Ch.2 Principal component analysis (PCA)

Books on PCA by Jolliffe (2002), Preisendorfer (1988).
PCA also called **empirical orthogonal function (EOF) analysis**.

2.1 Geometric approach to PCA [Book, Sect. 2.1.1]

Dataset with variables $y_1, \cdots, y_m$, each variable sampled n times, e.g. m time series each containing n observations in time.
For instance, one may have a dataset containing the monthly air temperature measured at m stations over n months.

Objective of PCA: If m is a large number, we would like to capture the essence of $y_1, \cdots, y_m$ by a smaller set of variables $z_1, \cdots, z_k$ (i.e. $k < m$; and hopefully $k \ll m$, for truly large m).

Begin with an intuitive geometric approach.
Start with only 2 variables, y_1 and y_2.

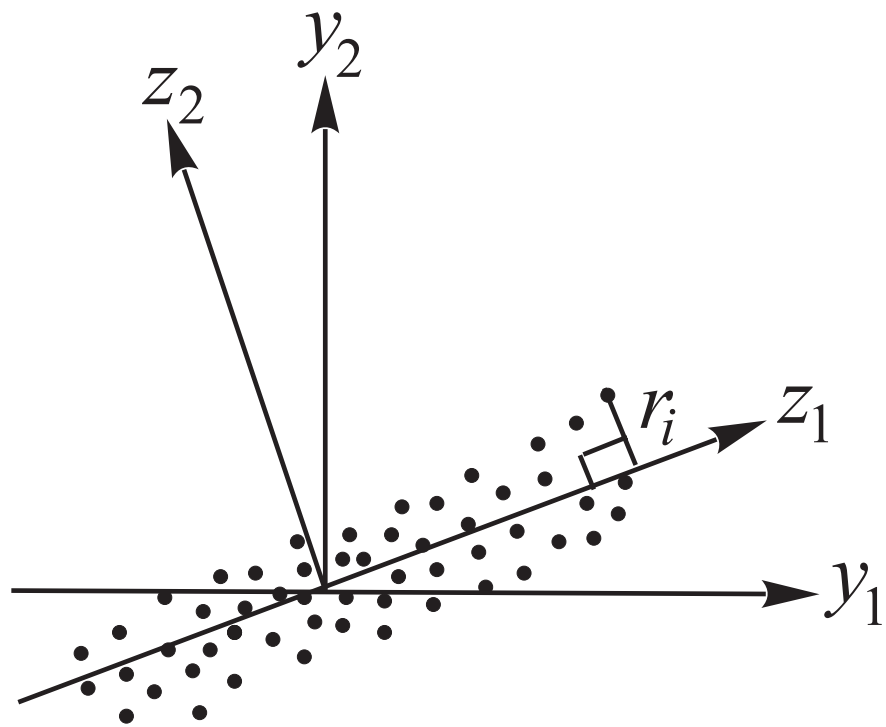


Figure : The PCA problem formulated as a minimization of the sum of r_i^2, where r_i is the shortest distance from the ith data point to the first PCA axis z_1.

Optimal z_1 found by minimizing $\sum_{i=1}^{n} r_i^2$.
This geometric approach to PCA due to Pearson (1901).

Note: PCA treats all variables equally, whereas regression divides variables into independent and dependent variables.

In 3-D, z_1 is the best 1-D line fit to the data, while z_1 and z_2 span a 2-D plane giving the best plane fit to the data.
In general, with an m-dimensional dataset, we want to find the k-dimensional hyperplane giving the best fit.

2.2 Eigenvector approach to PCA [Book, Sect.2.1.2]

The more systematic eigenvector approach to PCA is due to Hotelling (1933). In 2-D example, a data point is transformed from

its old coordinates (y_1, y_2) to new coordinates (z_1, z_2) via a rotation of the coordinate system:

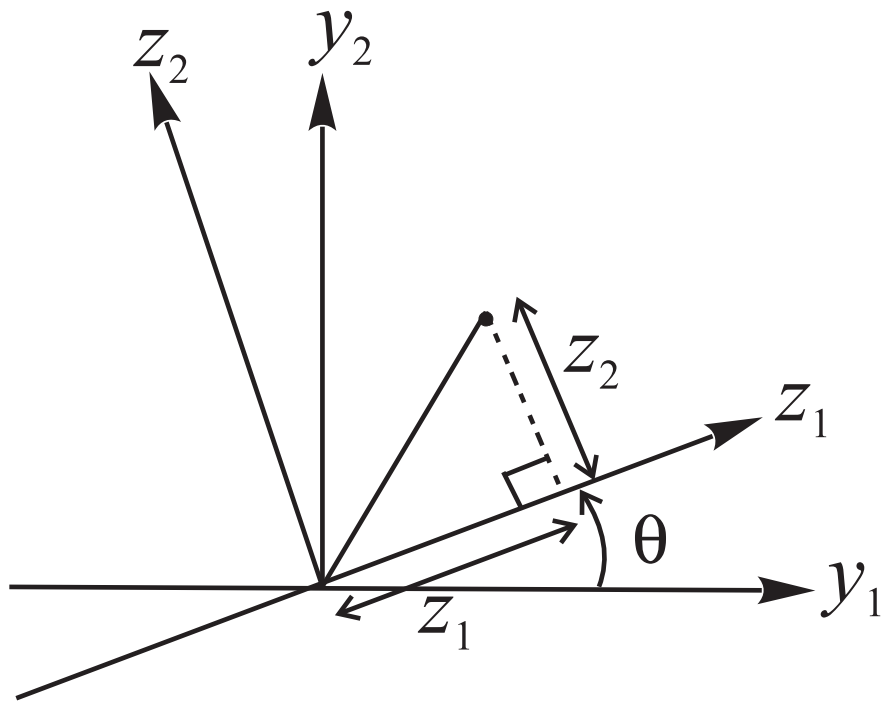


Figure : Rotation of coordinate axes by an angle θ in a 2-dimensional space.

$$
\begin{aligned}
z_1 &= y_1 \cos\theta + y_2 \sin\theta \\
z_2 &= -y_1 \sin\theta + y_2 \cos\theta
\end{aligned} \tag{1}
$$

In the general m-dimensional case, introduce new coordinates

$$
z_j = \sum_{l=1}^{m} e_{jl} y_l, \qquad j = 1, \cdots, m. \tag{2}
$$

The objective is to find

$$
\mathbf{e}_1 = \left[e_{11}, \cdots, e_{1m}\right]^{\mathrm{T}}, \tag{3}
$$

which maximizes $\mathrm{var}(z_1)$, i.e. find the coordinate transformation such that the variance of the dataset along the direction of the z_1 axis is maximized.
With

$$
z_1 = \sum_{l=1}^{m} e_{1l} y_l = \mathbf{e}_1^{\mathrm{T}} \mathbf{y}, \qquad \mathbf{y} = \left[y_1, \ldots, y_m\right]^{\mathrm{T}}, \tag{4}
$$

i.e. projecting the data point $\mathbf{y}$ onto the vector $\mathbf{e}_1$ gives a distance of z_1 along the $\mathbf{e}_1$ direction, we have

$$\mathrm{var}(z_1) = \mathrm{E}\left[(z_1 - \overline{z}_1)(z_1 - \overline{z}_1)\right] = \mathrm{E}\left[\mathbf{e}_1^{\mathrm{T}}(\mathbf{y} - \overline{\mathbf{y}})(\mathbf{y} - \overline{\mathbf{y}})^{\mathrm{T}}\mathbf{e}_1\right], \quad (5)$$

where we have used the vector property $\mathbf{a}^{\mathrm{T}}\mathbf{b} = \mathbf{b}^{\mathrm{T}}\mathbf{a}$. Thus,

$$\mathrm{var}(z_1) = \mathbf{e}_1^{\mathrm{T}}\,\mathrm{E}\left[(\mathbf{y} - \overline{\mathbf{y}})(\mathbf{y} - \overline{\mathbf{y}})^{\mathrm{T}}\right]\mathbf{e}_1 = \mathbf{e}_1^{\mathrm{T}}\mathbf{C}\mathbf{e}_1\,, \quad (6)$$

with the *covariance matrix* $\mathbf{C}$ given by

$$\mathbf{C} = \mathrm{E}\left[(\mathbf{y} - \overline{\mathbf{y}})(\mathbf{y} - \overline{\mathbf{y}})^{\mathrm{T}}\right]. \quad (7)$$

