Lecture 18

Varimax Factors
and
Empircal Orthogonal Functions

Syllabus

Lecture 01	Describing Inverse Problems
Lecture 02	Probability and Measurement Error, Part 1
Lecture 03	Probability and Measurement Error, Part 2
Lecture 04	The L_2 Norm and Simple Least Squares
Lecture 05	A Priori Information and Weighted Least Squared
Lecture 06	Resolution and Generalized Inverses
Lecture 07	Backus-Gilbert Inverse and the Trade Off of Resolution and Variance
Lecture 08	The Principle of Maximum Likelihood
Lecture 09	Inexact Theories
Lecture 10	Nonuniqueness and Localized Averages
Lecture 11	Vector Spaces and Singular Value Decomposition
Lecture 12	Equality and Inequality Constraints
Lecture 13	L_1 , L_∞ Norm Problems and Linear Programming
Lecture 14	Nonlinear Problems: Grid and Monte Carlo Searches
Lecture 15	Nonlinear Problems: Newton's Method
Lecture 16	Nonlinear Problems: Simulated Annealing and Bootstrap Confidence Intervals
Lecture 17	Factor Analysis
Lecture 18	**Varimax Factors, Empircal Orthogonal Functions**
Lecture 19	Backus-Gilbert Theory for Continuous Problems; Radon's Problem
Lecture 20	Linear Operators and Their Adjoints
Lecture 21	Fréchet Derivatives
Lecture 22	Exemplary Inverse Problems, incl. Filter Design
Lecture 23	Exemplary Inverse Problems, incl. Earthquake Location
Lecture 24	Exemplary Inverse Problems, incl. Vibrational Problems

Purpose of the Lecture

Choose Factors Satisfying A Priori Information of Spikiness (varimax factors)

Use Factor Analysis to Detect Patterns in data (EOF's)

Part 1: Creating Spiky Factors

can we find “better” factors

that those returned by `svd()`

?

mathematically

$\mathbf{S} = \mathbf{CF} = \mathbf{C'}\,\mathbf{F'}$

with $\mathbf{F'} = \mathbf{M}\,\mathbf{F}$ and $\mathbf{C'} = \mathbf{M}^{-1}\,\mathbf{C}$

where $\mathbf{M}$ is any P×P matrix with an inverse

must rely on prior information to choose $\mathbf{M}$

one possible type of prior information

factors should contain mainly just a few elements

example of rock and minerals

rocks contain minerals
minerals contain elements

Mineral	Composition
Quartz	SiO_2
Rutile	TiO_2
Anorthite	$CaAl_2Si_2O_8$
Fosterite	Mg_2SiO_4

example of rock and minerals

rocks contain minerals

factors

minerals contain elements

Mineral	Composition
Quartz	SiO_2
Rutile	TiO_2
Anorthite	$CaAl_2Si_2O_8$
Fosterite	Mg_2SiO_4

most of these minerals are “simple” in the sense that each contains just a few elements

spiky factors

factors containing mostly just a few elements

How to quantify spikiness?

variance as a measure of spikiness

$$\sigma_d^2 = \frac{1}{N}\left(\sum_{i=1}^{N}(d_i - \bar{d})^2\right) = \frac{1}{N^2}\left(N\sum_{i=1}^{N} d_i^2 - \left(\sum_{i=1}^{N} d_i\right)^2\right)$$

modification for factor analysis

$$\sigma_f^2 = \frac{1}{M^2}\left(M \sum_{i=1}^{M} f_i^{\,4} - \left(\sum_{i=1}^{M} f_i^{\,2} \right)^2 \right)$$

modification for factor analysis

$$\sigma_f^2 = \frac{1}{M^2}\left(M\sum_{i=1}^{M} f_i^4 - \left(\sum_{i=1}^{M} f_i^2\right)^2 \right)$$

