Appendix I

Names of 36 Subdivisions of India

1. Andaman & Nicobar Islands
2. Arunachal Pradesh
3. Assam &Meghalaya
4. Nagaland, Manipur, Mizoram & Tripura
5. Sub-Himalayan West Bengal & Sikkim
6. Gangetic West Bengal
7. Orissa
8. Jharkhand
9. Bihar
10. East Uttar Pradesh
11. West Uttar Pradesh
12. Uttarakhand
13. Haryana, Chd & Delhi
14. Punjab
15. Himachal Pradesh
16. Jammu & Kashmir
17. West Rajasthan
18. East Rajasthan
19. West Madhya Pradesh
20. East Madhya Pradesh
21. Gujarat
22. Saurashtra
23. Kutch
24. Madhya Maharashtra
25. Marathwada
26. Vidharbha
27. Chhattisgarh
28. Coastal Andhra Pradesh
29. Telangana
30. Rayalaseema
31. Tamilnadu & Pondi
32. Coastal Karnataka
33. North Interior Karnataka
34. South Interior Karnataka
35. Kerala
36. Lakshadweep
Appendix II

The Statistical Techniques used in CPT

The detail the techniques used in CPT for Principal Components Analysis (PCA) or Empirical Orthogonal Functions (EOF) and CCA. PCA or EOF is a technique used to perform data decomposition of an input data set. The input data may consist of a single variable, such as a field of gridded SSTs, or it can be a pooling of more than one field, of one field at two or more staggered times, or both types of combination together. It reduces the dimensionality of the data set, typically with a large number of interrelated variables, into principal components (PCs, or EOFs). A few PCs may often reproduce much of the variation present in the original data set. Each component is some weighted combination of the original variables—for example, 0.34 times variable 1, plus 0.51 time variable 2, minus 0.17 times variable 3, and so forth through all of the original variables. The weights are determined using the matrix of temporal covariances (or correlations) among all pairs of original variables. The components are thus efficient summaries of a large volume of data. A component involves a specific pattern of the weights that form the weighted combination (called the loading pattern). It also has an associated time series indicating the degree of presence of the loading pattern in the original data (called the principal component, temporal score, or amplitude). Thirdly, it includes a number representing the component’s proportion of the total variance of the original data set that it accounts for. The principal components have the following properties:

- They explain the maximum amount of variance represented in the original data, so that the several largest components together can often represent over half of then original variance.
- They are uncorrelated, which maximises both efficiency of usage of variance and stability of regression coefficients when their time series are used as predictors.
- Retaining only a few large components minimises the problem of multiplicity.
Both principal components regression (PCR) and canonical correlation analysis (CCA) first find the dominant modes, i.e. the PCs, of the predictor (X) variable by decomposing the data and reducing its dimensionality. In CCA this is done for the predictand (Y) also. Each of the dominant EOF modes, describable by spatial loading patterns spanning the original field, are used to create a time series for each EOF mode. In other words, each case in the data history (e.g. a single year out of the set of all years available) is given a score that indicates the degree to which it resembled the spatial loading pattern for the mode in question. These time series (temporal scores or amplitudes) show how similar the pattern for a given year was to the pattern represented by the mode (score would be positive, and magnitude would indicate degree of similarity), how dissimilar or opposite it was (score would be negative, and magnitude would indicate degree of dissimilarity), or how little relationship it bore to the pattern of the mode (score would be close to zero).

EOF analysis or PCA can best be understood if we begin with anomalies—i.e. departures from the average formed by subtracting the time-mean from each element of the time series in the original input data. A common way to determine the EOFs is by computing the eigenvalues and eigenvectors of a variance/covariance matrix of the data set. (If, after subtracting the mean from the original data, one also divides by the standard deviation so that the data are standardized, then the covariance matrix is also the correlation matrix. More will be said below about the decision of whether or not to standardize the data.) Assume an anomaly matrix $A$, where the rows represent time, increasing from top to bottom, and the columns represent the variable element number, increasing from 1 to $n$ from left to right. Then the covariance matrix $R$ of $A$ is

$$R = A^T A \quad (1)$$

where $(\cdot)^T$ denotes transpose, where the rows and columns of the matrix reverse their roles. In Equation (1), the product of the anomaly of each variable with each other variable (and with itself) is computed for each time, and accumulated over all the times. Division by the number of times has been omitted in (1) but is assumed. The resulting square matrix $R$ contains the covariance of each variable
with every other variable, and the diagonal of \( \mathbf{R} \) contains the variance of each of
the variables. \( \mathbf{R} \) is then used as input to an eigenvalue/eigenvector calculation.
The eigenvalue/eigenvector problem is set up as follows:

\[
\mathbf{RC} = \mathbf{C} \Lambda
\]  

