Summarizing space-time fields

Douglas Nychka
National Center for Atmospheric Research

Institute for Mathematics Applied to Geosciences

National Science Foundation *Math/Climate School, July 2010*

Outline

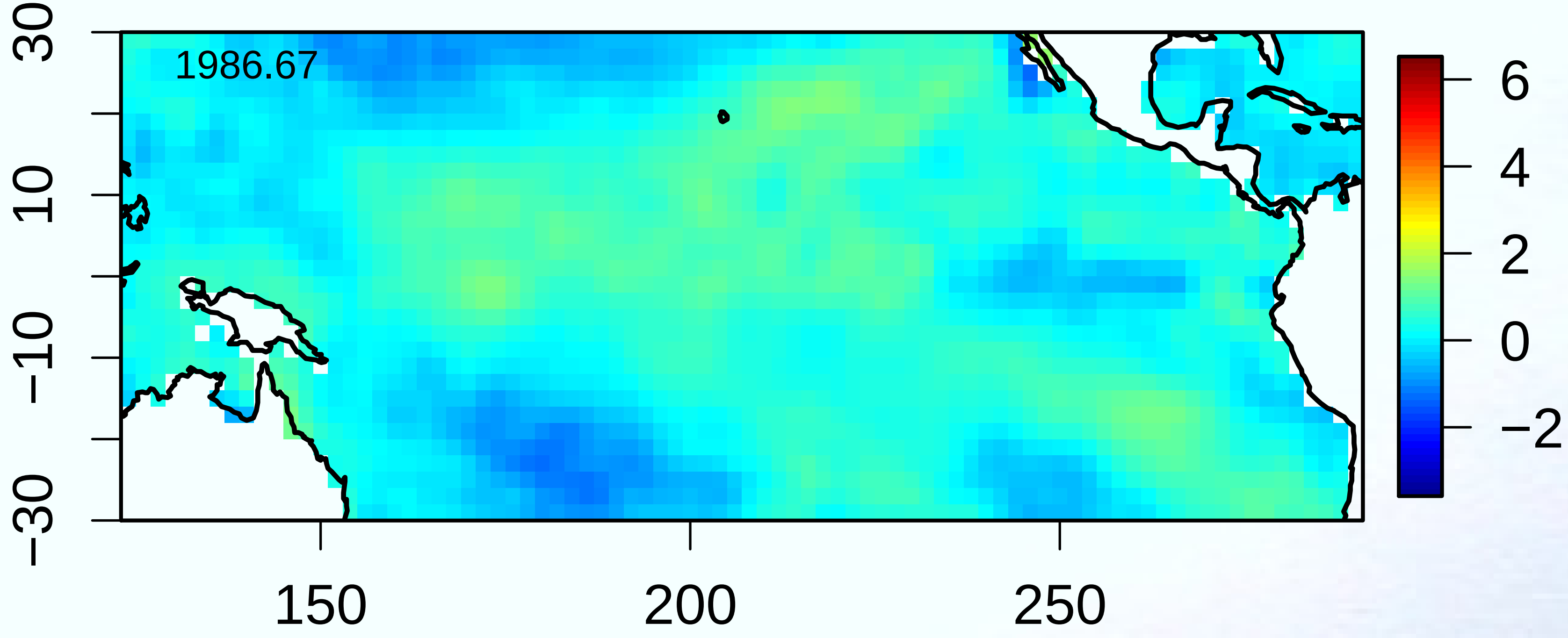
- Monthly SSTs
- The building blocks: basis functions and coefficients
- Where do the basis functions come from?
- Reconstructing a time slice
- Interpreting the principle modes.
- What can go wrong.

Big Idea:
Geophysical fields collected over time are difficult to visualize. EOFs can provide a few variables (coefficents) over time that summarize how the fields vary.

