

Why Use EOF Analysis?

Short answer: It is a useful way of organizing information.

Analogy to Fourier Analysis

I'll make an analogy to the more familiar Fourier analysis. Let's say we have a time series of N points. We can convert these N pieces of information into the different format of a time series mean and $N-1$ sines and cosines. All we are doing is representing the information as a linear combination of different amplitude sines and cosines instead of points in time with different amplitudes.

Why would we want to do this? Perhaps what we are interested in are the low-frequency variations in the time series. These are easily characterized by the amplitudes of the low-order sines and cosines. Perhaps we want to represent the data more compactly. If the high frequencies have little variance then we can cut out the high order sines and cosines and still have nearly the same original time series. But we cannot simply cut points out of the time series and still preserve nearly the same information.

It is important to keep in mind we are not restricted to using sines and cosines. If the time series has "jumps" in it then we would need large-amplitude high-order sines and cosines to characterize those jumps. Instead we would be better off using a different set of orthogonal functions to characterize the time series. A set of functions with "jumps" in them could allow us to characterize the time series primarily with the lower-order functions. The appropriate set of *orthogonal functions* to use could be *empirically* determined from the time series itself.

Empirical Orthogonal Function Analysis

All EOF analysis does is determine a set of orthogonal functions that characterizes the covariability of time series for a set of grid points in a simple way. Instead of X grid points each with N values in time we have X EOF patterns each with N values in time. In fact, we need not even use time series and spatial grid points. EOF analysis could just as well be applied to concentrations of X different chemical compounds from N different experiments.

How are the EOFs chosen? Well we could ask the question, what index time series could we construct that produces regression/correlation maps with the overall strongest amplitudes? Such a regression map and index time series would turn out to be the leading EOF. We could then ask the question, if we subtracted the variability associated with EOF 1 out of the data, what index time series would produce regression/correlation maps with the overall strongest amplitudes? That would turn out to be EOF 2, and so on.

Why is this useful? Going back to the Fourier analysis analogy, the data are more compactly characterized by EOFs. Large-scale variability will be in the low-order EOFs and the higher-order EOFs will have low-amplitude spatially-incoherent "noise". One application of EOFs is reconstruction of SST variations. EOF patterns are calculated for recent decades when SST data are available globally and used to transfer information from data-rich to data-poor regions in earlier times. I.e., if EOF analysis tells is SST is low in region A when is it high in region B, we can estimate variations in region A by knowing variations in region B (assuming SST variability is stationary).

Advantages of EOF analysis

- Produces the index time series which explains the greatest amount of variability (i.e., amplitude of regression map)
- Convenient method for characterizing dominant spatial pattern of variability
- Compact representation of data
- EOF patterns and time series are linearly independent

Disadvantages of EOF analysis

- Can be sensitive to choice of spatial domain and time period
- Features can be mixed between EOFs if their eigenvalues are similar and the degrees of freedom in the time series are too small.
- No guarantee the EOF pattern has physical meaning (i.e., EOF analysis can create patterns from “noise”)

Advantages of Simple Time Indices

- Simpler to construct and interpret
- Insensitive to details of statistical techniques

Disadvantages of Simple Time Indices

- Do not explain the most variance
- Not certain to characterize the dominant spatial pattern of variability
- Can be linearly dependent (i.e., different time indices do not uniquely contribute to variability)

In my own research I use both EOF analysis and simpler time indices depending on what provides the most advantages and least disadvantages for the particular situation. Often, EOF analysis can be used to identify the best way to construct a simple time index.