(2)

where \( \Lambda \) is a diagonal matrix containing the eigenvalues \( \lambda_i \) of \( \mathbf{R} \) where \( i \) runs
from 1 to the number of variables, \( n \). (Note that \( \lambda \) is the lower case of symbol \( \Lambda \),
used to denote a scalar element along the diagonal of matrix \( \Lambda \).) The eigenvalues
in matrix \( \Lambda \) will be in descending order of magnitude. The column vectors \( \mathbf{c}_i \) of
\( \mathbf{C} \) are the eigenvectors of \( \mathbf{R} \) corresponding to the eigenvalues \( \lambda_i \). The \( \Lambda \) and \( \mathbf{R} \)
matrices are square, and are the same size, \( n \times n \). While the index \( i \) initially
denoted the variable number, upon solving equation 2 it also denotes the mode
number for the eigenvalues and associated eigenvectors resulting from the EOF
analysis. Thus, there are potentially as many EOF modes as there are original
variables, but we will see that we will be able to discard all of them except for
the first few. The eigenvectors \( \mathbf{c}_i \) corresponding to the eigenvalues \( \lambda_i \) are the
loading patterns, ordered by eigenvalue size. These eigenvectors are the EOFs
(for our purposes, spatial patterns) that we seek to understand physically, and are
sometimes considered as maps because they can be shown using the same
geographical background as that used for the original predictor data. The pattern
described by the first EOF (i.e., \( \mathbf{c}_i \)) is therefore associated with the largest
eigenvalue, that described by the second EOF the second biggest eigenvalue,
etc., where the eigenvalues give a measure of the fraction of the total variance
in \( \mathbf{R} \) explained by the respective EOF. This fraction is found by dividing an \( \lambda_i \) by
the sum of all the eigenvalues (i.e. the trace of \( \Lambda \)). The eigenvector matrix \( \mathbf{C} \) has
the property \( \mathbf{C}^T \mathbf{C} = \mathbf{C} \mathbf{C}^T = \mathbf{I} \), where \( \mathbf{I} \) is the identity matrix—i.e. all diagonal
elements are unity and all off-diagonal elements are zero—which means that the
EOFs are spatially uncorrelated or orthogonal to each other. Multiplying the
original anomaly matrix \( \mathbf{A} \) by the corresponding eigenvector gives the projection
or transformation function \( \mathbf{b}_i \) of \( \mathbf{A} \) on \( \mathbf{c}_i \) which is the time series showing the
case-to-case variation of each \( \mathbf{c}_i \):
The \( b_i = A \cdot c_i \). \hspace{1cm} (3)

The \( b_i \) are the principal component coefficients (PCs), or amplitudes or temporal scores of the EOFs. In matrix form, encompassing all values of \( i \), we have \( B = AC \). Note that postmultiplying both sides by \( C^T \) gives \( BC^T = ACC^T = A \). The scores for the first few (say \( p \)) modes can be transformed back to partially reconstitute \( A \) by

\[
A = \sum_{i=1}^{p} b_i (c_i), \quad 1 \leq p \leq n, \text{ the dimension of the original matrix.}
\]

There are as many eigenvalues (each with an associated loading pattern and a temporal series) as there are original variables. If all of them are used, then the reconstitution of \( A \) using the formula above would be exact. However, only a relatively small subset of the first \( p \) EOFs, having the largest eigenvalues, are typically used.

We now show in greater detail how we solve for \( \Lambda \) and for \( C \) in (2): \( RC = \Lambda C \).

Moving \( C \Lambda \) to the left and using scalar expressions for an individual mode \( i \), we get \( (R - \lambda_i I)c_i = 0 \). This equation for \( c_i \) has a nonzero solution when \( (R - \lambda_i I)c_i \) is singular, i.e. when the determinant of \( (R - \lambda_i I) \) is 0. It is a system of \( n \) equations in \( n \) unknowns. The determinant solution for the \( \lambda_i \), containing the variance/covariance coefficient matrix \( R \), is

\[
\det (R - \lambda I) = 0 \quad \text{i.e.,} \quad \det \begin{pmatrix}
\lambda_1 - \hat{\lambda} & r_{12} & \cdots & r_{1n} \\
r_{21} & \lambda_2 - \hat{\lambda} & \cdots & r_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
r_{n1} & r_{n2} & \cdots & \lambda_n - \hat{\lambda}
\end{pmatrix} = 0
\]

The above determinant expands into a polynomial of degree \( n \) in the variable \( \lambda \). In solving the equation, we find \( n \) roots for \( \lambda \), these roots being the eigenvalues. Once the \( \lambda_i \) are found, then for each individual root \( \lambda_i \) we solve the equation \( R C_i = \lambda_i C_i \), where \( C_i \) is the eigenvector associated with eigenvalue \( \lambda_i \).
We do not include the zero vector, despite that it satisfies the equation, since it has no structure and therefore no meaning to us.