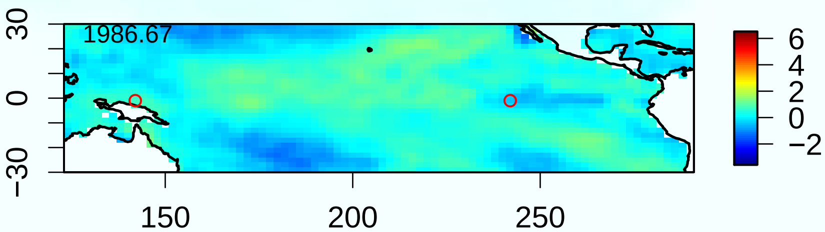
The mysterious aspect is where the basis functions come from …

Sea Surface Temperatures (SST)

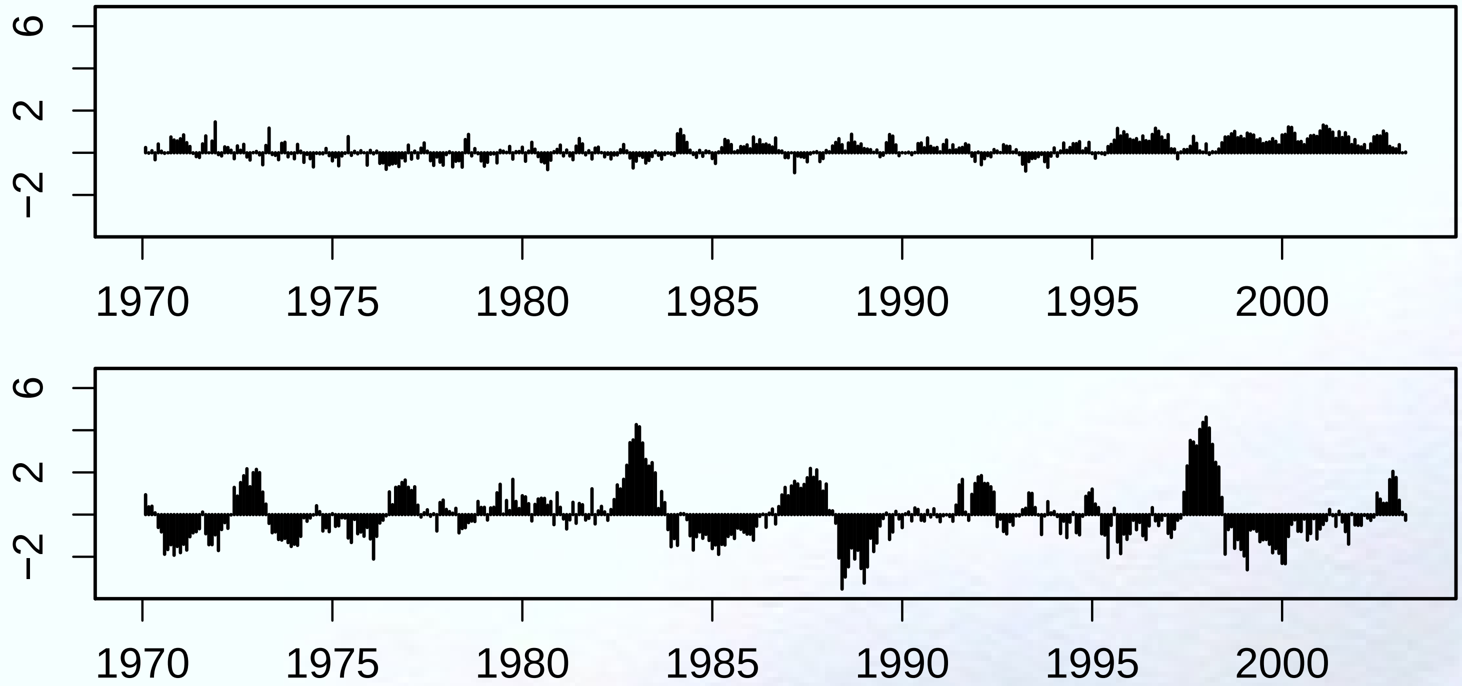
A snapshot of SST anomalies for a single month (JAN 1970)



Sea Surface Temperatures (SST)



Two monthly time series from two grid boxes



The Emipirical Orthogonal Function Decomposition

Two parts:

- basis functions – actually spatial fields
- coefficients for each time point – actually time series

Basis functions determined in an optimal (least squares) way from the same data. Thus they are empirical.