The larger is the vector norm $\|\mathbf{e}_1\|$, the larger $\mathrm{var}(z_1)$ will be. Hence, need constraint on $\|\mathbf{e}_1\|$ while maximizing $\mathrm{var}(z_1)$. Impose a normalization constraint $\|\mathbf{e}_1\| = 1$, i.e.

$$\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_1 = 1\,. \quad (8)$$

Optimization problem is to find $\mathbf{e}_1$ which maximizes $\mathbf{e}_1^{\mathrm{T}}\mathbf{C}\mathbf{e}_1$, subject to the constraint

$$\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_1 - 1 = 0\,. \tag{9}$$

Method of Lagrange multipliers commonly used to do optimization under constraints (Book, Appendix B). Instead of finding stationary points of $\mathbf{e}_1^{\mathrm{T}}\mathbf{C}\mathbf{e}_1$, we search for the stationary points of the Lagrange function L,

$$L = \mathbf{e}_1^{\mathrm{T}}\mathbf{C}\mathbf{e}_1 - \lambda(\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_1 - 1)\,, \tag{10}$$

where λ is a Lagrange multiplier.

Differentiating L by the elements of $\mathbf{e}_1$, and setting the derivatives to zero:

$$\mathbf{C}\mathbf{e}_1 - \lambda\mathbf{e}_1 = 0\,, \tag{11}$$

i.e. λ is an eigenvalue of the covariance matrix $\mathbf{C}$, with $\mathbf{e}_1$ the eigenvector.

Multiplying this equation by $\mathbf{e}_1^{\mathrm{T}}$ on the left,

$$\lambda = \mathbf{e}_1^{\mathrm{T}} \mathbf{C} \mathbf{e}_1 = \mathrm{var}(z_1)\,. \tag{12}$$

Since $\mathbf{e}_1^{\mathrm{T}} \mathbf{C} \mathbf{e}_1$ is maximized, so are $\mathrm{var}(z_1)$ and λ.

New coordinate z_1, called the *principal component* (PC), is found from (4).

- - - - - - - - -

Next, find z_2:

Our task is to find $\mathbf{e}_2$ which maximizes $\mathrm{var}(z_2) = \mathbf{e}_2^{\mathrm{T}}\mathbf{C}\mathbf{e}_2$, subject to the constraint $\mathbf{e}_2^{\mathrm{T}}\mathbf{e}_2 = 1$, and the constraint that z_2 be uncorrelated with z_1, i.e. the covariance between z_2 and z_1 be zero,

$$\mathrm{cov}(z_1, z_2) = 0\,. \tag{13}$$

As $\mathbf{C} = \mathbf{C}^{\mathrm{T}}$, we can write

$$\begin{aligned} 0 &= \mathrm{cov}(z_1, z_2) = \mathrm{cov}(\mathbf{e}_1^{\mathrm{T}}\mathbf{y}, \mathbf{e}_2^{\mathrm{T}}\mathbf{y}) \\ &= \mathrm{E}[\mathbf{e}_1^{\mathrm{T}}(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^{\mathrm{T}}\mathbf{e}_2] = \mathbf{e}_1^{\mathrm{T}}\mathrm{E}[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^{\mathrm{T}}]\mathbf{e}_2 \\ &= \mathbf{e}_1^{\mathrm{T}}\mathbf{C}\mathbf{e}_2 = \mathbf{e}_2^{\mathrm{T}}\mathbf{C}\mathbf{e}_1 = \mathbf{e}_2^{\mathrm{T}}\lambda_1\mathbf{e}_1 = \lambda_1\mathbf{e}_2^{\mathrm{T}}\mathbf{e}_1 = \lambda_1\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_2\ . \end{aligned} \tag{14}$$

The orthogonality condition

$$\mathbf{e}_2^{\mathrm{T}}\mathbf{e}_1 = 0\,, \tag{15}$$

can be used as a constraint in place of (13).

Upon introducing another Lagrange multiplier γ, we want to find an $\mathbf{e}_2$ which gives a stationary point of the Lagrange function L,

$$L = \mathbf{e}_2^{\mathrm{T}}\mathbf{C}\mathbf{e}_2 - \lambda(\mathbf{e}_2^{\mathrm{T}}\mathbf{e}_2 - 1) - \gamma\,\mathbf{e}_2^{\mathrm{T}}\mathbf{e}_1\,. \tag{16}$$

Differentiating L by the elements of $\mathbf{e}_2$, and setting the derivatives to zero:

$$\mathbf{C}\mathbf{e}_2 - \lambda\mathbf{e}_2 - \gamma\mathbf{e}_1 = 0\,. \tag{17}$$

Left multiplying this equation by $\mathbf{e}_1^{\mathrm{T}}$ yields

$$\mathbf{e}_1^{\mathrm{T}}\mathbf{C}\mathbf{e}_2 - \lambda\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_2 - \gamma\,\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_1 = 0\,. \tag{18}$$

On the left hand side, the first two terms are both zero from (14) while the third term is simply γ, so we have $\gamma = 0$, and (17) reduces to

$$\mathbf{C}\mathbf{e}_2 - \lambda\mathbf{e}_2 = 0\,, \tag{19}$$

i.e. λ is again an eigenvalue of $\mathbf{C}$, with $\mathbf{e}_2$ the eigenvector. As

$$\lambda = \mathbf{e}_2^{\mathrm{T}} \mathbf{C} \mathbf{e}_2 = \mathrm{var}(z_2)\,, \tag{20}$$

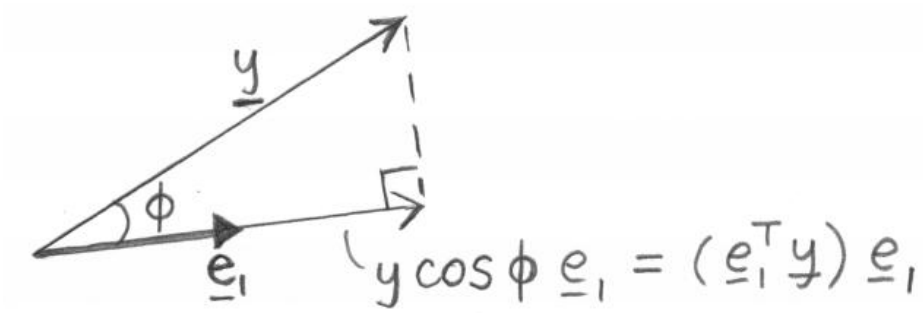
which is maximized, this $\lambda = \lambda_2$ is as large as possible with $\lambda_2 < \lambda_1$. (The case $\lambda_2 = \lambda_1$ is degenerate and will be discussed later). Hence, λ_2 is the second largest eigenvalue of $\mathbf{C}$, with $\lambda_2 = \mathrm{var}(z_2)$. This process can be repeated for z_3, $z_4, \ldots$

- - - - - - - -

How to reconcile the geometric approach and the eigenvector approach?