depends on the square, so positive and negative values are treated the same

$$\mathbf{f}^{(1)} = [1, 0, 1, 0, 1, 0]^T$$

is much spikier than

$$\mathbf{f}^{(2)} = [1, 1, 1, 1, 1, 1]^T$$

$$\mathbf{f}^{(2)} = [1, 1, 1, 1, 1, 1]^T$$

is just as spiky as

$$\mathbf{f}^{(3)} = [1, -1, 1, -1, -1, 1]^T$$

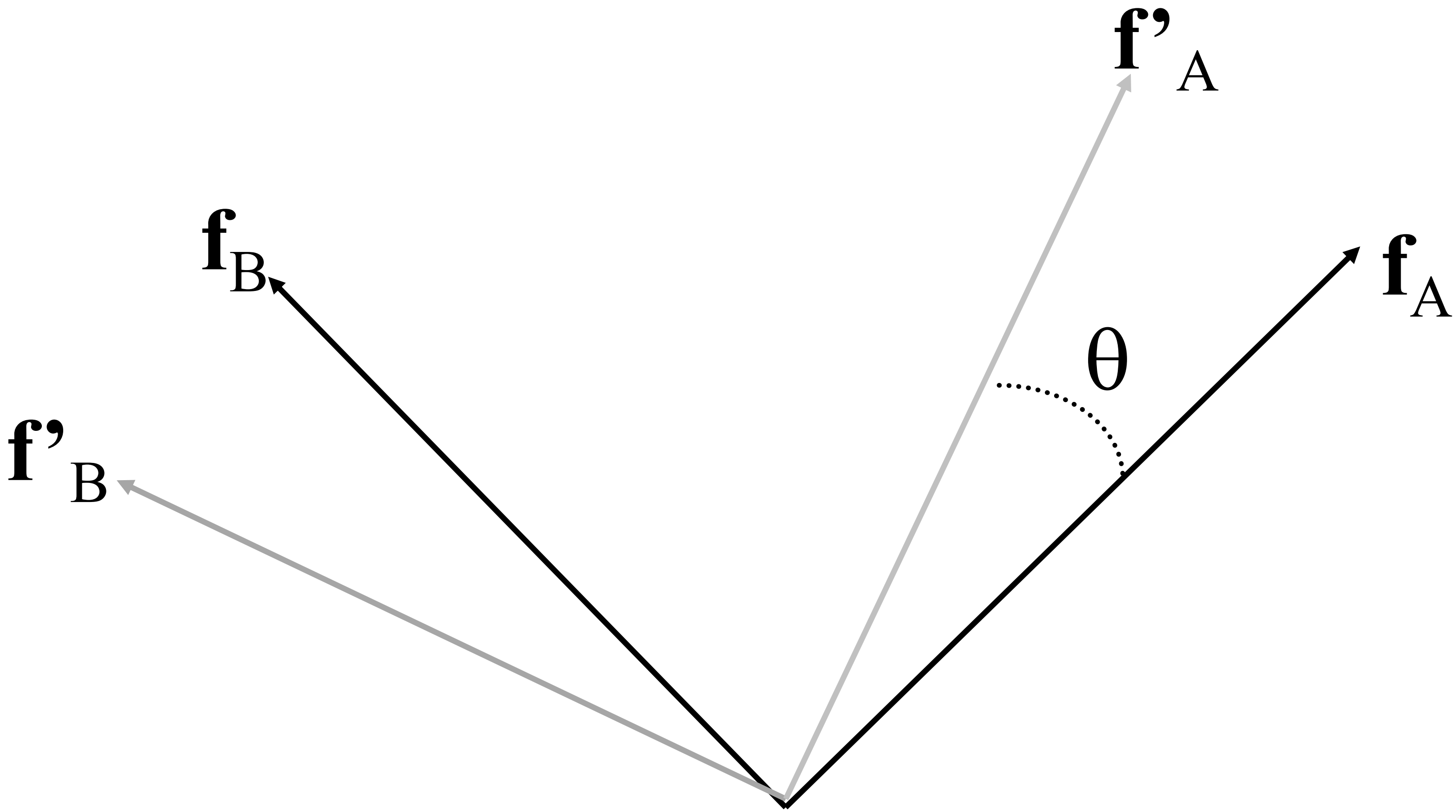
“varimax” procedure

find spiky factors without changing P

start with P `svd()` factors

rotate pairs of them in their plane by angle θ

to maximize the overall spikiness



f'A
fB
fA
θ
f'B

determine θ by maximizing

$\Phi(\theta)=M^2(\sigma_{fA}{}^2+\sigma_{fB}{}^2)$ with respect to θ

after tedious trig the solution can be shown to be

$$\theta = ¼\ \tan^{-1}\frac{2M\sum_i u_i v_i - \sum_i u_i \sum_i v_i}{M\sum_i (u_i^2 - v_i^2) - ((\sum_i u_i)^2 - (\sum_i v_i)^2)}$$

with

$$u_i = \left(f_i^A\right)^2 - (f_i^B)^2 \ \ and \ \ v_i = 2f_i^A f_i^B$$

and the new factors are

$$\begin{bmatrix} \mathbf{f}_1^T \\ \mathbf{f}_2^T \\ \cos(\theta)\mathbf{f}_3^T + \sin(\theta)\mathbf{f}_5^T \\ \mathbf{f}_4^T \\ -\sin(\theta)\mathbf{f}_3^T + \cos(\theta)\mathbf{f}_5^T \\ \mathbf{f}_6^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(\theta) & 0 & \sin(\theta) & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -\sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{f}_1^T \\ \mathbf{f}_2^T \\ \mathbf{f}_3^T \\ \mathbf{f}_4^T \\ \mathbf{f}_5^T \\ \mathbf{f}_6^T \end{bmatrix}$$