Coefficients found by least squares.

The problem

How we capture the dynamics seen in the animation using statistics?

Expand the fields in basis functions and coefficients

$$T(x,t) = \sum_j a_j(t)\phi_j(x)$$

If these were 1-d polynomials then

Temperature at location x and time t

$$= a_1 + a_2 x + a_3 x^2 + \ldots$$

• The a's are the coefficients (or amplitudes) and 1 x x^2 are the three basis functions.
• We want these for many time points – so a different set of a's for each time.

My quirky slide

The basis functions (EOFs) are building blocks:

- Individually they don't have to look like the geophysical process
- The fact they are orthogonal is not particularly important.
- Try to interpret them being combined with specific sizes coefficients.

EOFS: singular value decomposition

T is a big matrix , M rows are spatial locations and N columns are time.

$$T = UDV^T$$

U: $M \times N$ matrix with orthonormal columns, V: $M \times M$ orthogonal martix
and D: is a diagonal matrix with positive elements sorted from largest to smallest.

- M Columns of U are the EOFs.
- M Columns of VD are the time series of coefficients

Centering and scaling
It is usually a good idea to subtract off the mean – i.e. work with anomoalies. Standardizing is sometimes done but more difficult to interpret.

Another way using regression:

By a series of regressions
Find the single field that best explains the different times.

$$\min_{a,E} \sum_{t,x} (T(x,t) - a_t E(x))^2$$

$\longrightarrow$ *first basis function and coefficients*

1) regress the data on E – to get the updated a
2) regress data on normalized a to get the updated E.

Repeat until convergence.

$\longrightarrow$ *second basis function and coefficients.*
Subtract this estimate from the data and repeat algorithm.

Keep on going ... until there are as many basis functions as time points.

NOTE:
These basis functions will not be orthogonal and the coefficients are not the standard time series.

- Use the Gram-Schmidt technique to get an orthogonal set.

- Recompute all the coefficients by regressing the data on each EOF.

Basis functions for the SST data

We use a standard "EOF" decomposition to get these.

How much stuff should we look at?

Full EOF decomposition:

- as many basis functions as time points
- a time series of coefficients for each basis function.

This is as much information as the original data!

The decomposition is successful if we only have to look at a few basis functions and their time series of coefficients.

Strategy:

$$T(x,t) \approx a_1(t)EOF_1(x) + a_2(t)EOF_2(x)$$

Then we can interpret the whole data set by just two time series and looking at two basis functions. Typically the basis function have some spatial/geophysical interpretation.

What is $\approx$? We want the approximation to explain as much of the variance in the original data as possible.

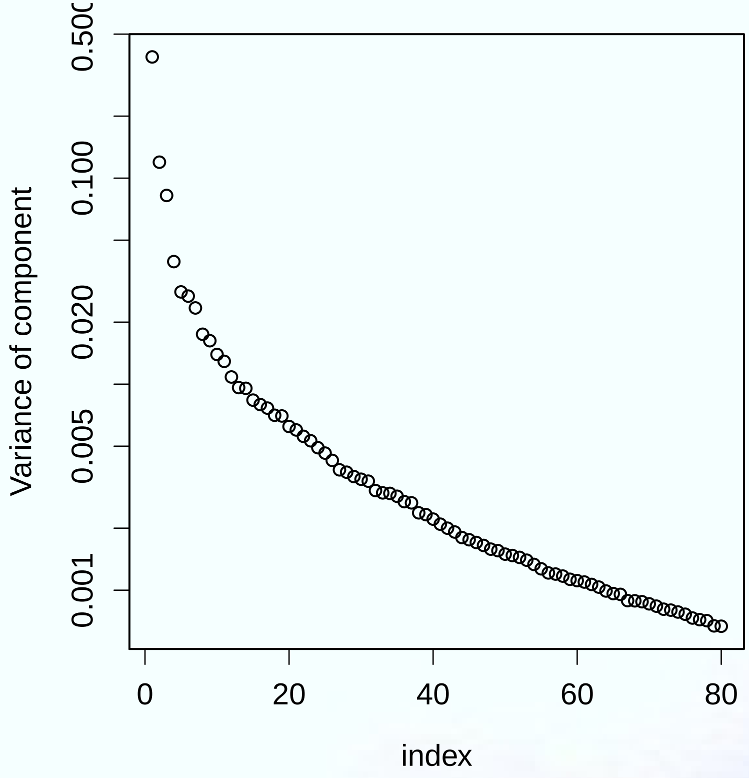
Usually if more than a few EOFs are required to approximate the data – we try something else!

