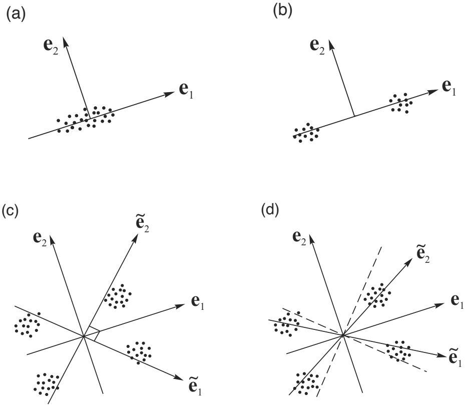
2.13 Rotated principal component analysis [Book, Sect. 2.2]

Fig.: PCA applied to a dataset composed of (a) 1 cluster, (b) 2 clusters, (c) and (d) 4 clusters. In (c), an orthonormal rotation and (d) an oblique rotation gave rotated eigenvectors $\tilde{\mathbf{e}}_j$, (j =1, 2).



PCA finds the linear mode which captures the most variance of the dataset. The eigenvectors may not align close to local data clusters, so the eigenvectors may not represent actual physical states well.

The rotated PCA (RPCA) methods rotate the PCA eigenvectors, so they point closer to the local clusters of data points.
Thus the rotated eigenvectors may bear greater resemblance to actual physical states (though they account for less variance) than the unrotated eigenvectors.

RPCA, also called rotated EOF analysis, and in statistics, PC factor analysis, is a more general but more subjective technique than PCA.

Review the rotation of vectors and matrices:

Given a matrix $\mathbf{P}$ composed of the column vectors $\mathbf{p}_1, \ldots, \mathbf{p}_m$, and a matrix $\mathbf{Q}$ containing the column vectors $\mathbf{q}_1, \ldots, \mathbf{q}_m$, then $\mathbf{P}$ can be transformed into $\mathbf{Q}$ by $\mathbf{Q} = \mathbf{PR}$, i.e.

$$q_{il} = \sum_j p_{ij} r_{jl} \,, \tag{1}$$

where $\mathbf{R}$ is a rotation matrix with elements r_{jl}. When $\mathbf{R}$ is orthonormal, i.e.

$$\mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I} \,, \tag{2}$$

the rotation is called an *orthonormal rotation*. Clearly,

$$\mathbf{R}^{-1} = \mathbf{R}^{\mathrm{T}} \tag{3}$$

for an orthonormal rotation. If $\mathbf{R}$ is not orthonormal, the rotation is an *oblique rotation*.

Q6: Which of the following matrices are orthonormal?
(a) $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and (b) $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Given the data matrix $\mathbf{Y}$,

$$\mathbf{Y} = (y_{il}) = \left(\sum_{j=1}^{m} e_{ij} a_{jl} \right) = \sum_j \mathbf{e}_j \mathbf{a}_j^{\mathrm{T}} = \mathbf{E}\mathbf{A}^{\mathrm{T}}\,, \tag{4}$$

we rewrite it as

$$\mathbf{Y} = \mathbf{E}\mathbf{R}\mathbf{R}^{-1}\mathbf{A}^{\mathrm{T}} = \tilde{\mathbf{E}}\tilde{\mathbf{A}}^{\mathrm{T}}\,, \tag{5}$$

with

$$\tilde{\mathbf{E}} = \mathbf{E}\mathbf{R} \tag{6}$$

and

$$\tilde{\mathbf{A}}^{\mathrm{T}} = \mathbf{R}^{-1}\mathbf{A}^{\mathrm{T}}\,. \tag{7}$$

$\mathbf{E}$ has been rotated into $\tilde{\mathbf{E}}$, and $\mathbf{A}$ into $\tilde{\mathbf{A}}$.
If $\mathbf{R}$ is orthonormal, (3) and (7) yield

$$\tilde{\mathbf{A}} = \mathbf{A}\mathbf{R}\,. \tag{8}$$

To see the orthogonality properties of the rotated eigenvectors, we note that

$$\tilde{\mathbf{E}}^{\mathrm{T}}\tilde{\mathbf{E}} = \mathbf{R}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}\mathbf{E}\mathbf{R} = \mathbf{R}^{\mathrm{T}}\mathbf{D}\mathbf{R}\,, \tag{9}$$

where the diagonal matrix $\mathbf{D}$ is

$$\mathbf{D} = \mathrm{diag}(\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_1, \ldots, \mathbf{e}_m^{\mathrm{T}}\mathbf{e}_m)\,. \tag{10}$$