First we subtract the mean $\overline{\mathbf{y}}$ from $\mathbf{y}$, so the transformed data are centred around the origin with $\overline{\mathbf{y}} = 0$. In the geometric approach, we minimize the distance between the data points and the new axis. If the unit vector $\mathbf{e}_1$ gives the direction of the new axis, then the

projection of a data point (described by the vector $\mathbf{y}$) onto $\mathbf{e}_1$ is $(\mathbf{e}_1^{\mathrm{T}}\mathbf{y})\mathbf{e}_1$. The component of $\mathbf{y}$ normal to $\mathbf{e}_1$ is $\mathbf{y} - (\mathbf{e}_1^{\mathrm{T}}\mathbf{y})\mathbf{e}_1$.



Thus minimizing the distance between the data points and the new axis amounts to minimizing

$$\epsilon = \mathrm{E}[\,\|\,\mathbf{y} - (\mathbf{e}_1^{\mathrm{T}}\mathbf{y})\mathbf{e}_1\|^2\,]\,. \tag{21}$$

$$\epsilon = \mathrm{E}[\,\|\mathbf{y}\|^2 - 2(\mathbf{e}_1^{\mathrm{T}}\mathbf{y})(\mathbf{e}_1^{\mathrm{T}}\mathbf{y}) + (\mathbf{e}_1^{\mathrm{T}}\mathbf{y})\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_1(\mathbf{e}_1^{\mathrm{T}}\mathbf{y})\,], \tag{22}$$

$$\epsilon = \mathrm{E}[\,\|\mathbf{y}\|^2 - (\mathbf{e}_1^{\mathrm{T}}\mathbf{y})^2\,] = \mathrm{var}(\mathbf{y}) - \mathrm{var}(\mathbf{e}_1^{\mathrm{T}}\mathbf{y})\,, \tag{23}$$

where $\mathrm{var}(\mathbf{y}) \equiv \mathrm{E}[\,\|\mathbf{y}\|^2]$, with $\overline{\mathbf{y}}$ assumed to be zero. Since $\mathrm{var}(\mathbf{y})$ is constant, minimizing ϵ is equivalent to maximizing $\mathrm{var}(\mathbf{e}_1^{\mathrm{T}}\mathbf{y})$, which is equivalent to maximizing $\mathrm{var}(z_1)$.

Hence the geometric approach of minimizing the distance between the data points and the new axis is equivalent to the eigenvector approach in finding the largest eigenvalue λ, which is simply $\max[\mathrm{var}(z_1)]$.

Data covariance matrix versus data correlation matrix

So far, $\mathbf{C}$ is the data covariance matrix, but it can also be the data correlation matrix, if one prefers.

In *combined PCA*, where two or more variables with different units are combined into one large data matrix for PCA— e.g. finding the PCA modes of the combined sea surface temperature data and the

sea level pressure data— then one needs to standardize the variables, so $\mathbf{C}$ is the correlation matrix.

2.3 Real and complex data [Book, Sect.2.1.3]

In general, for $\mathbf{y}$ real,

$$\mathbf{C} \equiv \mathrm{E}[(\mathbf{y} - \overline{\mathbf{y}})(\mathbf{y} - \overline{\mathbf{y}})^{\mathrm{T}}] \,, \tag{24}$$

implies that $\mathbf{C}^{\mathrm{T}} = \mathbf{C}$, i.e. $\mathbf{C}$ is a real, symmetric matrix.

A *positive semi-definite matrix* $\mathbf{A}$ is defined by the property that for any vector $\mathbf{v} \neq \mathbf{0}$, it follows that $\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v} \geq 0$ (Strang, 2005).

Q1: Prove that $\mathbf{C}$ is a real, symmetric, positive semi-definite matrix.

If $\mathbf{y}$ is complex, then

$$\mathbf{C} \equiv \mathrm{E}[(\mathbf{y} - \overline{\mathbf{y}})(\mathbf{y} - \overline{\mathbf{y}})^{\mathrm{T}*}], \tag{25}$$

with complex conjugation denoted by the superscript asterisk. As $\mathbf{C}^{\mathrm{T}*} = \mathbf{C}$, $\mathbf{C}$ is a *Hermitian matrix*. It is also a positive semi-definite matrix.

Theorems on Hermitian, positive semi-definite matrices tell us: $\mathbf{C}$ has real eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0, \qquad \sum_{j=1}^{m} \lambda_j = \mathrm{var}(\mathbf{y}), \tag{26}$$

with corresponding orthonormal eigenvectors, $\mathbf{e}_1, \ldots, \mathbf{e}_m$. The k eigenvectors corresponding to $\lambda_1, \ldots, \lambda_k$ minimize

$$\epsilon_k = \mathrm{E}[\, \|(\mathbf{y} - \overline{\mathbf{y}}) - \sum_{j=1}^{k} (\mathbf{e}_j^{\mathrm{T}}(\mathbf{y} - \overline{\mathbf{y}}))\mathbf{e}_j\|^2 \,], \tag{27}$$

which can be expressed as

$$\epsilon_k = \mathrm{var}(\mathbf{y}) - \sum_{j=1}^{k} \lambda_j \,. \tag{28}$$

Hence λ_j is the variance explained by mode j.

2.4 Orthogonality relations [Book, Sect.2.1.4]

PCA finds the eigenvectors and eigenvalues of $\mathbf{C}$.

The orthonormal eigenvectors then provide a basis, i.e. the data $\mathbf{y}$ can be expanded in terms of the eigenvectors $\mathbf{e}_j$:

$$\boxed{\mathbf{y} - \overline{\mathbf{y}} = \sum_{j=1}^{m} a_j(t)\mathbf{e}_j \,,} \tag{29}$$

where $a_j(t)$ are the expansion coefficients. To obtain $a_j(t)$, left multiply the above equation by $\mathbf{e}_i^{\mathrm{T}}$, and use the orthonormal relation of the eigenvectors,

$$\mathbf{e}_i^{\mathrm{T}}\mathbf{e}_j = \delta_{ij} \,, \tag{30}$$

(with δ_{ij} denoting the Kronecker delta function, which equals 1 if $i = j$, and 0 otherwise) to get

$$\boxed{a_j(t) = \mathbf{e}_j^{\mathrm{T}}(\mathbf{y} - \overline{\mathbf{y}}) \,,} \tag{31}$$

i.e. $a_j(t)$ is obtained by the projection of the data vector $\mathbf{y} - \overline{\mathbf{y}}$ onto the eigenvector $\mathbf{e}_j$, as the right hand side of this equation is simply a dot product between the two vectors.

Nomenclature varies in the literature:
a_j are called principal components, scores, temporal coefficients and amplitudes;
eigenvectors $\mathbf{e}_j$ are also referred to as principal vectors, loadings, spatial patterns and EOFs (Empirical Orthogonal Functions).

We prefer calling a_j *principal components* (PCs), $\mathbf{e}_j$ eigenvectors or EOFs (and the elements e_{ij} *loadings*), and j the mode number.

Note that for time series, a_j is a function of time while $\mathbf{e}_j$ is a function of space, hence the names temporal coefficients and spatial patterns describe them well.