$$\text{or} \quad \mathbf{F}' = \mathbf{MF}$$

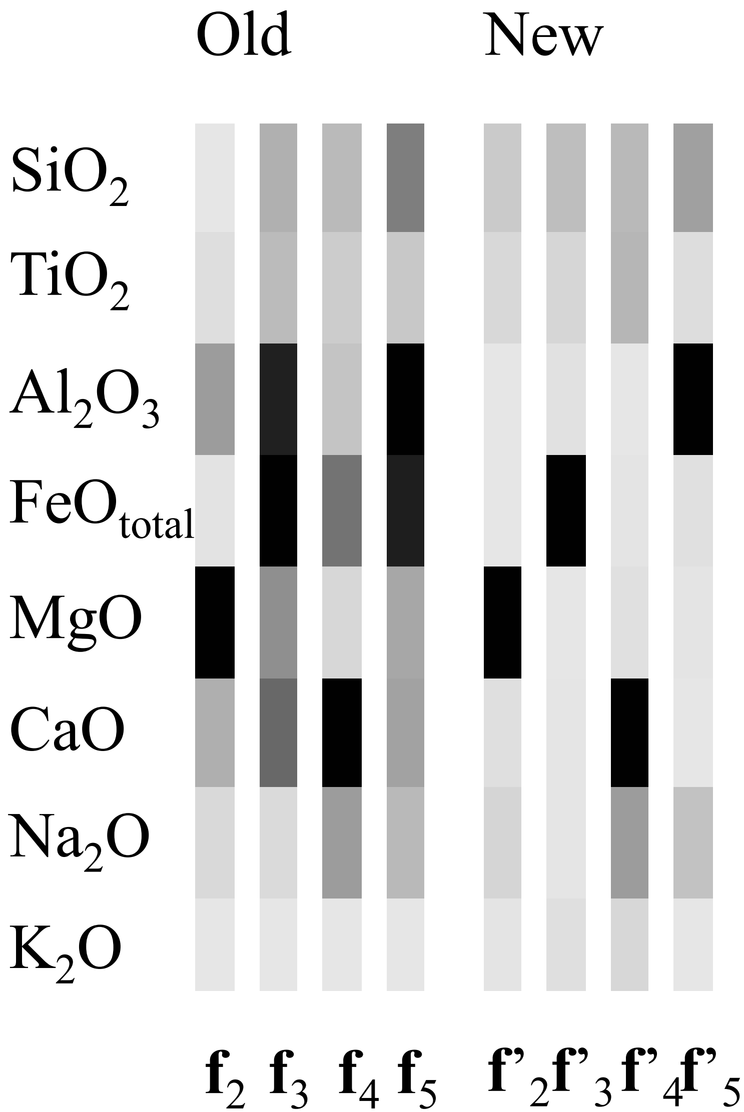
in this example A=3 and B=5