If $\mathbf{e}_j^{\mathrm{T}}\mathbf{e}_j = 1$, for all j, then $\mathbf{D} = \mathbf{I}$ and (9) reduces to

$$\tilde{\mathbf{E}}^{\mathrm{T}}\tilde{\mathbf{E}} = \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}\,, \tag{11}$$

i.e. the $\{\tilde{\mathbf{e}}_j\}$ are orthonormal. Hence the rotated eigenvectors $\{\tilde{\mathbf{e}}_j\}$ are orthonormal only if the original eigenvectors $\{\mathbf{e}_j\}$ are orthonormal.

If $\{\mathbf{e}_j\}$ are orthogonal but not orthonormal, then $\{\tilde{\mathbf{e}}_j\}$ are in general not orthogonal.

Recall the PCs $\{a_j(t_l)\}$ are uncorrelated, i.e. the covariance matrix

$$\mathbf{C}_{\mathbf{AA}} = \mathrm{diag}(\alpha_1^2, \ldots, \alpha_m^2)\,, \tag{12}$$

where

$$\mathbf{a}_j^{\mathrm{T}}\mathbf{a}_j = \alpha_j^2\,. \tag{13}$$

With the rotated PCs, their covariance matrix is

$$\mathbf{C}_{\tilde{\mathbf{A}}\tilde{\mathbf{A}}} = \mathrm{cov}(\mathbf{R}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}, \mathbf{A}\mathbf{R}) = \mathbf{R}^{\mathrm{T}}\,\mathrm{cov}(\mathbf{A}^{\mathrm{T}}, \mathbf{A})\,\mathbf{R} = \mathbf{R}^{\mathrm{T}}\,\mathbf{C}_{\mathbf{AA}}\,\mathbf{R}\,. \tag{14}$$

Hence $\mathbf{C}_{\tilde{\mathbf{A}}\tilde{\mathbf{A}}}$ is diagonal only if $\mathbf{C}_{\mathbf{AA}} = \mathbf{I}$, i.e. $\mathbf{a}_j^{\mathrm{T}}\mathbf{a}_j = 1$, for all j.

2 cases:

Case a: If we choose $\mathbf{e}_j^{\mathrm{T}}\mathbf{e}_j = 1$, for all j, we cannot have $\mathbf{a}_j^{\mathrm{T}}\mathbf{a}_j = 1$, for all j. This implies that $\{\tilde{\mathbf{a}}_j\}$ are not uncorrelated, but $\{\tilde{\mathbf{e}}_j\}$ are orthonormal.

Case b: If we choose $\mathbf{a}_j^{\mathrm{T}}\mathbf{a}_j = 1$, for all j, we cannot have $\mathbf{e}_j^{\mathrm{T}}\mathbf{e}_j = 1$, for all j. This implies that $\{\tilde{\mathbf{a}}_j\}$ are uncorrelated, but $\{\tilde{\mathbf{e}}_j\}$ are not orthonormal.

PCA can have both $\{\mathbf{e}_j\}$ orthonormal and $\{\mathbf{a}_j\}$ uncorrelated, but RPCA can only possess one of these two properties.
In general, out of a total of m PCA modes, only the k leading ones are selected for rotation, with the rest discarded.

Many possible criteria for rotation ⇒ many RPCA schemes.

Richman (1986) listed 5 orthogonal and 14 oblique rotation schemes.

Varimax scheme (Kaiser, 1958) is most popular among orthogonal rotation schemes.

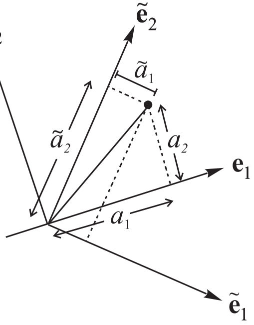
E.g. the first 2 eigenvectors are chosen for rotation. The data are first projected onto the 2 PCA eigenvectors $\mathbf{e}_j$ $(j = 1, 2)$ to get the first two PCs

$$a_j(t_l) = \sum_i e_{ji} y_{il} \,. \tag{15}$$

With rotated eigenvectors $\tilde{\mathbf{e}}_j$, the rotated PCs are

$$\tilde{a}_j(t_l) = \sum_i \tilde{e}_{ji} y_{il} \,. \tag{16}$$

A common objective in rotation is to make $\tilde{a}_j^2(t_l)$ either as large as possible, or as close to zero as possible, i.e. to maximize the variance of the square of the rotated PCs.