However, in many cases, the dataset may not consist of time series. For instance, the dataset could be plankton collected from various oceanographic stations— t then becomes the label for a station, while 'space' here could represent the various plankton species, and the data $\mathbf{y}(t) = [y_1(t), \ldots, y_m(t)]^{\mathrm{T}}$ could be the amount of species $1, \ldots, m$ found in station t.

Another example comes from the multi-channel satellite image data, where images of the earth's surface have been collected at several frequency channels. Here t becomes the location label for a pixel in an image, and 'space' indicates the various frequency channels.

There are two important properties of PCAs. The expansion $\sum_{j=1}^{k} a_j(t)\mathbf{e}_j(\mathbf{x})$, with $k \leq m$, explains more of the variance of the data than any other linear combination $\sum_{j=1}^{k} b_j(t)\mathbf{f}_j(\mathbf{x})$.

Thus PCA provides the most efficient way to compress data, using k eigenvectors $\mathbf{e}_j$ and corresponding time series a_j.

The second important property is that the time series in the set $\{a_j\}$ are uncorrelated. We can write

$$a_j(t) = \mathbf{e}_j^{\mathrm{T}}(\mathbf{y} - \overline{\mathbf{y}}) = (\mathbf{y} - \overline{\mathbf{y}})^{\mathrm{T}}\mathbf{e}_j \,. \tag{32}$$

For $i \neq j$,

$$\begin{aligned} \mathrm{cov}(a_i, a_j) &= \mathrm{E}[\mathbf{e}_i^{\mathrm{T}}(\mathbf{y} - \overline{\mathbf{y}})(\mathbf{y} - \overline{\mathbf{y}})^{\mathrm{T}}\mathbf{e}_j] = \mathbf{e}_i^{\mathrm{T}}\mathrm{E}[(\mathbf{y} - \overline{\mathbf{y}})(\mathbf{y} - \overline{\mathbf{y}})^{\mathrm{T}}]\mathbf{e}_j \\ &= \mathbf{e}_i^{\mathrm{T}}\mathbf{C}\mathbf{e}_j = \mathbf{e}_i^{\mathrm{T}}\lambda_j\mathbf{e}_j = \lambda_j\mathbf{e}_i^{\mathrm{T}}\mathbf{e}_j = 0\,, \end{aligned} \tag{33}$$

implying zero correlation between $a_i(t)$ and $a_j(t)$. Hence PCA extracts the uncorrelated modes of variability of the data field.

2.5 Example: PCA of the tropical Pacific climate variability

[Book, Sect. 2.1.5]

Tropical Pacific has the *El Niño* phenomenon (Philander, 1990). Every 2-10 years, a sudden warming of the coastal waters occurs off Peru (El Niño). Sometimes a *La Niña* develops, i.e. anomalously cool waters appear in the equatorial Pacific.

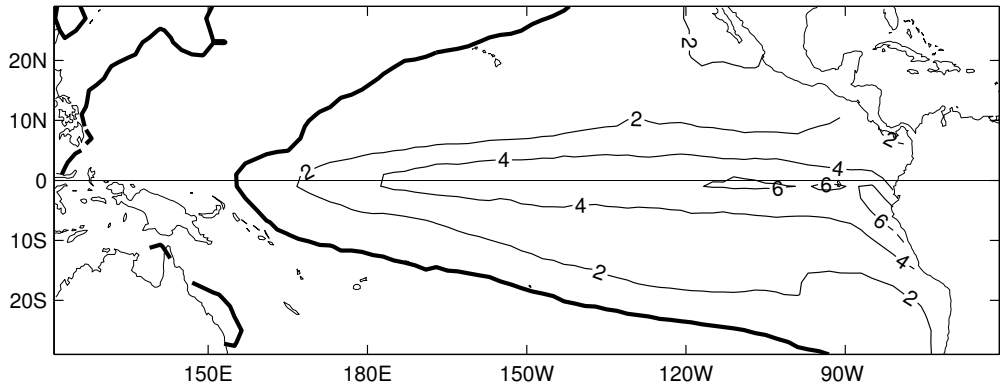
Ocean oscillation coupled to atmospheric oscillation (called the Southern Oscillation). Hence the name ENSO (El Niño-Southern Oscillation).

Take the monthly tropical Pacific sea surface temperature (SST) from NOAA (1950-2000), (with climatological seasonal cycle removed, and smoothed in time by a 3-month moving average). The SST field has 2 spatial dimensions, but can easily be rearranged into the form of $\mathbf{y}(t)$ for the analysis with PCA.

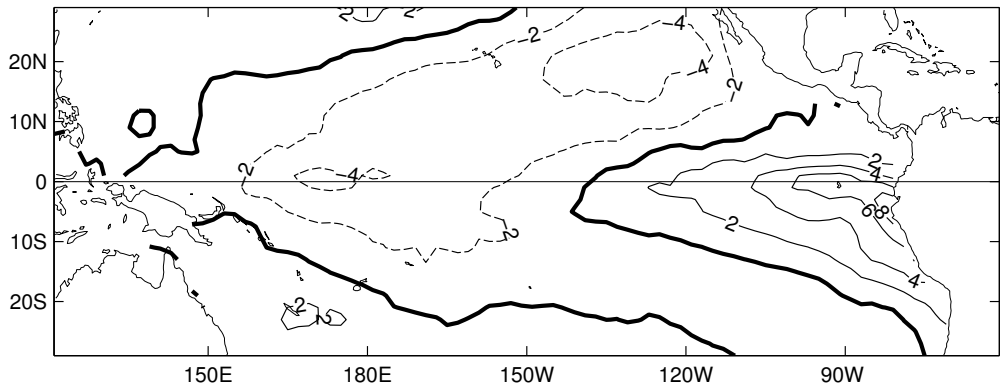
The first six PCA modes account for 51.8%, 10.1%, 7.3%, 4.3%, 3.5% and 3.1%, respectively, of the total SST variance.

The spatial patterns (i.e. the eigenvectors or EOFs) for the first 3 modes are shown below. (Positive contours are indicated by the solid curves, negative contours by dashed curves, and the zero contour by the thick solid curve. The contour unit is 0.01°C. The eigenvectors have been normalized to unit norm.)