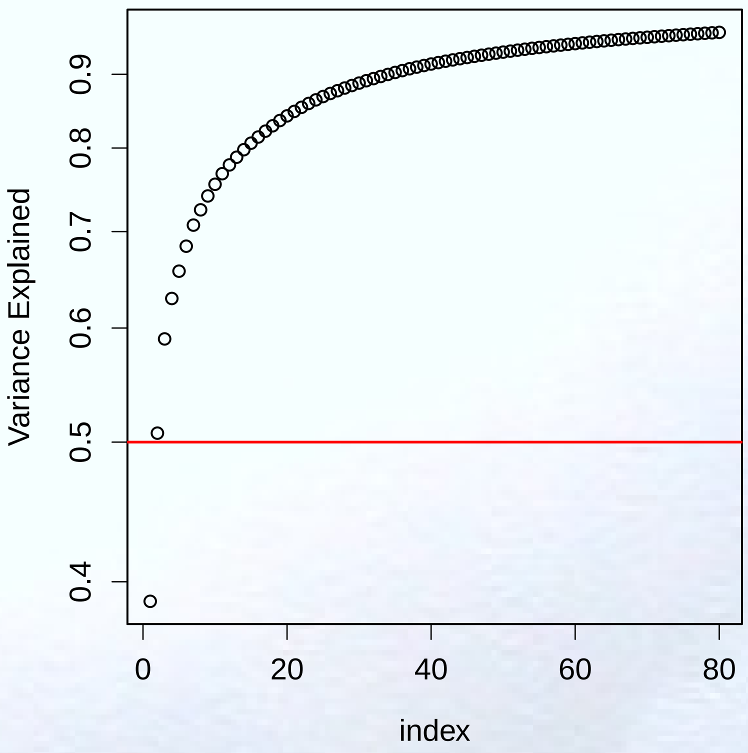
Simplicity vs. accuracy

How much of the data is explained by the EOFs?