The rotation which has yielded $|\tilde{a}_1| < |a_1|, |a_2| < |\tilde{a}_2|$, i.e. instead of intermediate magnitudes for a_1, a_2, the rotated PCs have either larger or smaller magnitudes.

Geometrically, this means the rotated axes (i.e. the eigenvectors) point closer to actual data points than the unrotated axes.

If the rotated vector $\tilde{\mathbf{e}}_2$ actually passes through the data point in the Fig., then $|\tilde{a}_1|$ is zero, while $|\tilde{a}_2|$ assumes its largest possible value.

The varimax criterion is to maximize $f(\tilde{\mathbf{A}}) = \sum_{j=1}^{k} \mathrm{var}(\tilde{a}_j^2)$, i.e.

$$f(\tilde{\mathbf{A}}) = \sum_{j=1}^{k} \left\{ \frac{1}{n} \sum_{l=1}^{n} [\tilde{a}_j^2(t_l)]^2 - \left[\frac{1}{n} \sum_{l=1}^{n} \tilde{a}_j^2(t_l) \right]^2 \right\} . \qquad (17)$$

Kaiser (1958) found an iterative algorithm for finding the rotation matrix **R** (see also Preisendorfer, 1988, pp.273-277).

In the above varimax criterion, (17) maximizes the variance of the rotated squared PCs. An alternative (which is actually the one used in traditional factor analysis) is to maximize the variance of the rotated squared loadings $\tilde{e}_{ji}^2$, i.e. maximize

$$f(\tilde{\mathbf{E}}) = \sum_{j=1}^{k} \left\{ \frac{1}{m} \sum_{i=1}^{m} [\tilde{e}_{ji}^2]^2 - \left[\frac{1}{m} \sum_{i=1}^{m} \tilde{e}_{ji}^2 \right]^2 \right\} . \qquad (18)$$

That the squared loadings are made as large as possible or as close to zero as possible means that many of the loadings are essentially set to zero, yielding loading patterns which have more localized features than the unrotated patterns.

E.g. Horel (1981) showed both the rotated and unrotated loading patterns for the 500 mb height data of winter months in the Northern Hemisphere, with the rotated patterns showing more regionalized anomalies than the unrotated patterns, where the anomalies were spread all over the Northern Hemisphere.

RPCA used to extract large scale teleconnection patterns/indices in N.Hemisphere atmosphere
http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml

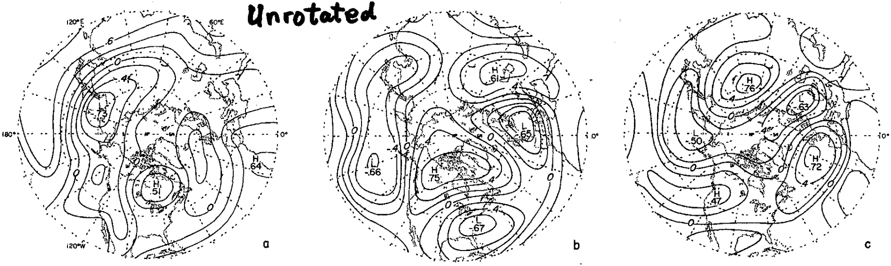


FIG. 1. Eigenvectors of the first ~~four~~ 3 principal components derived from 500 mb height time series consisting of 45 winter months (from Wallace and Gutzler, 1981). The value of the *k*th eigenvector at any location represents the temporal correlation between the *k*th principal component and the time series of local 500 mb height. The gridpoints used are indicated in Fig. 4b in the paper by Wallace and Gutzler (1981). The sign convention is arbitrary, and contour interval is 0.2.