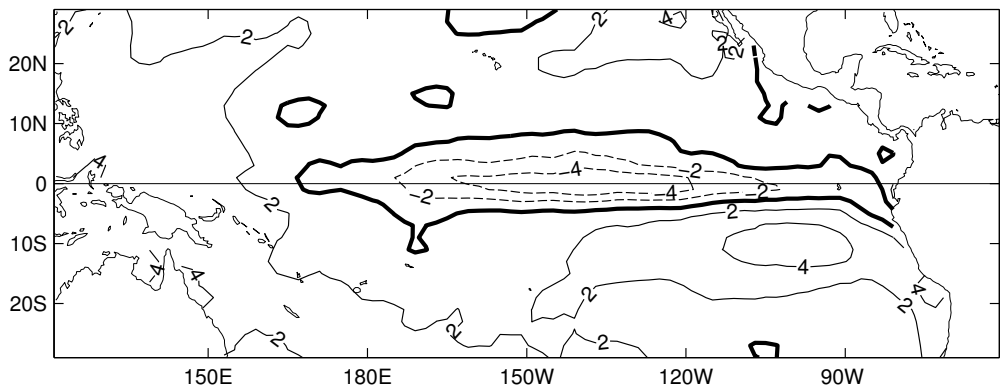
(a) SST mode 1

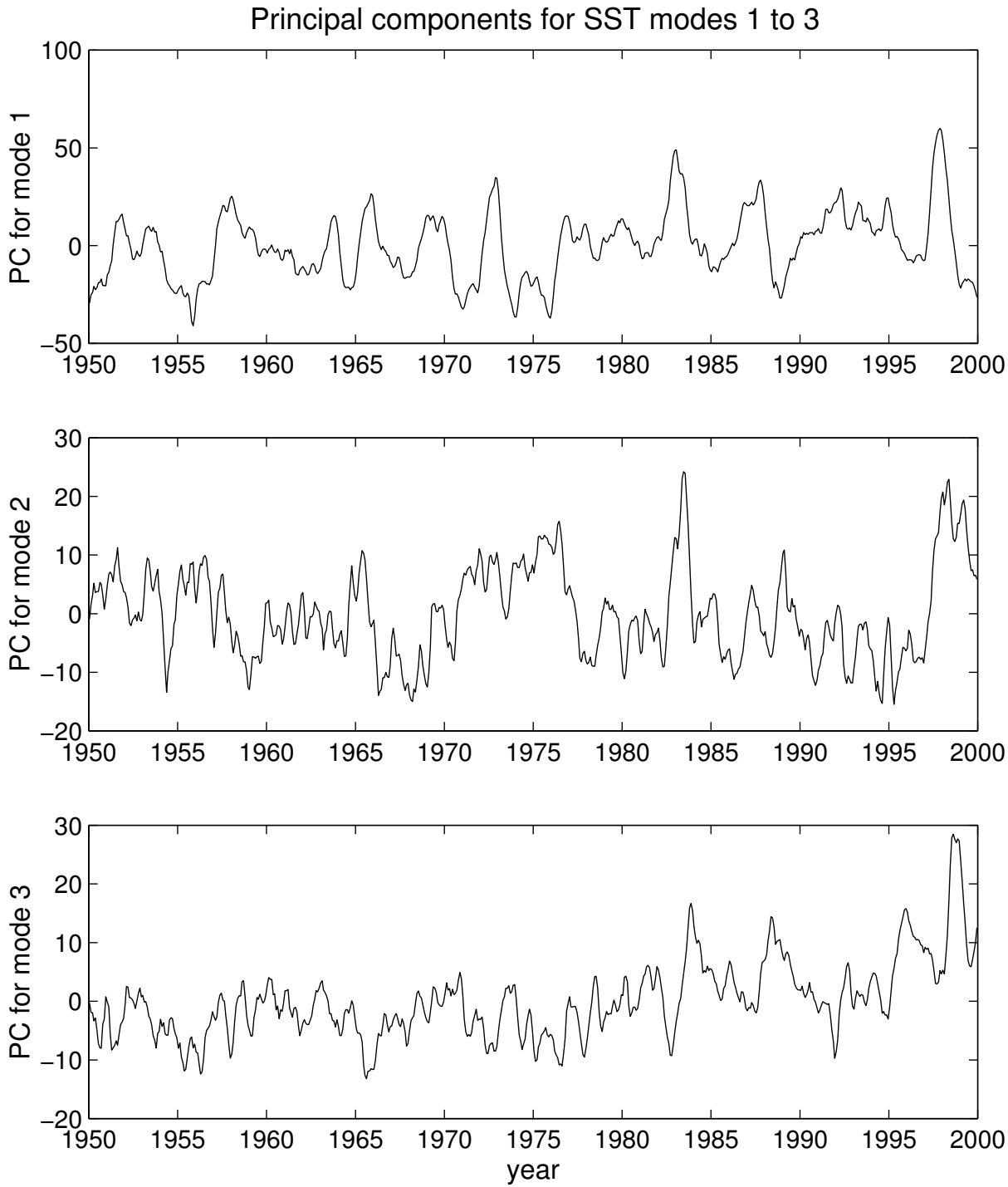


(b) SST mode 2



(c) SST mode 3




Principal components for SST modes 1 to 3
PC for mode 1
PC for mode 2
PC for mode 3
year
100
50
0
−50
30
20
10
0
−10
−20
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000

Example: Sea surface temperature (SST) off Vancouver Island

Satellite SST images for summer 1984–1991 analyzed by PCA (Fang and Hsieh, 1993). Coastal upwelling off Vancouver Island shows up as cool water.

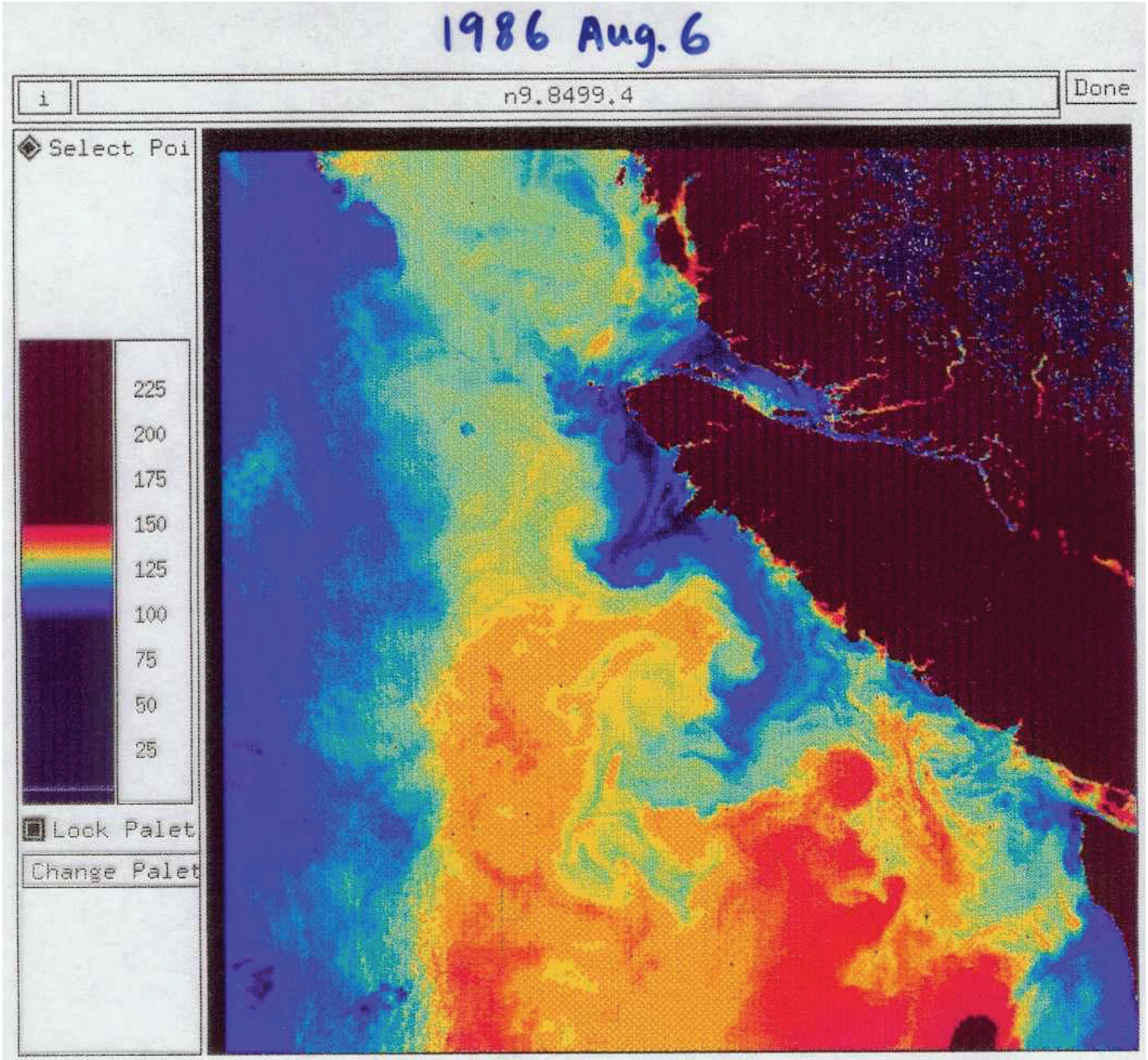


Figure : Cool temperature is shown as blue, warm temperature red.