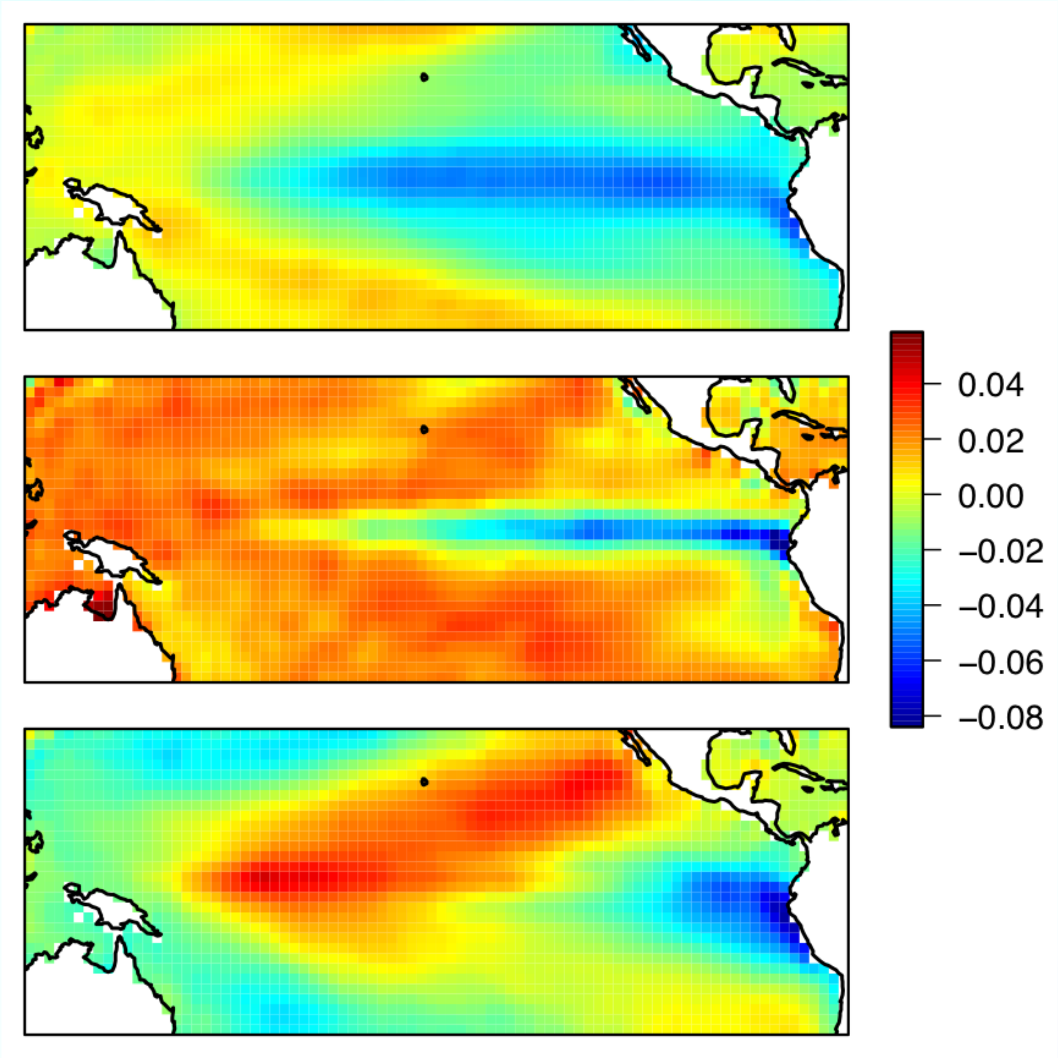
EOFs are ordered by their ability to explain the data – usually from the best EOF to the less important.

Log variance explained individually and cumulatively





Focus on first three EOFs



EOF1 The classic ENSO pattern.

EOF2 A modulation of the center.