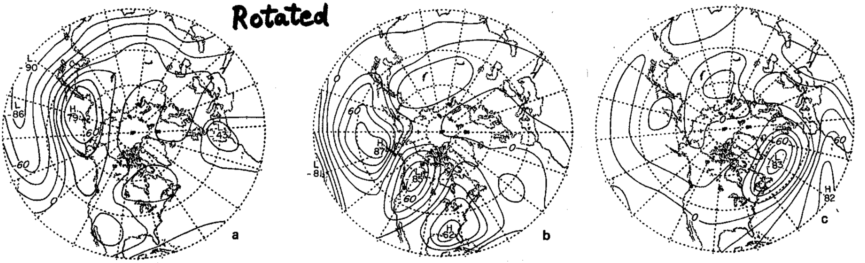


FIG. 2. Loading vectors of the first ~~four~~ 3 rotated principal components derived from the same data set and principal components described in Fig. 1. The rotated principal components are a varimax rotation of the 19 principal components whose first four eigenvectors are shown in Fig. 1. The value of the *k*th loading vector at any location represents the correlation between the *k*th rotated principal component and local 500 mb height. The contour interval is 0.2.

These two ways of performing rotation can be seen as working with either the data matrix or the transpose of the data matrix. In PCA, using the transpose of the data matrix does not change the results (but can be exploited to save considerable computional time by working with the smaller data covariance matrix).

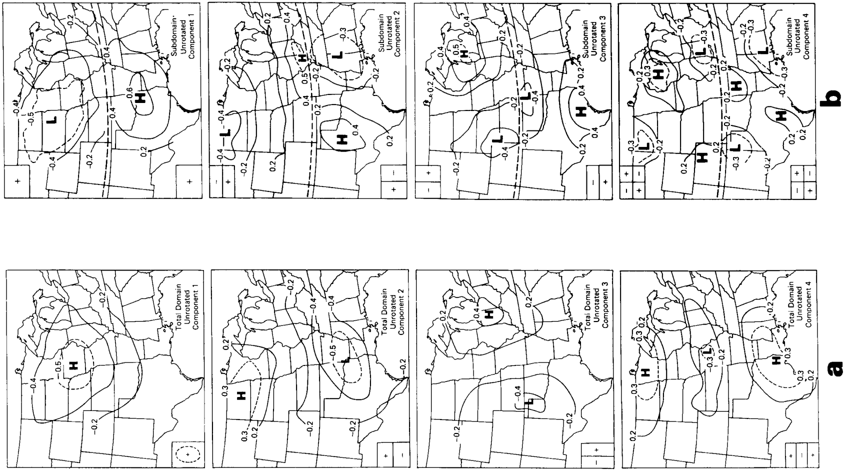
In RPCA, taking the transpose reverses the role of the PCs and the loadings, thereby changing from a rotational criterion on the loadings to one on the PCs.

Richman (1986): applying rotational criterion on the loadings yielded loading patterns with far fewer anomaly centres than observed in typical 700 mb height anomaly maps of the Northern Hemisphere, whereas applying rotational criterion on PCs yielded loading patterns in good agreement with commonly observed patterns.

Four main disadvantages with PCA:

(i) Domain shape dependence: Often the PCA spatial modes are related simply to the spatial harmonics than to physical states.
(ii) Subdomain instability: If the domain is divided into two parts, then the PCA mode 1 spatial patterns for the subdomains may not be similar to the spatial mode calculated for the whole domain.

Richman (1986) shows first 4 PCA spatial modes of the 3-day May–August precipitation over central U.S.A.
(a) Left panels show the four modes computed for the whole domain;
(b) right panels, the modes computed separately for the northern and southern halves of the full domain. The dash lines in (b) indicate the boundary of the two halves.
Insets show the basic harmonic patterns found by the modes.


Subdomain Unrotated Component 1
Subdomain Unrotated Component 2
Subdomain Unrotated Component 3
Subdomain Unrotated Component 4
b
Total Domain Unrotated Component 1
Total Domain Unrotated Component 2
Total Domain Unrotated Component 3
Total Domain Unrotated Component 4
a

(iii) Degeneracy: If $\lambda_i \approx \lambda_j$, near degeneracy means eigenvectors $\mathbf{e}_i$ and $\mathbf{e}_j$ cannot be estimated accurately by PCA.
(iv) Neglect of regional correlated patterns: Small regional correlated patterns tend to be ignored by PCA, as PCA spatial modes tend to be related to the dominant spatial harmonics.

RPCA improves on all (i) to (iv).

Four disadvantages with RPCA:

(i) Many possible choices for the rotation criterion: Richman (1986) listed 19 types of rotation schemes. Critics complain that rotation is too subjective. Furthermore, the rotational criterion can be applied to the loadings or to the PCs, yielding different results.