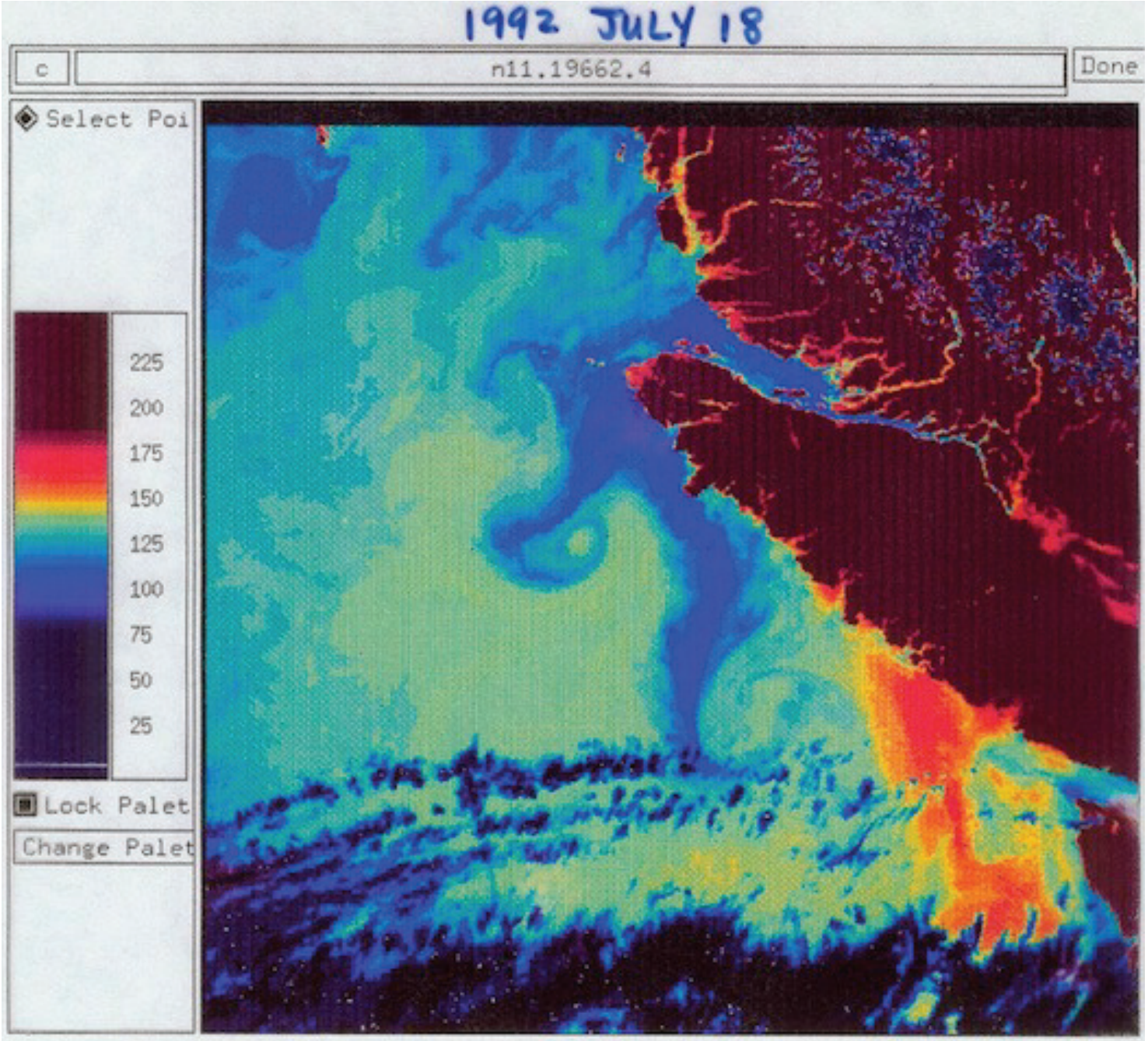


Figure : Clouds to the south.

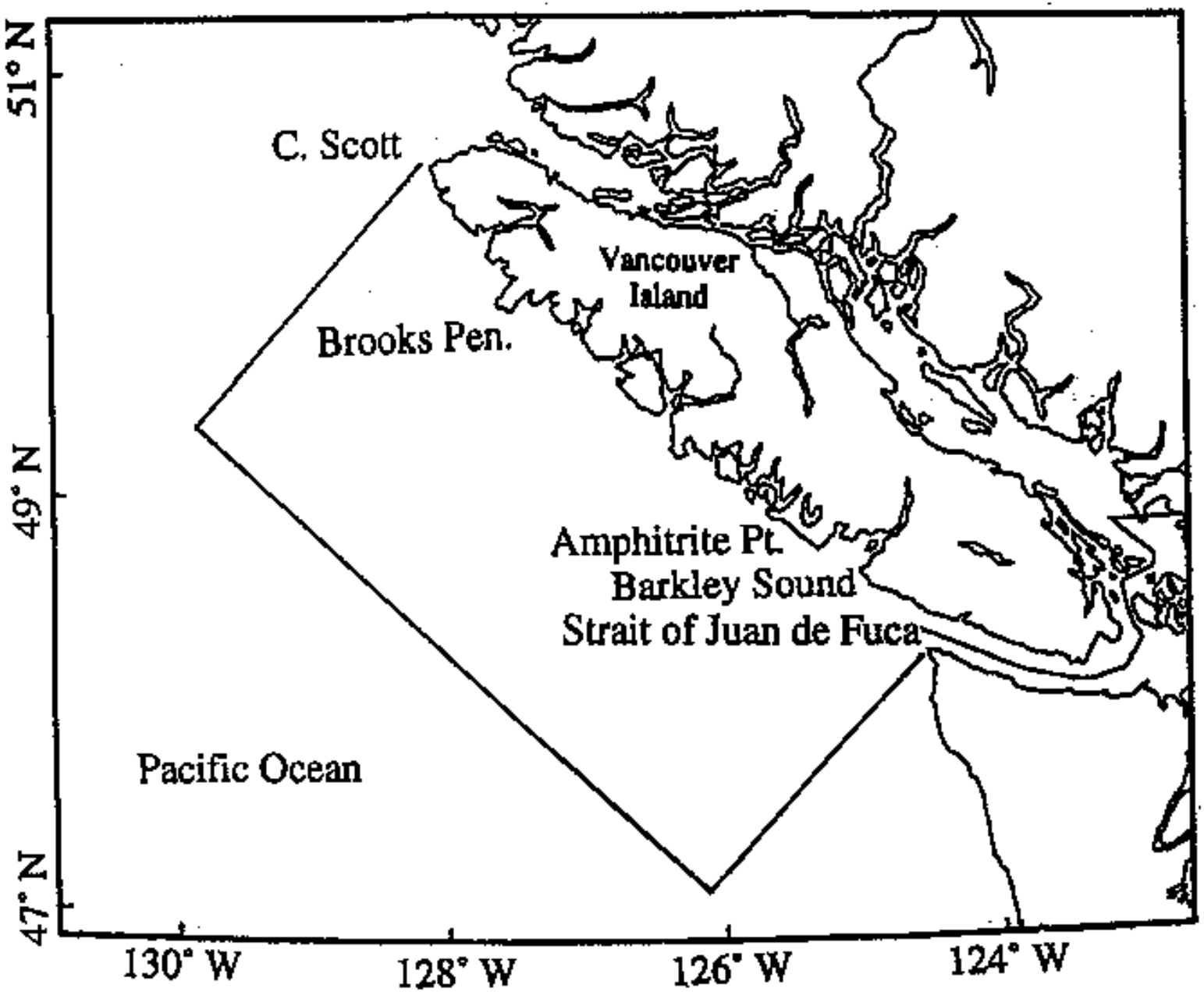


Fig. 1. Study area off the west coast of Vancouver Island.