now one repeats for every pair of factors

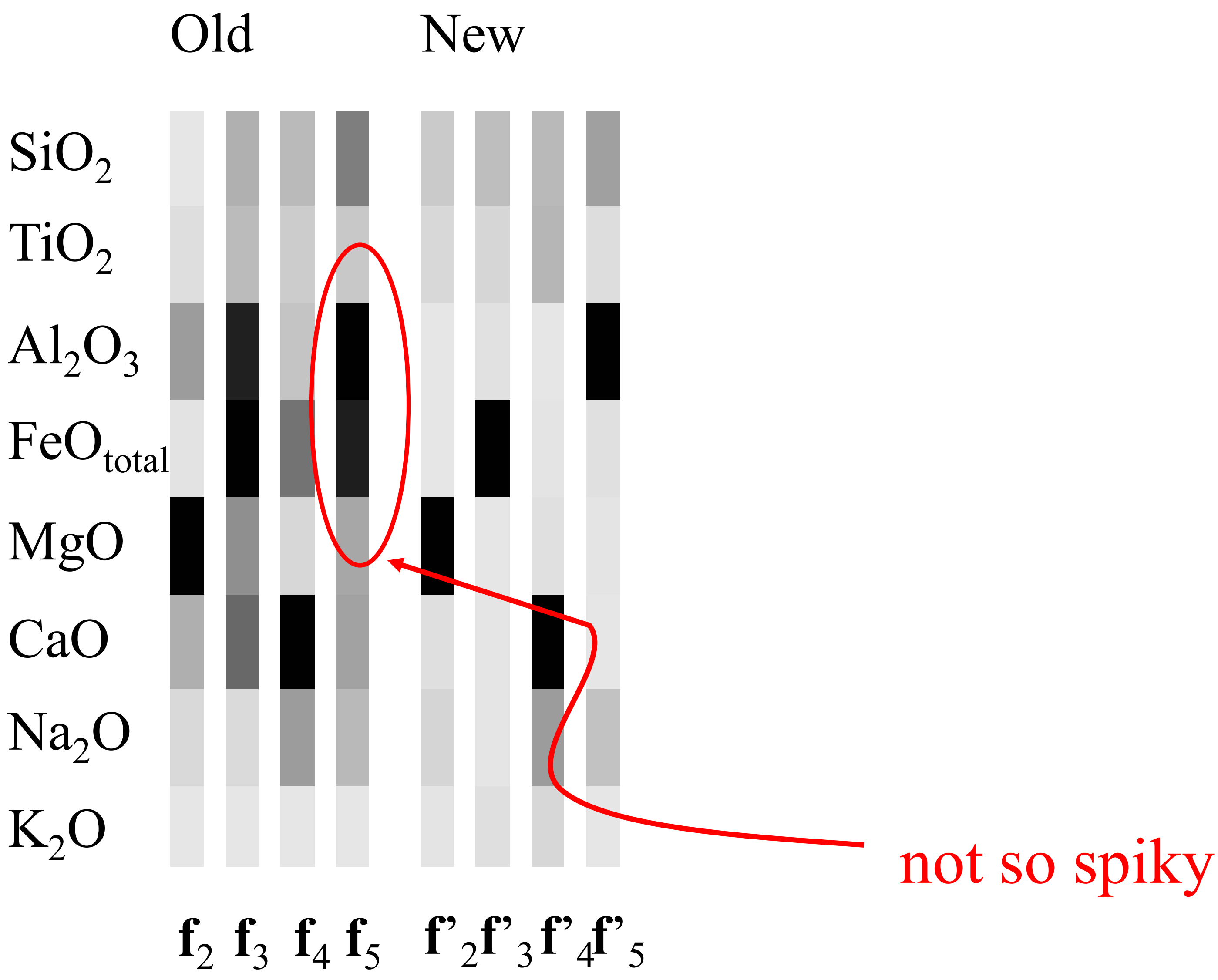
and then iterates the whole process several times