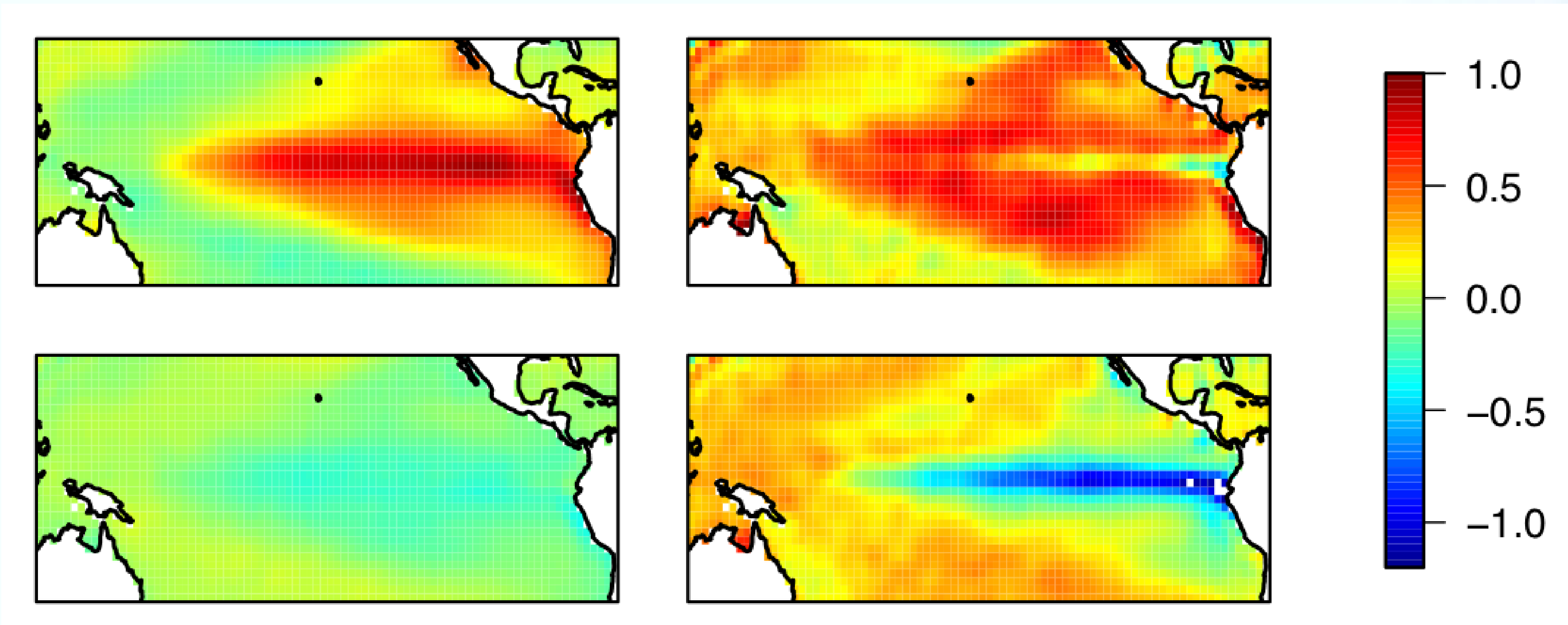
EOF3 Messing with the coast of SA and the Northern Pacific.

Note: All the EOFs are scaled to have same range (MSE).

How EOF 1 and 2 interact

Four different cases of adding together EOF1 an EOF2. Using "small" and "large" coefficients.

