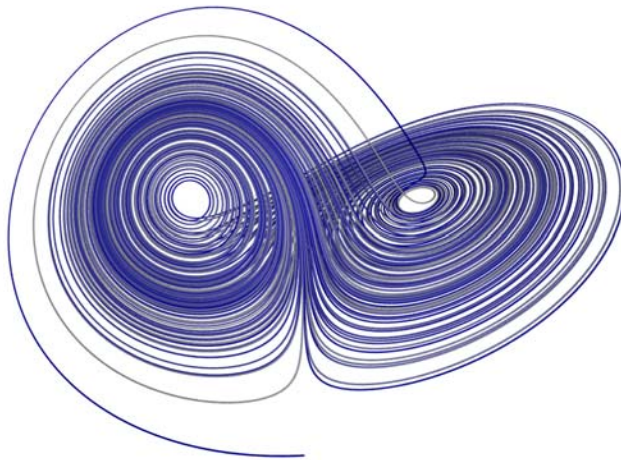


## NetCDF, mapping and analysis of space/time variance

### How can we define climate?

- Recall week 3 Lorenz attractor, and week 10 QG model



## Data decomposition

Consider spectral model:

$$Q(x,y,t) = \sum q_m(t) Y_m(x,y)$$

( $Y_m$  are spherical harmonics in longitude and “mu”)

More generally some variable “Z”

$$Z(\mathbf{x},t) = \sum a(t) E(\mathbf{x})$$

We are free to choose the basis functions (E) and derive the accompanying amplitudes (a).

*We will find E are “EOFs” and a are the principal components*

## Consider correlation

- Wish to choose basis that maximizes variance along primary axis.
- Can show that this is just the eigenvectors of the correlation/covariance matrix.
- $EC = \lambda C$  where  $C = Z^T Z$
- E are the eigenvectors (EOFs), and  $\lambda$  the eigenvalues. Notice  $\lambda^2$  is proportional to variance!

# Recipe for PCA

Eigenvector method ( $EC = \lambda C$ )

- Get some time space data
- Remove time mean
- Compute correlation (“scatter”) matrix
- Solve eigenvalue problem

SVD method ( $Z = USV^T$ )

*Equivalent!*

- Get some time space data
- Remove time mean
- Perform SVD

Notice S are singular values (like eigenvalues), and a property of SVD is that V is made of columns of eigenvectors of  $Z^T Z$ , and U the eigenvectors of  $Z Z^T$

# Lab assignment

- Perform principal component analysis of monthly mean data from 1981-1990.
- Choose some region of the earth (e.g., North America, polar regions, tropics ...)
- Make plots of:
  1. Percentage variance explained by top few (say 10) PCs
  2. Amplitude versus time of first 3 PCs
  3. Maps of first 2 or 3 EOFs

# The data

- We will use data from the NCEP reanalysis.
- This is monthly mean data from 1981-1990 (thus a total of 120 months)
- See NetCDF file  
~dcn/ATOC7500/project/data/ncepmonthly.nc
- You do not need to copy this data, read it from this directory

- A unix program **ncdump** shows “contents” of a NetCDF file
- (tip give `-h` option to see only header)
- `% ncdump -h ncepmonthly.nc`

```
netcdf ncepmonthly {
dimensions:
    lon = 64 ;
    lat = 32 ;
    lev = 3 ;
time = UNLIMITED ; // (120 currently)
variables:
    float lon(lon) ;
        lon:long_name = "Longitude" ;
        lon:units = "degrees_east" ;
    float lat(lat) ;
        lat:long_name = "Latitude" ;
        lat:units = "degrees_north" ;
    double time(time) ;
        time:long_name = "Time" ;
        time:units = "months since 01-01-1981 00:00:00" ;
    float Z(time, lev, lat, lon) ;
        Z:long_name = "Geopotential height" ;
        Z:units = "m" ;
    MORE VARIABLES HERE
}
```

Dimensions

Coordinate data

Variables

# How to read data in IDL

NetCDF files are binary, and use “ID” codes to locate the data.