PCA shows the first 4 modes account for 33%, 12%, 10% and 5% of the variance.

The eigenvectors $\mathbf{e}_1^{\mathrm{T}}, \ldots, \mathbf{e}_4^{\mathrm{T}}$ give 4 spatial patterns.

The PC $(z_1, \ldots, z_4)$ indicate the strength of the particular spatial pattern at a given time.

Note the mean of SST at each pixel was not subtracted prior to performing the PCA. Hence first spatial mode $(\mathbf{e}_1^{\mathrm{T}})$ dominated by the mean SST pattern. More meaningful if the mean SST pattern was removed prior to PCA.

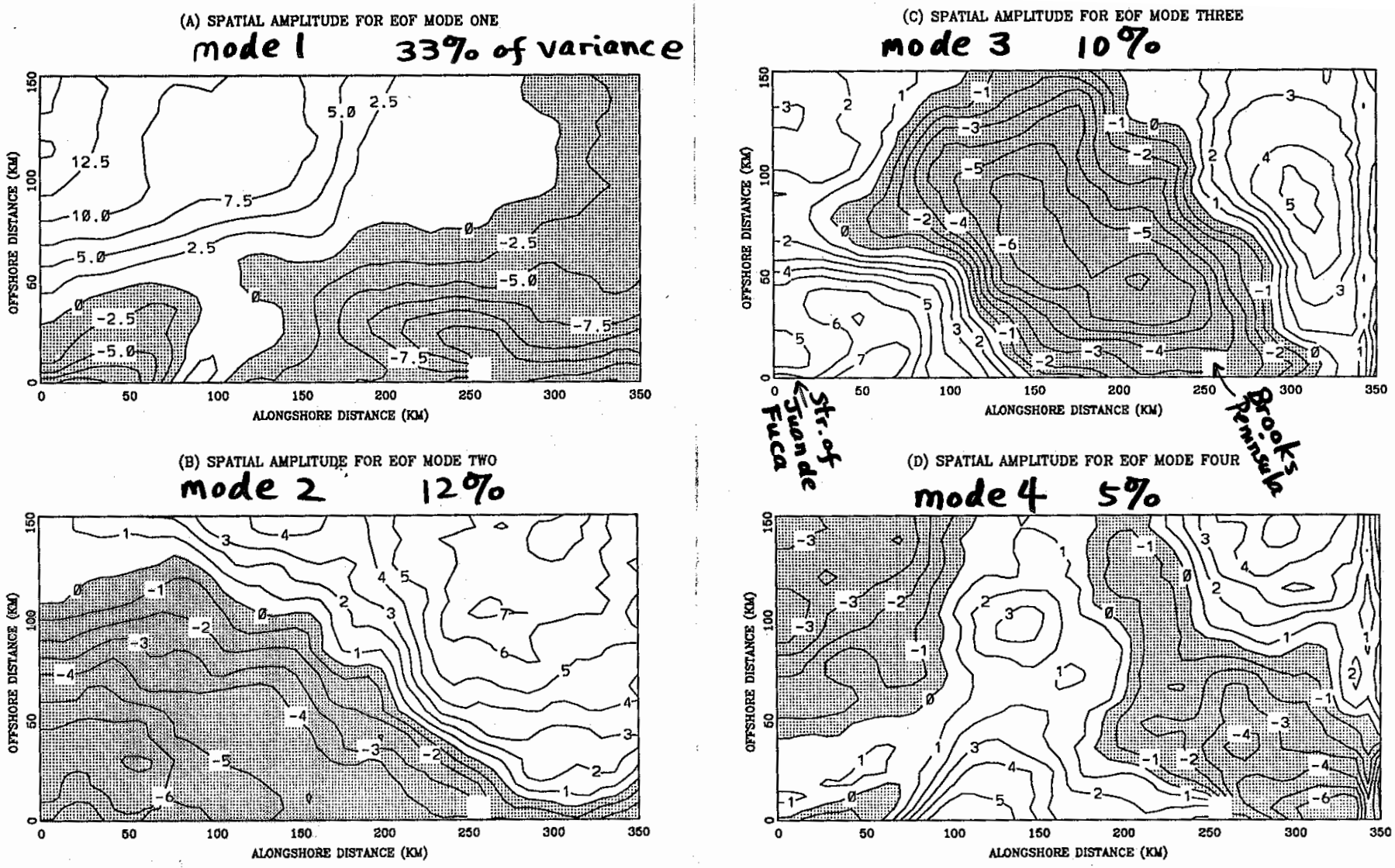


Fig. 3. Spatial amplitude patterns for the SST gradient EOF modes 1 to 4. The spatial domain extends 150 km offshore and 350 km alongshore. In the alongshore direction, the mouth of the Juan de Fuca Strait stretches from 0 to 30 km, Barkley sound from 70 to 90 km, and Brooks peninsula protrudes from the coast at 250 km alongshore.

Fang & Hsieh JGR 1993 98:14391-14400

Fig. 3. (continued)

asymptotic theory for large data sets can be used
dorfer, 1988, p. 204-205]. As the data set in this
arge, we used the the asymptotic method to test
icance of the EOFs. From *Preisendorfer* [1988,

increases monotonically from June to Septe
the intensification of the mode 1 pattern (Fi
summer progresses. Mode 2 also intensif
peaking in August. Mode 3 shows a decline
amplitude in August and especially in Septe