until the whole set of factors is as spiky as possible

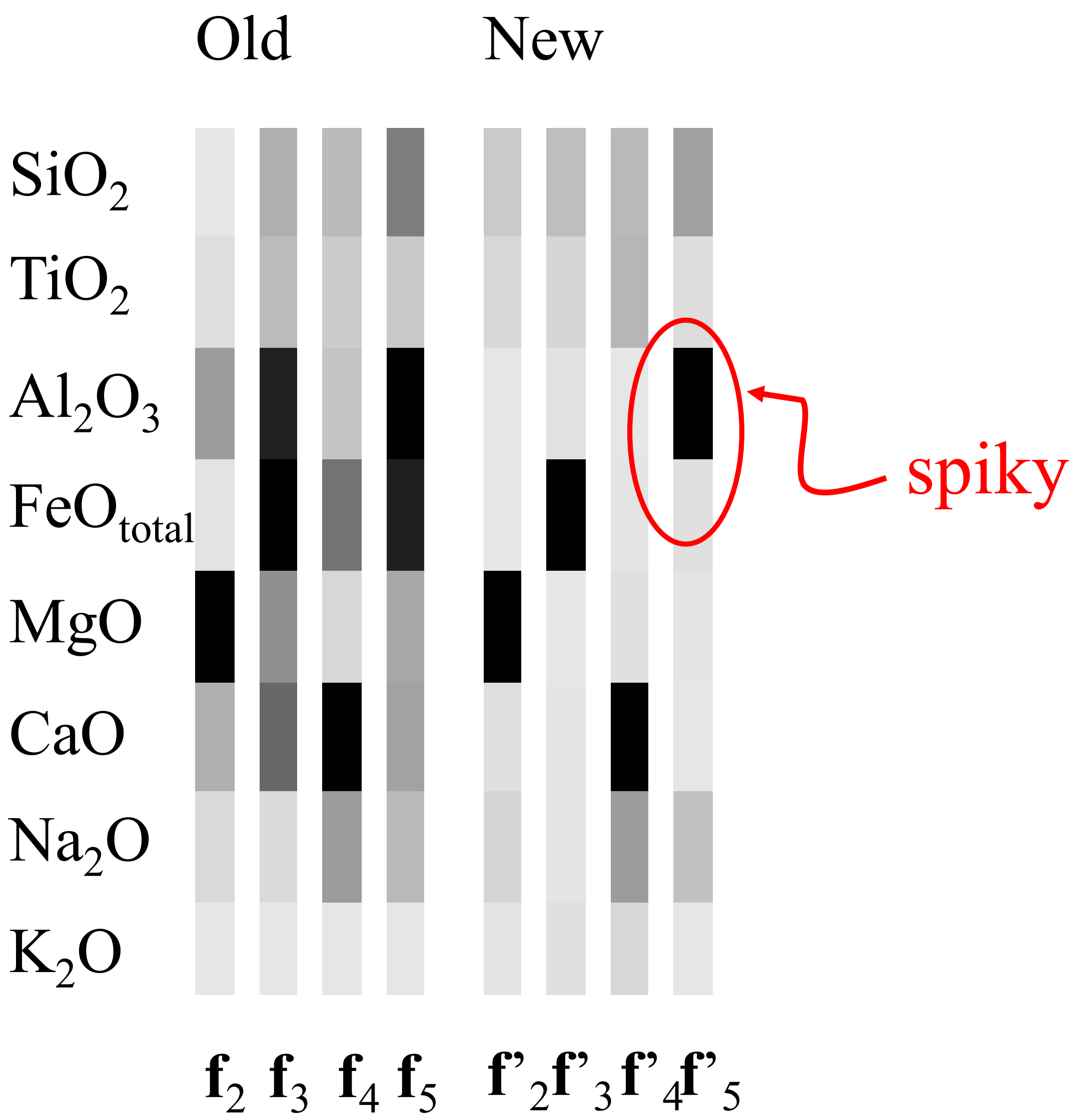
example: Atlantic Rock dataset



example: Atlantic Rock dataset



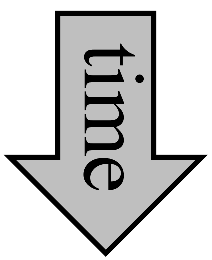
example: Atlantic Rock dataset



Part 2: Empirical Orthogonal Functions

row number in the sample matrix could be meaningful

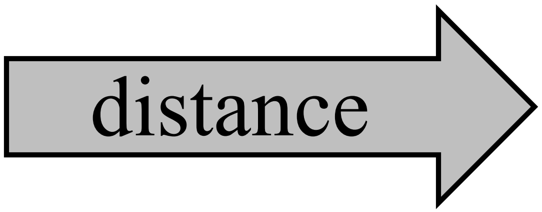
example: samples collected at a succession of times



$$\mathbf{S} = \begin{bmatrix} \text{element 1 in sample 1} & \text{element 2 in sample 1} & \cdots & \text{element } M \text{ in sample 1} \\ \cdots & \cdots & \cdots & \cdots \\ \text{element 1 in sample } N & \text{element 2 in sample } N & \cdots & \text{element } M \text{ in sample } N \end{bmatrix}$$

column number in the sample matrix could be meaningful

example: concentration of the same chemical element at a sequence of positions