- Open the file, and return file ID  
`ncid = ncdf_open(filename)`
- Get the ID of the variable  
`varid = ncdf_varid(ncid, 'Z')`
- Read the data  
`ncdf_getvar(ncid, varid, zg)`
- Close file  
`ncdf_close,ncid`
  
- Hint: you need to know the variable name ‘Z’ first.  
Get this from `ncdump!`
- [http://idlastro.gsfc.nasa.gov/idl\\_html\\_help/NetCDF-routines10.html](http://idlastro.gsfc.nasa.gov/idl_html_help/NetCDF-routines10.html)

# Making map projection plots

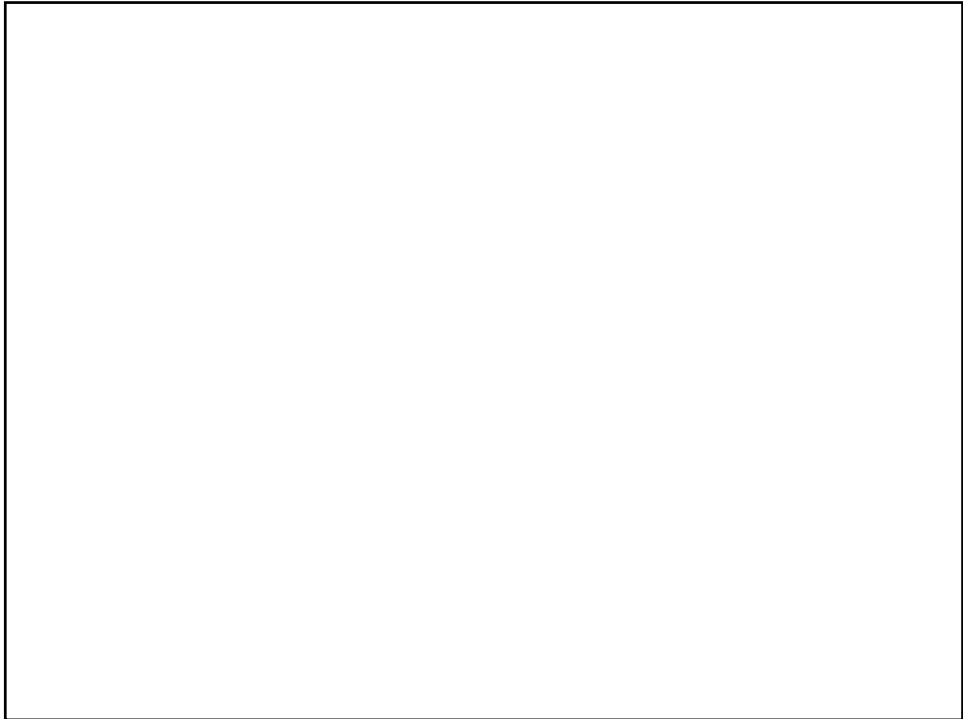
- IDL does the work of figuring out the projections.
- `map_set,lat0,lon0,projection_type`
- E.g.,  
`map_set,30,300,/orthographic,/isotropic,/horizon  
contour, something  
map_continents`
  
- `map_set` has many options to play with
- Notice: this is troublesome with `!p.multi`
- [http://idlastro.gsfc.nasa.gov/idl\\_html\\_help/projections2.html](http://idlastro.gsfc.nasa.gov/idl_html_help/projections2.html)