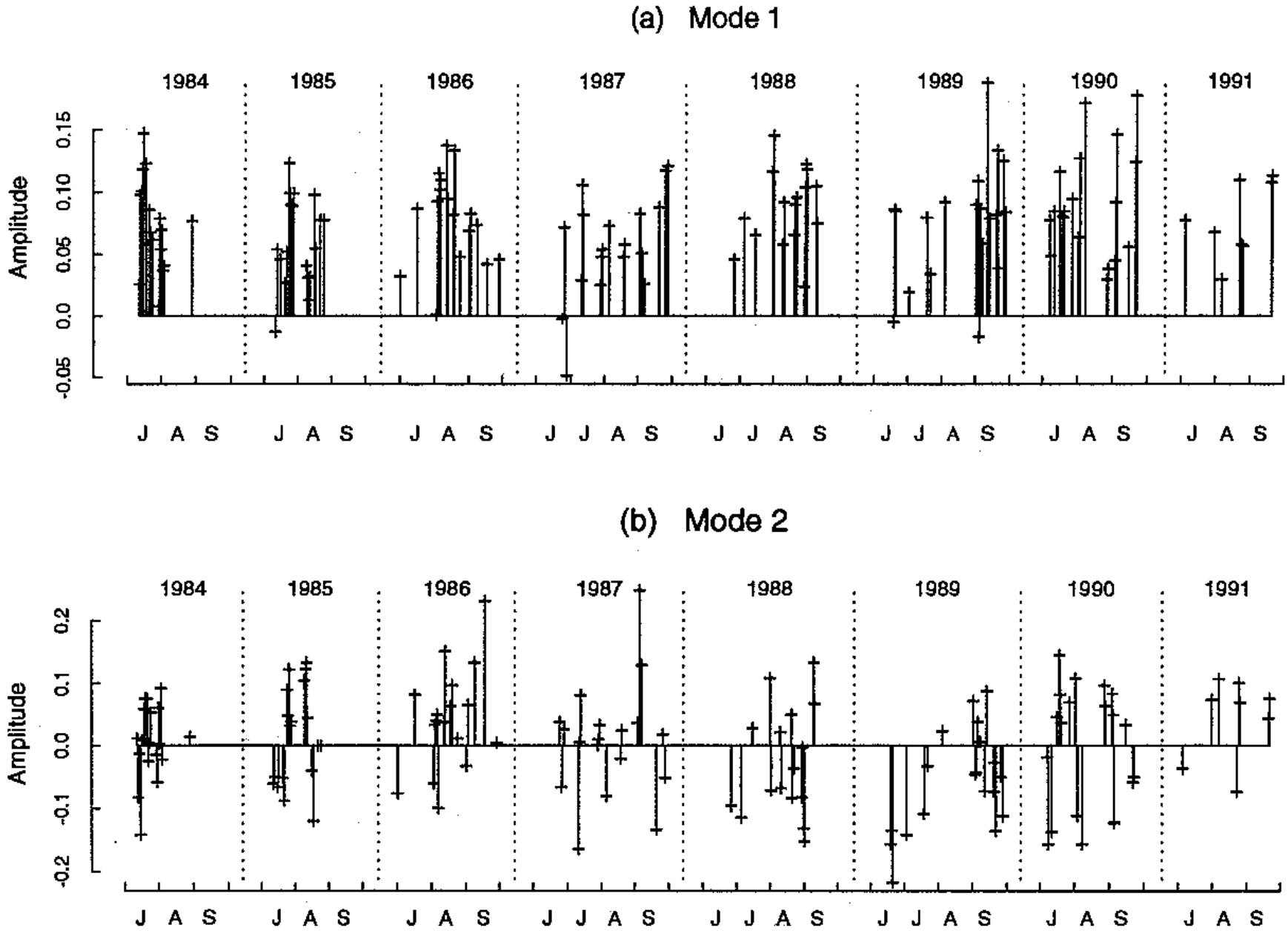


Figure : Principal components 1 and 2.

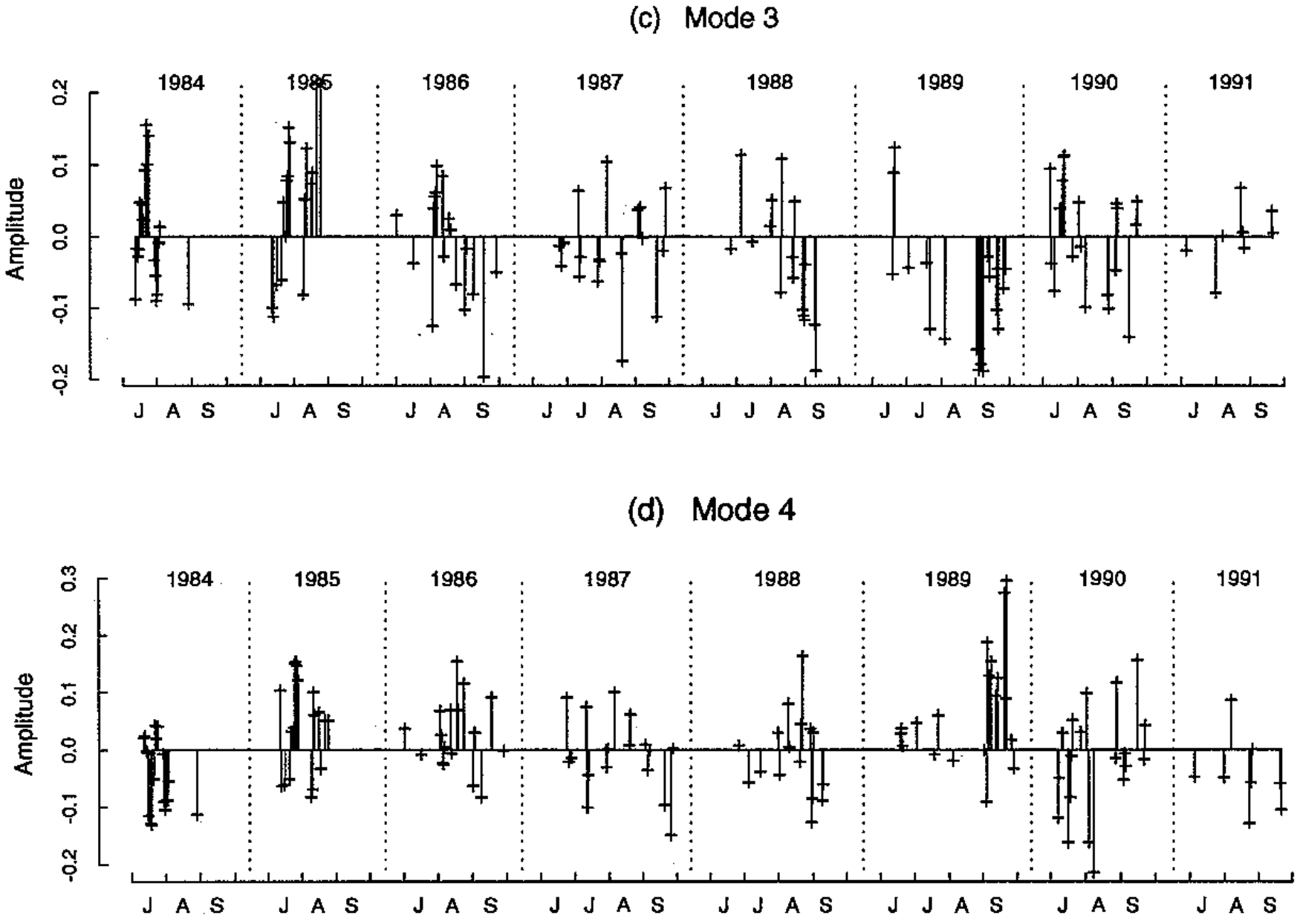


Figure : Principal components 3 and 4

Q2: From the principal components during 1984–1991, determine which year the strongest plume of cool water was found to extend offshore from (a) Brooks Peninsula and (b) Cape Scott (at the northern tip of Vancouver Island)?

PCA functions in Matlab

`www.mathworks.com/help/toolbox/stats/princomp.html`
[eigenvectors, PCs, eigenvalues] = princomp(X)

`www.mathworks.com/help/toolbox/stats/pcacov.html`
(user provides data covariance or correlation matrix).

References:

Fang, W. and Hsieh, W. W. (1993). Summer sea Surface temperature variability off Vancouver Island from satellite data. *Journal of Geophysical Research*, 98(C8):14391–14400.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441.

Pearson, K. (1901). On lines and planes of closest fit to system of points in space. *Philosophical Magazine, Ser. 6*, 2:559–572.

Philander, S. G. (1990). *El Niño, La Niña, and the Southern Oscillation*. Academic Pr., San Diego.

Strang, G. (2005). *Linear Algebra and Its Applications*. Cole Brooks.