$$\mathbf{S} = \begin{bmatrix} \text{element 1 in sample 1} & \text{element 2 in sample 1} & \cdots & \text{element } M \text{ in sample 1} \\ \cdots & \cdots & \cdots & \cdots \\ \text{element 1 in sample } N & \text{element 2 in sample } N & \cdots & \text{element } M \text{ in sample } N \end{bmatrix}$$

$$\mathbf{S} = \mathbf{C}\mathbf{F}$$

becomes

$$s(x_j, t_i) = \sum_{k=1}^{P} C_k(t_i)\, f_k(x_j)$$

$$\mathbf{S} = \mathbf{CF}$$

becomes

distance dependence

time dependence

$$s(x_j, t_i) = \sum_{k=1}^{P} C_k(t_i)\, f_k(x_j)$$

$$\mathbf{S} = \mathbf{CF}$$

each factor:
a spatial pattern
of variability of
the element

becomes

each loading: a
temporal pattern of
variability of the
corresponding factor

$$s(x_j, t_i) = \sum_{k=1}^{P} C_k(t_i)\, f_k(x_j)$$

S = CF

becomes

there are P patterns and they are sorted into order of importance

$$s(x_j, t_i) = \sum_{k=1}^{P} C_k(t_i)\, f_k(x_j)$$

$$\mathbf{S} = \mathbf{CF}$$

becomes

factors now called EOF's (empirical orthogonal functions)