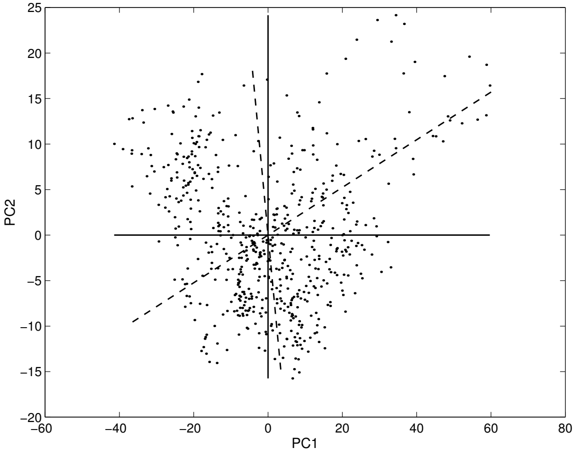
(ii) Dependence on k, the number of PCA modes chosen for rotation: If the first k PCA modes are selected for rotation, changing k can lead to large changes in the RPCAs.

E.g., in RPCA, if one first chooses $k = 3$, then one chooses $k = 4$, the first three RPCAs are changed. In PCA, if one first chooses $k = 3$, then $k = 4$, the first three PCAs are unchanged.

(iii) Dependence on how the PCA eigenvectors and PCs are normalized before rotation is performed.

(iv) Less variance explained: variance of the data accounted for by the first k RPCA modes is $\leq$ variance explained by the first k PCA modes.

Return to the tropical Pacific SST PCA modes. PC1-PC2 values are shown as dots in a scatter plot, with La Niña states lie in the upper left corner, and El Niño states in the upper right corner.



The first PCA eigenvector lies along the horizontal line, and the second PCA, along the vertical line, neither of which would come close to the El Niño nor the La Niña states.

Using the varimax criterion on the squared PCs, a rotation is performed on the first 3 PCA eigenvectors.

The first RPCA eigenvector, shown as a dashed line, spears through the cluster of El Niño states in the upper right corner.

In terms of variance explained, the first RPCA mode explained only 91.7% as much variance as the first PCA mode.

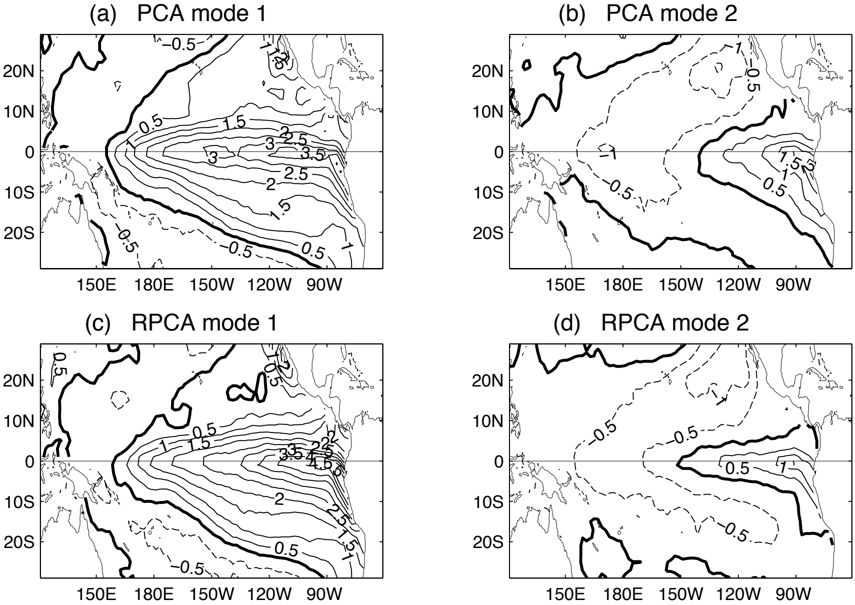


Figure : SST anomalies pattern for PCA modes (a) 1 and (b) 2, and RPCA modes (c) 1 and (d) 2, when their corresponding PC is at max.

RPCA functions in Matlab

`www.mathworks.com/help/toolbox/stats/rotatefactors.html`

[Erotate, R] = rotatefactors(E);
for varimax rotation (can also do other types of rotation). E is an $m \times k$ matrix, containing the first k PCA eigenvectors, R is the $k \times k$ rotation matrix, and Erotate = ER gives the rotated eigenvectors. Note: my notation $[m, n, k]$ equals Matlab's notation $[d, n, m]$.