The temporal scores for each of the two modes are determined by multiplying the original anomaly matrix $A$ by the eigenvector corresponding to the mode in question, as per (3): $b_i = A c_i$. For mode 1, for example, with $c_1 = c_2 = 1$, cases in which both $c$ elements have positive anomalies would have a positive temporal score, both $c$ elements are negative would have a negative score, and when they are mixed the sign of the element having greater magnitude would “win” in determining the sign of the temporal score. Since both elements have a loading of 1, it does not matter for the score which element has which anomaly, as they could be reversed without changing the score. This feature is unique to this very simple example; generally, the pattern of the anomalies does matter to the temporal score. Note that the polarity of the mode can be reversed without changing the meaning or interpretation of the mode. For example, the eigenvector associated with mode 1 could just as easily have been written $c_1 = c_2 = -1$ instead of $c_1 = c_2 = 1$. The temporal scores for the mode would then all be of sign opposite to what we had originally, but the sequence of variations would remain intact as a mirror image of the original sequence and the original input anomalies for each case would still be properly reproducible. For the second mode, temporal scores will have highest magnitude in cases when one element has the opposite anomaly sign of the other (and the higher the magnitudes, the higher the score), and the sign of the score depends on which element has which anomaly sign.

In the above example, there were two original variables, and the eigenvalues, eigenvectors and temporal scores were computed for two modes—the maximum possible, as the number of modes possible is the number of original variables. When all possible modes are calculated, the original data set can be perfectly reproduced from its decomposition. What is atypical about the given example was that all possible modes were computed, because the original number of variables was designed to be only two to illustrate a by-hand EOF analysis. Usually the original data set consists of a much larger number of variables, of which only a few of the most dominant modes are used. These may reproduce a moderate portion, but not all, of the original data. What is perhaps more important is that even on occasions when only a small to medium
proportion of the original variability is represented in the first few retained modes, this represents a large proportion of the coherent or physically meaningful variability, the remaining variability often being largely random, representing “noise”—unsystematic or internal variability of the ocean or atmosphere that cannot be harvested for predictions.

Some more matrix algebra helps illustrate the decomposition of the covariance matrix, $R$, into its components. Since $R$ is quadratic, and $C^T C = C C^T = I$, it follows that $C$ has the property $C^{-1} = C^T$ such that equation (3), $b_i = A c_i$, can be written as $R = C \Lambda C^T$. $R$ can also be decomposed in the form of eigenvalues and eigenvectors, as follows:

$$R = \Lambda_1 c_1 c_1^T + \Lambda_2 c_2 c_2^T + \ldots + \Lambda_p c_p c_p^T$$  \hspace{1cm} (4)

where $\Lambda_1/\Sigma \Lambda_i$ is the fractional variance associated with the first EOF. Since just a few EOFs often account for more than half of the total variance, it means that the data matrix can be explained by a few basis vectors which is the same as reducing the data to a few different modes of variability. This means that if we have a maximum of $n$ eigenvalues, then we can have a $p$-dimensional space (where $p \ll n$) that reproduces much of the data set, each EOF representing a single independent component mode of variability. In addition to the technique of computing principal components analysis or EOFs shown above, they can also be computed in another way, using singular value decomposition (SVD). As will be discussed below, SVD is a technique quite similar to CCA in that it relates patterns of a predictor pattern to patterns of a predictand pattern. The SVD of matrix $A$ is $A = U \Sigma V^T$ where $U$ and $V$ are orthogonal matrices and the $\Sigma$ matrix is diagonal. Then

$$R = A^T A = V \Sigma^T U^T V \Sigma V^T = V \Sigma^T \Sigma V^T.$$  

Note that $RV = V \Sigma^T \Sigma$. Recalling from equation (2) above that $RC = C \Lambda$, we can identify $C$ with $V$ and $\Lambda$ with $\Sigma^T \Sigma$, and $B = AC = U \Sigma V^T V = U \Sigma$. Conceptually, SVD is an operation in which the covariances between one field and another field are computed as covariances between each element of one field and each element of the other field, but not among elements of the same field. However,
when using SVD to compute EOFs, the two fields are identical, both being the field for which EOFs are to be computed. In effect, the covariances become those among the elements of the one field, and the “covariance” of an element with itself is its variance. The elements of the matrix $\Sigma$ reflect the relative proportions of variance accounted for each of the resulting vectors, representing the eigenvalues associated with their corresponding eigenvectors in the direct computation method described earlier.