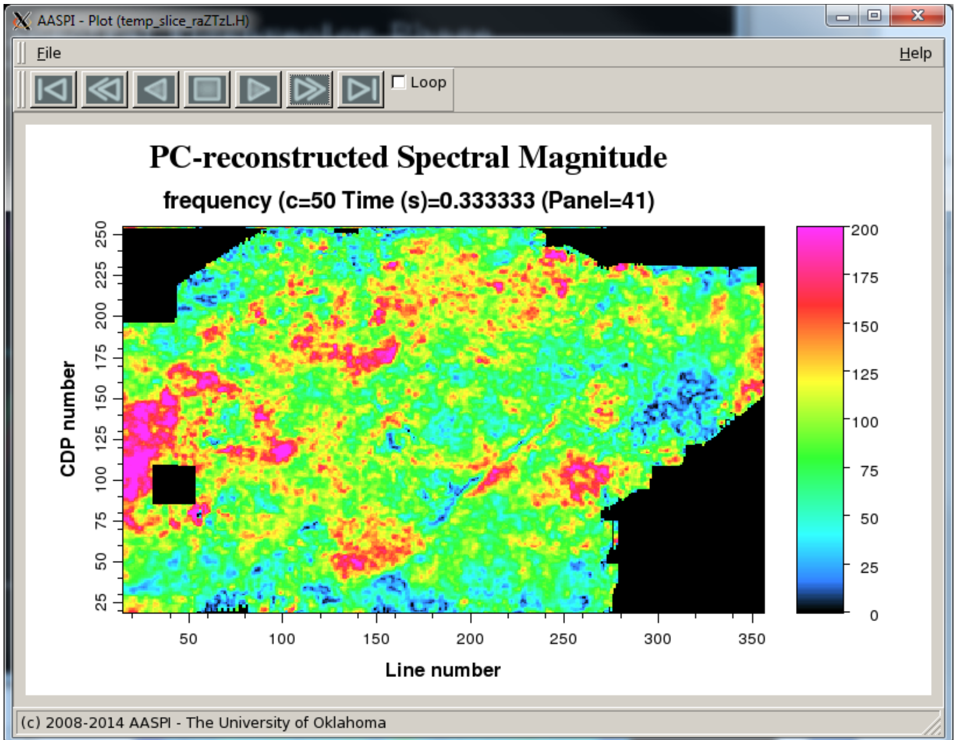

We can apply AASPI plot to draw the results, using magnitude slices as an instance. Here, we show the first two principal components.

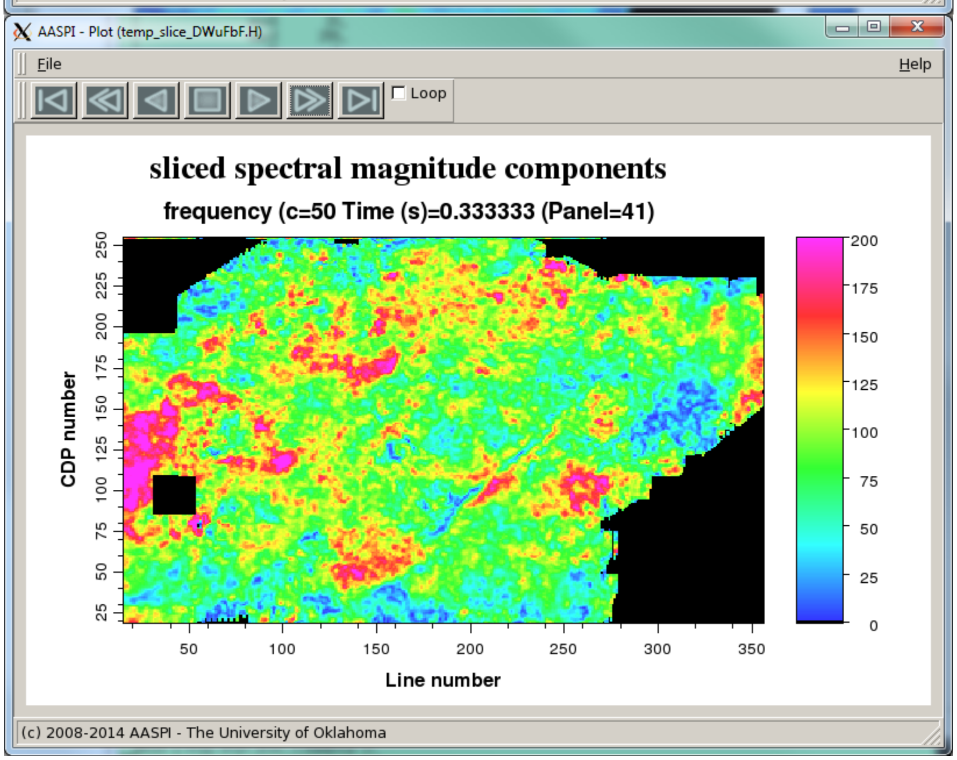





We also output the PC reconstructed data. The reconstruction data have the same structure as the input data. As comparison, we show the original data as well. Since 5 principal components contain most of the information, the difference between them is not obvious.

