ESS/ATMOS/OCEAN 589, Spring 2009

Notes:
Principal Component Analysis and Statistical Climate Reconstruction

A standard method for doing statistical climate field reconstruction is a simple three-step process:

1) break down the data field into its principal "patterns" or "modes" of variability, usually called the (spatial) empirical orthogonal functions (EOFs), and the associated time varying principal components (PCs),

If the data are arranged as a matrix M*N where M rows are the locations, and N columns are the spaces, then a simple calculation of the eofs is as follows (in Matlab):

```
c=data*data'/N; % covariance matrix
[v lamda]=eig(c); % get the EOFs (eigenvectors) (v) and
eigenvalues
                  % (lamda)
v = fliplr(v); % sort them so that the most important is number
one
               % (that is, v(:,1);

lamda = fliplr(flipud(lamda)); % sort the eigenvalues too
Lamda=diag(lamda); Lamda=abs(Lamda); %'Lamba' is the diagononal
of the matrix 'lamda'

p = v' * data; % Project the eigenvectors onto the data, to
obtain the
               % principal components.

% Finally, scale the PCs and EOFs so that the PCs have unit
variance,
% and the EOFs have the same units as the original data (e.g. for
SST, % this would be the number of degrees C of SST that amounts
to a  one
% standard deviation anomaly)

for i = 1:M
    v(:,i)=v(:,i)*sqrt(Lamda(i));
    p(i,:)=p(i,:)/std(p(i,:)');
end
```

2) Assume that the patterns of variability are fixed in space (e.g. El Nino always has more warming in the Eastern Pacific than the Western

Pacific). Find some way to extend the length of the principal components back in time. A typical way to do this by least squares regression between predictor variables (e.g. some tree ring data) and the principle components.

To make this calculation easy, we put everything in matrix form:

P'X = Y'

where P is our matrix of principal components, Y is our matrix of proxy variables (during the overlap time period with P), and X is a weighting matrix, that tells us how the two are related. Note that the prime (') denotes the transpose (which flips the matrix on its side): we have flipped P and Y so time is the vertical dimension so that the matrix multiplcation works. If there are three principal components we want to reconstruct, then we need at least 4 proxy variables so that the problem is overdetermined. If we happen to have 5, then the matrices look like this:

$$\begin{bmatrix} p11 & p12 & p13 \\ p21 & p22 & p23 \\ p31 & p32 & p33 \\ p41 & p42 & p43 \\ p51 & p52 & p53 \\ p61 & p62 & p63 \\ p71 & p72 & p73 \end{bmatrix} \begin{bmatrix} X11 & X12 & X13 & X14 & X15 \\ X21 & X22 & X23 & X24 & X25 \\ X31 & X32 & X33 & X34 & X35 \end{bmatrix} = \begin{bmatrix} y11 & y12 & y13 & y14 & y15 \\ y21 & y22 & y23 & y24 & y25 \\ & & & & \\ & etc. & & & \\ & & & & \\ & & & & \\ & & & & y75 \end{bmatrix}$$

Matlab has a superb way to solve this kind of problem for X

Just 'back divide':

P'X = Y'
X = P'\Y'

Note that this is **not** the same as doing Y/P, which in matrix multiplication terms can't actually be done. Read "help mldivide" in matlab to learn more about this function.

A very nice property of matrices is that in general, for two matrices A and B, we can write:

AB = (B'A')'

so if P'X = Y' then X'P = Y and so

P = X'\Y

Note that there is no limit on the N dimension of the proxy matrix, Y, once we know X.

"Inner matrix dimensions" must agree, so in the above example

[N*3][3*5]=[N*5] and N can be anything.

This means that is we have a longer set of proxy variables, Ylong, then

Plong = X'\Ylong

allows us to extend the length of P (the principle components).

3) Reconstruct the data by combining the PCs and EOFs.
reconstruction = v(:,1:R)*Plong(1,R,:))

where "R" is the number of PCs we choose to retain.