Can also use
`www.mathworks.com/help/toolbox/stats/factoran.html`

2.14 PCA for vectors [Book, Sect. 2.3]

For vector variables, e.g. wind velocity (u, v), there are several options for performing PCA:
(a) Simply apply PCA to the u field and to the v field separately.

(b) Do a *combined PCA*, i.e. treat the v variables as though they were extra u variables, so the data matrix becomes

$$\mathbf{Y} = \begin{bmatrix} u_{11} & \cdots & u_{1n} \\ \cdots & \cdots & \cdots \\ u_{m1} & \cdots & u_{mn} \\ v_{11} & \cdots & v_{1n} \\ \cdots & \cdots & \cdots \\ v_{m1} & \cdots & v_{mn} \end{bmatrix}, \qquad (19)$$

where m is the number of spatial points and n the number of time points. In cases (a) and (b), the vector can be generalized to > 2 dimensions.

If the vector is 2-dimensional, one has option (c): combine u and v into a complex variable, and perform a complex PCA.
Let

$$w = u + \mathrm{i}v \,. \tag{20}$$

PCA applied to w allows the data matrix to be expressed as

$$\mathbf{Y} = \sum_j \mathbf{e}_j \mathbf{a}_j^{*\mathrm{T}} \,, \tag{21}$$

where the superscript ${}^{*\mathrm{T}}$ denotes the complex conjugate transpose. Since the covariance matrix is Hermitian and positive-semi definite, eigenvalues of $\mathbf{C}$ are real and non-negative, though $\mathbf{e}_j$ and $\mathbf{a}_j$ are in general complex.

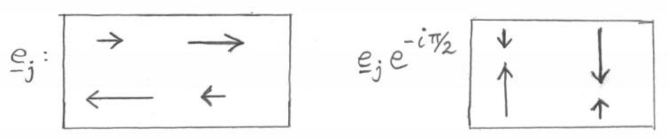
Write the lth component of $\mathbf{a}_j$ as

$$a_{jl} = |a_{jl}|\, \mathrm{e}^{\mathrm{i}\theta_{jl}}\,, \tag{22}$$

then

$$Y_{il} = \sum_j e_{ij} \mathrm{e}^{-\mathrm{i}\theta_{jl}} |a_{jl}|\,. \tag{23}$$

Interpret $e_{ij}\mathrm{e}^{-\mathrm{i}\theta_{jl}}$ as each complex element of $\mathbf{e}_j$ being rotated by the same angle θ_{jl} during the lth time interval. Similarly, each element of $\mathbf{e}_j$ is amplified by the same factor $|a_{jl}|$.



Q7: Complex PCA is used to analyze the horizontal wind vectors from two stations, with the first station located to the west of the second station. The first eigenvector gives $\mathbf{e}_1^{\mathrm{T}} = [1 + \mathrm{i}, \ -\mathrm{i}]$. Sketch the horizontal wind field at time (a) $t = 1$ when the first principal component takes on the value -1, and at (b) $t = 2$, when the first PC $= -\mathrm{i}$.

When PCA is applied to real variables, the real $\mathbf{e}_j$ and $\mathbf{a}_j$ can both be multiplied by -1. When PCA is applied to complex variables, an arbitrary phase ϕ_j can be attached to the complex $\mathbf{e}_j$ and $\mathbf{a}_j$, as follows

$$\mathbf{Y} = \sum_j (\mathbf{e}_j e^{i\phi_j})(e^{-i\phi_j}\mathbf{a}_j^{*\mathrm{T}}) \,. \tag{24}$$

Often the arbitrary phase is chosen to make the interpretation of the modes easier.

E.g., in the analysis of the tropical Pacific wind field, Legler (1983) chose ϕ_j so that $e^{-i\phi_j}\mathbf{a}_j^{*\mathrm{T}}$ lies mainly along the real axis.