XX. Extract complex PCA spectra – Program **complex_pca_spectra**

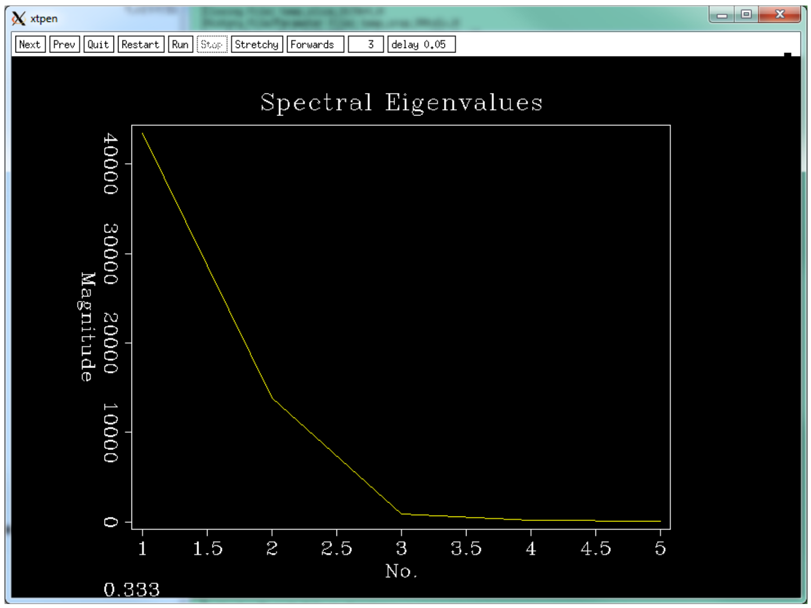



To better understand how big portion every principal component cover, we plot the eigenvalues's magnitude. For now, we use some simple SEP commands and plot the spectral wavelets can be plotted by typing:

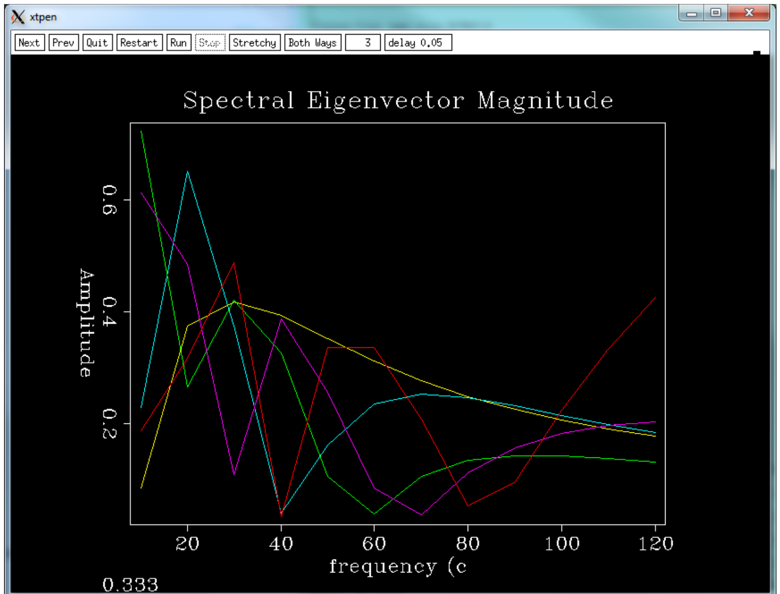
*Graph < spectra_eigenvalue_docu_0.H* | *Stube &*

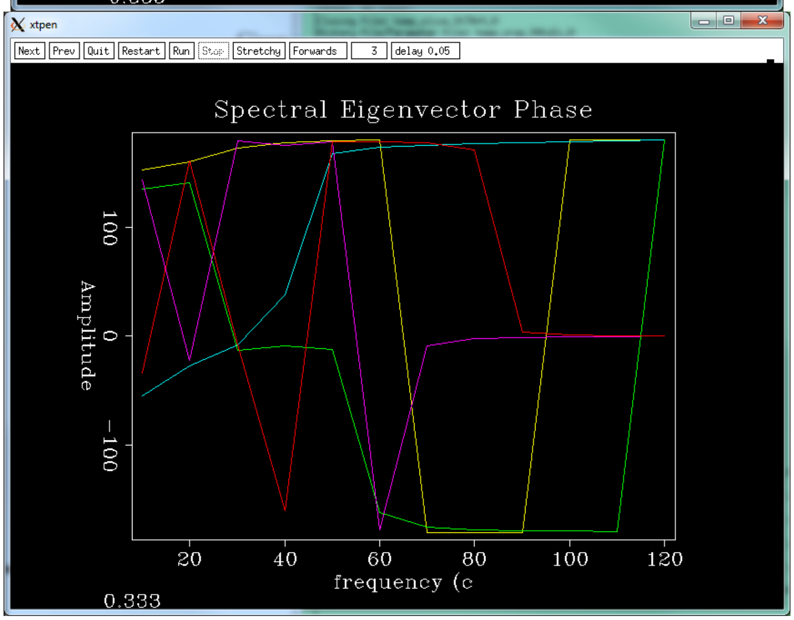
For the t = 0.333s slice, note that the first 3 components have the most energy of the whole data, so for further filtering, you can just use 1 or 2 principal components.




We can also use the SEP commands to show the eigenvectors' magnitude and phase:

XX. Extract complex PCA spectra – Program **complex_pca_spectra**




Other outputs can be shown in a similar way.

## References

Guo, H., K. J. Marfurt, and J. Liu, 2009, Principal component spectral analysis, Geophysics, **74**, 35-43.