ESS/ATMOS/OCEAN 589, Spring 2009

Principal Component Analysis and Statistical Climate Reconstruction
PROBLEM SET:

Get the file ssta_short.mat from
courseswashington.edu/pcc589/2009/notes/ProblemSet6

This is a matrix of sea surface temperatures (SST) from 1950 through 1997. They are monthly data, with the seasonal cycle removed ('monthly anomalies').

You will also want to download the other m-files there:
sst_eofs and three_two.m, and the file 'predictors'.

You can use three_two.m as follows to take this three dimensional matrix (time, latitude, longitude) and make it two dimensional (location,time):

```
[data2,index] = three_two(data3);
```

This function also gets rid of the places where there is land (no data there).

data2 is the matrix you want to do eof calculations on.

If you want to look at the data, remap it to the Lat Long grid and view it using pcolor.

```
map = ones(19,36)*NaN; % which creates an empty matrix of the right
                       % size and then
map(index) = data2;  % where you could replace data2 with e.g. your
                     % reconstruction of the data
```

pcolor(Lon,Lat,map) will make a pretty map, with the continents all in white.

To make life easy, most of the first part of the problem set is already solved:

run sst_eofs

Figure 1 shows the 4 EOFs of sea surface temperature, based just on North Pacific variations (20 N to 60 N, 120 W to 280 W).

Figure 2 shows the first two normalized PCs.

Figure 3 shows the cumulative and individual variances of the eigenvalues.

Figure 4 shows the projection (sst*PC1'/N) of the first principal component on the entire data field.

PROBLEMS:

1) Do a spectral analysis on the first PC, which nominally represents the "Pacific Decadal Oscillation." Is this actually oscillatory or decadal? Explain what you mean.

2) Determine and plot the uncertainty in each eigenvalue (lamda), using the formulation given in David's EOF notes: lamda*(2/Neff). A good way to calculate the effective degrees of freedom is Neff = N*(1-r)/(1+r) where r is the lag-one-autocorrelation coefficient of each PC. You can make a nice plot using "errorbar" in place of "plot" in matlab. errorbar(x,y,E) plots y with error bars [y-E, y+E]. Comment on the significance of the eigenvalues (and therefore the associated EOFs). Are the higher-order EOFs meaningful?

3) Edit the sst_eof code so that it calculates the EOFs based only on Tropical Pacific SSTs (20 N to 20 S, 120 W to 280 W). You are capturing nominally the "El Nino" or ENSO variability this way. Redo the calculations and questions in (1) and (2). Also, project the first PC onto the entire data field. Is there any similarity to the "PDO" pattern?

4) Now, use the proxy variables in the file 'predictors' to reconstruct the Tropical SST field, back to 1900. ('predictors' has data from 1900 through 1997, ssta_short has data from 1950 to 1997, so the last 564 rows of 'predcitors' overlaps in time with the SST data.) Be thoughtful in how many PCs you choose to retain. Be similarly thoughtful about how many predictors to use. (Are any of the predictors useful? Useless? How do you known in advance?) Explain your reasoning in using all the predictors or just some of them.

5) Optional but hopefully fun:

Download the full SST data set (sst_3_full) to see how well you did. Comment, based on correlation coefficients, on the appearance of individual reconstructed months, etc. Choose whatever metric you like for this. I'm just looking for a qualitative assessment.