# Lab assignment

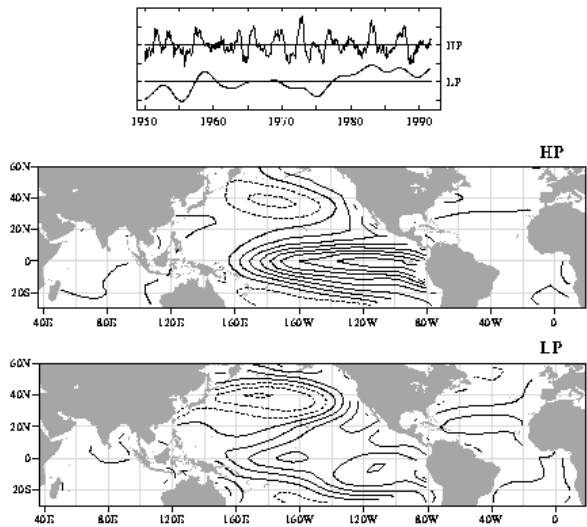
- Perform principal component analysis of monthly mean data from 1981-1990.
- Choose some region of the earth (e.g., North America, polar regions, tropics ...)
- Make plots of:
  1. Percentage variance explained by top few (say 10) PCs
  2. Amplitude versus time of first 3 PCs
  3. Maps of first 3 EOFs
- *Output as images, and upload to Wiki*
- *Modify PCA code to area weight the data to avoid area bias (i.e., cosine of latitude). Compare results.*
- Note: 2d “space” could be long-lat, lat-lev, or lon-lev!

# Key tasks/lessons for today

- Interrogate NetCDF files (ncdump)
- Read NetCDF files (in IDL)
- Subset the data in IDL
- Perform PCA
- Contour plots with map projections
  
- See NetCDF file  
~dcn/ATOC7500/project/data/ncepmonthly.nc
- PCA routine and “stub” (pcanal.pro, pcastub.pro)  
in ~dcn/ATOC7500/week12

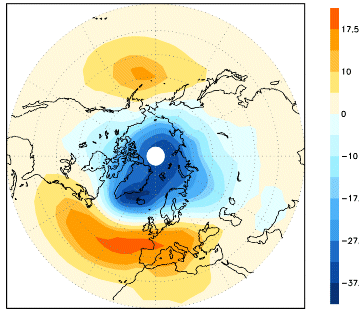


## EOF of tropical sea surface temperature



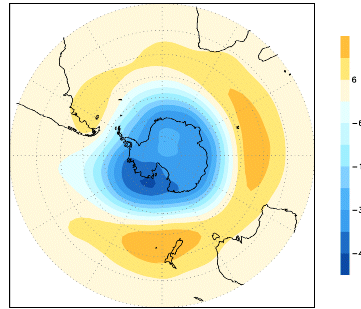
# First EOF of 500 hPa height “Annual modes”

The Northern Hemisphere annular mode



The surface signature of the Northern Hemisphere annular mode. The NAM is defined here as the leading EOF of NH monthly-mean 1000-hPa height anomalies. Units are m/std of the principal component time series.

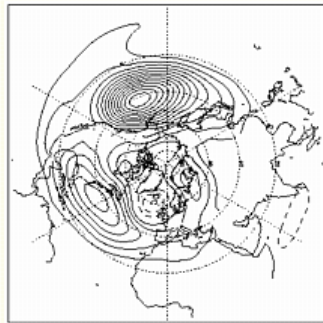
The Southern Hemisphere annular mode



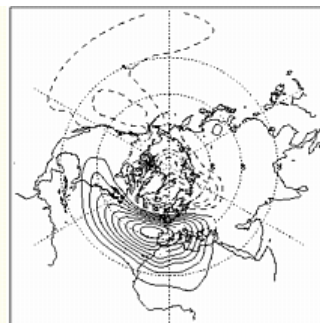
The surface signature of the Southern Hemisphere annular mode. SAM is defined here as the leading EOF of SH monthly-mean 350-hPa height anomalies. Units are m/std of the principal component time series.

Thompson and Wallace, 2001: *Science*, **293**, 85-89.

## Others! 850 hPa streamfunction



EOF 1 - PNA  
Pacific North American pattern



EOF 2 - NAO  
North Atlantic Oscillation

*Are they real, or statistical artifacts?*

# Summary

- Notice the NCEP data file has been interpolate to 64x32 grid and 3 layers of our climate model. This may be useful for validation
- See IDL plotting script:
- ~dcn/ATOC7500/project