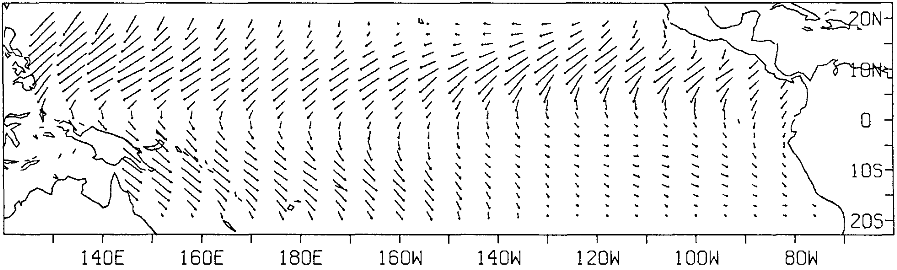


FIG. 3. Spatial eigenvector, E_1. Vectors are dimensionless.

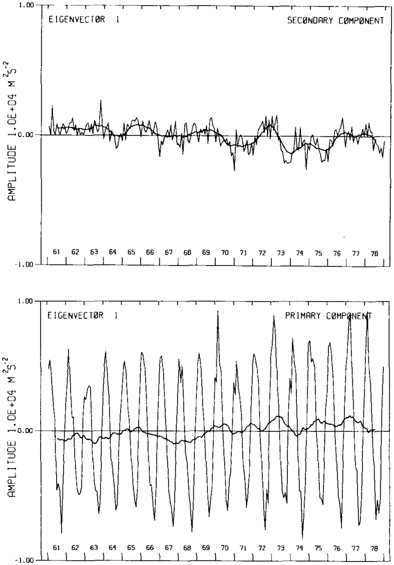


FIG. 4. Primary and secondary components corresponding to E_1. Twelve-month running mean is indicated by a heavy line. The amplitude is normalized.

Figure : Imaginary component (top) & real component (bottom) of PC1.

In the ocean, dynamical theory predicts the near-surface wind-driven current to spiral and diminish with depth, in the shape of an ‘Ekman spiral’. This fascinating spiral shape was detected by the complex PCA (Stacey et al., 1986).

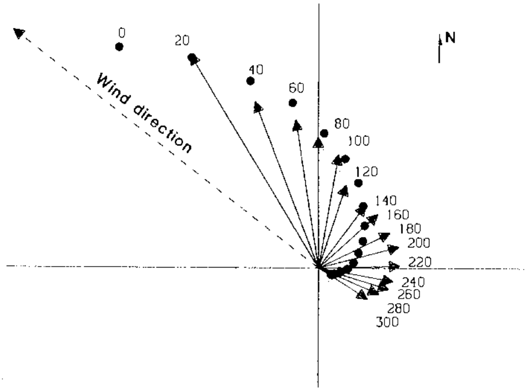


Fig. 3. The mode 1 eigenfunction. Its direction has been fixed by picking the arrow at 20 m to lie along the direction of the principal axis of the covariance matrix of the currents at 20 m. The dots show the amplitude and direction as given by the Ekman spiral. The wind direction is the direction of the principal axis of the covariance matrix of the wind.

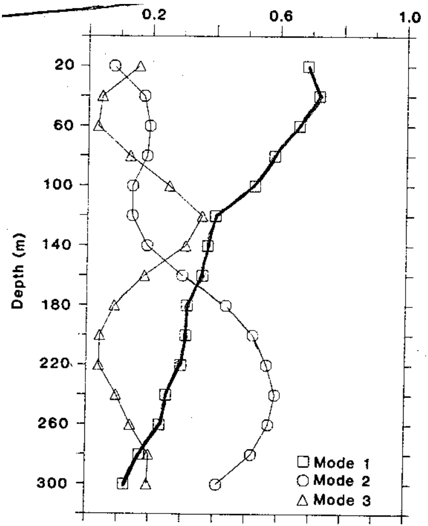


Fig. 2. The fraction of the variance explained by the first three orthogonal modes as a function of depth.

References:

Horel, J. D. (1981). A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field. *Monthly Weather Review*, 109:2080–2092.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23:187–200.

Legler, D. M. (1983). Empirical orthogonal function analysis of wind vectors over the tropical Pacific region. *Bulletin of the American Meteorological Society*, 64(3):234–241.

Preisendorfer, R. W. (1988). *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, New York.

Richman, M. B. (1986). Rotation of principal components. *Journal of Climatology*, 6:293–335.

Stacey, M. W., Pond, S., and LeBlond, P. H. (1986). A wind-forced Ekman spiral as a good statistical fit to low-frequency currents in a coastal strait. *Science*, 233:470–472.