ATMS 587: Notes on Empirical Orthogonal Functions and Principal Components
D.L. Hartmann November 2002.

Suppose we have a variable expressed in terms of space and time y(x,t). In Fourier analysis we can express this data set in terms of a specified set of functions, normally signs and cosines. If y(x,t)is defined on the interval 0 < x < L, then we can write:

$$y(x,t) = \sum_{n=0}^{N} A_n(t)\cos(2\pi n/L) + B_n(t)\sin(2\pi n/L)$$

Now instead of being expressed as a function of space and time, the spatial dependence is expressed in terms of the functional forms of sines and cosines with different frequencies. Each function has n maxima and n minima on the interval $0<x<L$. The time dependence is expressed in the coefficients $A_n(t)$ and $B_n(t)$ that express the amplitudes of the set of basis functions $\cos(2\pi n/L)$ *and* $\sin(2\pi n/L)$ as functions of time. So now the same information that was contained in *y(x,t)* is contained in $A_n (t)$ and $B_n (t)$, but expressed differently.

In Empirical Orthogonal Function (EOF) analysis we do a very similar thing, except that the spatial structures are not specified beforehand, but rather are determined by the structure of the data itself. The first EOF expresses that maximum fraction of the variance of the original data set that is possible with a single functional form. The second explains the maximum amount of the variance remaining with a function that is orthogonal to the first, and so on. Finding the structures that do this is done by diagonalizing the covariance matrix of the original data, and the eigenfunctions and eigenvalues of this mathematical process are the EOF's and the variance explained by that eigenfunction, respectively. So the first EOF is the spatial structure that explains the maximum possible amount of the variance, and its eigenvector is the amount of variance that it explains. If you plot the eigenvalues as a function of their index value (1, 2, 3, . . . N), then this eigenvalue spectrum gives you an idea of how exceptional the first few eigenvectors are. You can divide each eigenvalue by the sum of all the eigenvalues and then the result is the fraction of the total variance explained by each eigenvalue. To be useful EOF analysis must result in an decomposition of the data in which a big fraction of the variance is explained by the first few EOFs.

Since the EOFs are an orthogonal and complete function set within the data set, you can project the data on these functions like they were sines and cosines. The resulting amplitudes as a function of time correspond to the $A_n (t)$ and $B_n (t)$ of Fourier analysis, but we call them the Principal Components, or PCs. So the PCs show the amplitude versus time of the individual EOFs.

You can learn all about these procedures and much more in ATMS 552 next quarter.