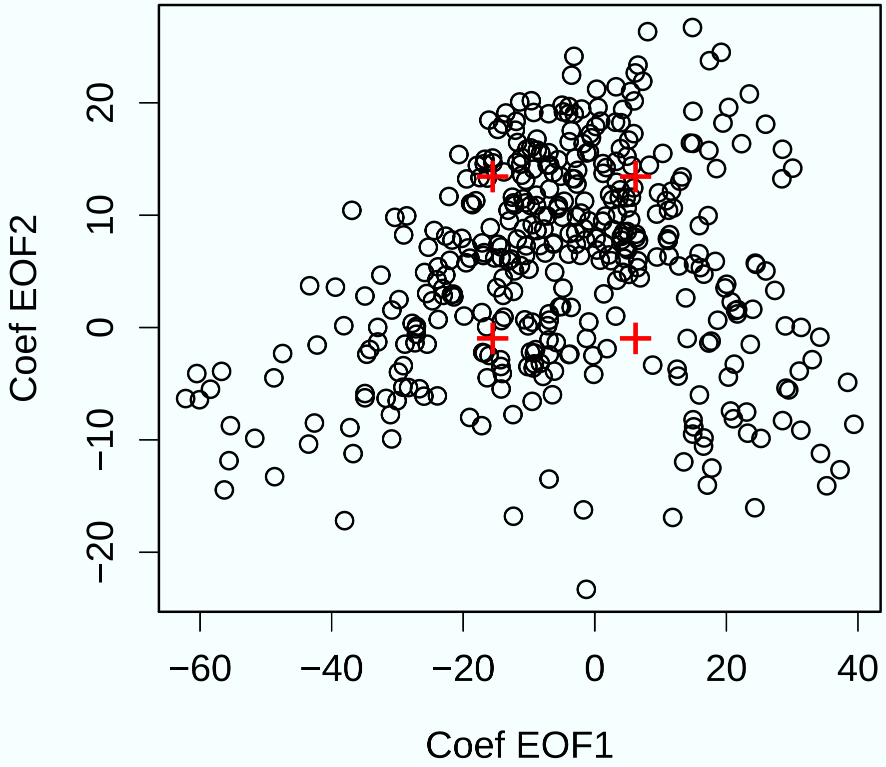
Small/ Small   Small/Large



Large/ Small   Large/Large

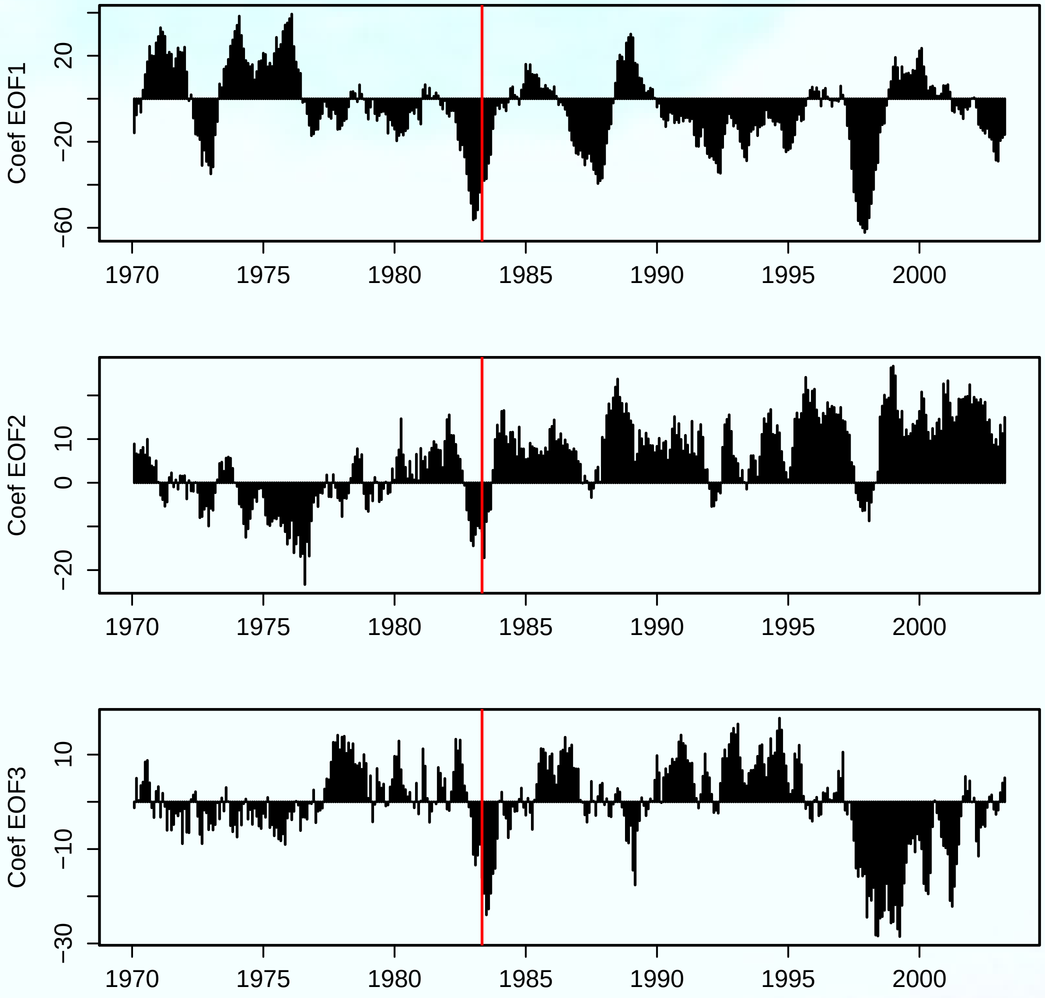
Coefficients for EOF 1 and 2

A pair of coefficients for all 399 monthly time points



four cases from last panel

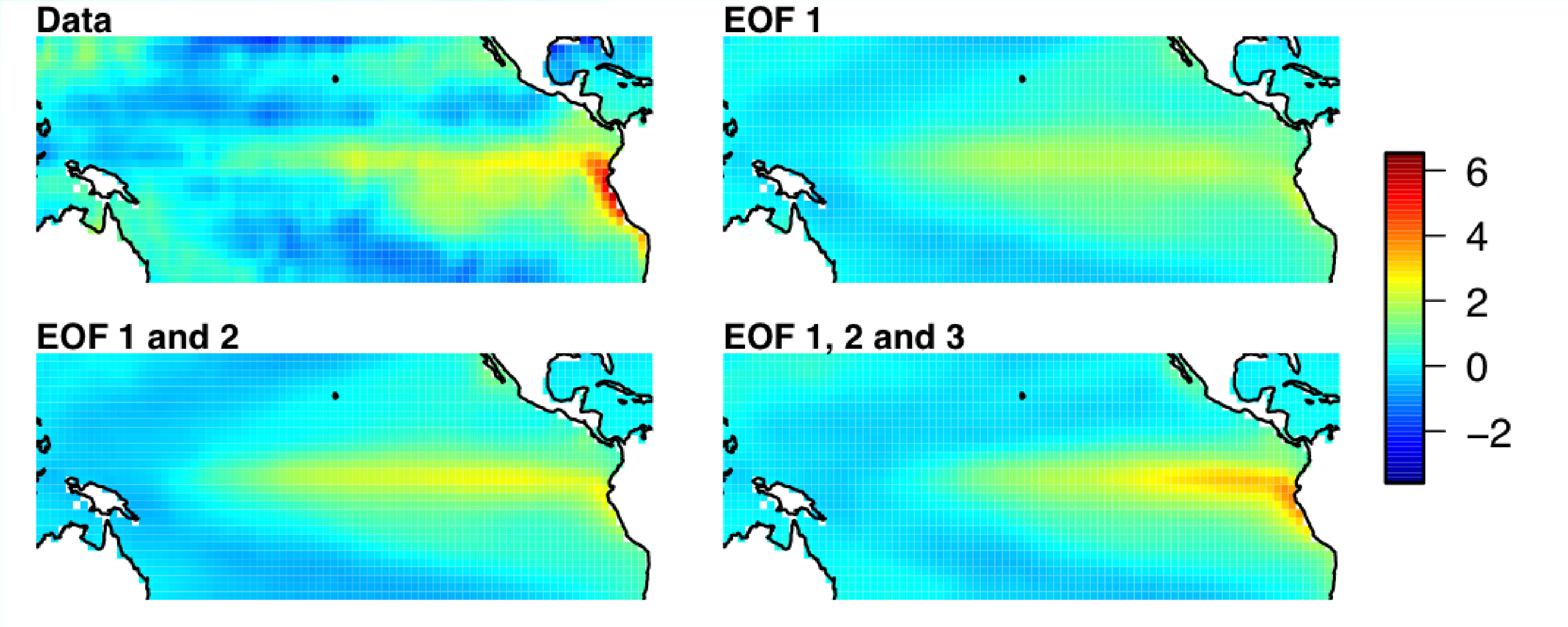
Coefficients over time



Interaction of the three coefficients suggests how pattern changes.

How good is it?

An example in 1983



EOF3 seems to add important features for this time but the match is not perfect.

What can go wrong.

Nature does feel limited to a few basis functions!
– and probably does not care about orthogonality either.

• Features such as a traveling pulse may not be captured by a few EOFs

• Basis functions typically selected based on mean squared error. The MSE criterion can be deceptive in capturing features.

• EOFs are sensitive to outliers in the data and also to the exact spatial domain chosen.

• Seeing an illustration of EOFs where they work well and expecting this method to always be successful!

Thank you!