$$s(x_j, t_i) = \sum_{k=1}^{P} C_k(t_i)\, f_k(x_j)$$

example 1

hypothetical mountain profiles

what are the most important spatial patterns
that characterize mountain profiles

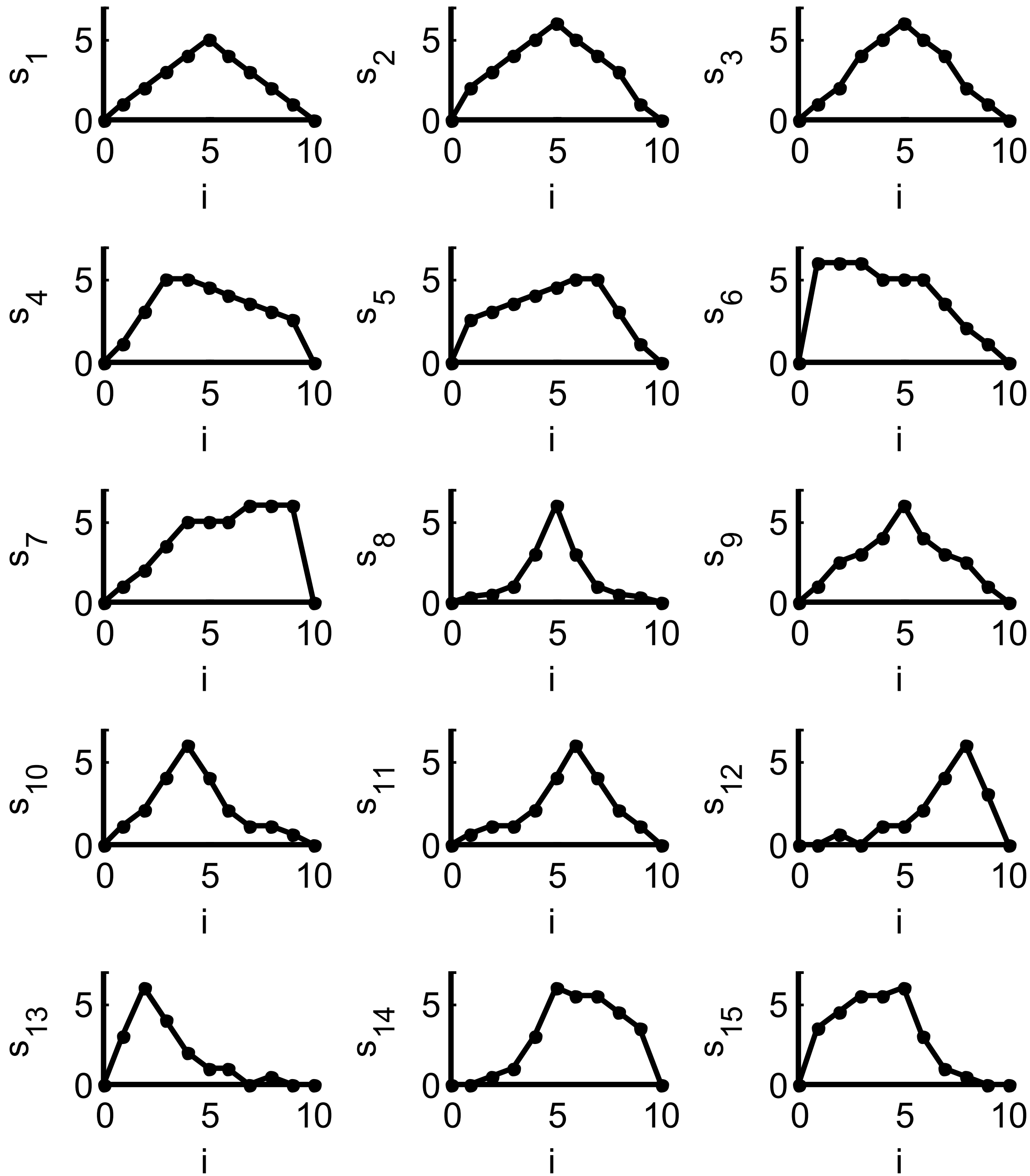
this problem has space but not time

$$s(x_j, i) = \Sigma_{k=1}^{p} c_{ki} f^{(k)}(x_j)$$

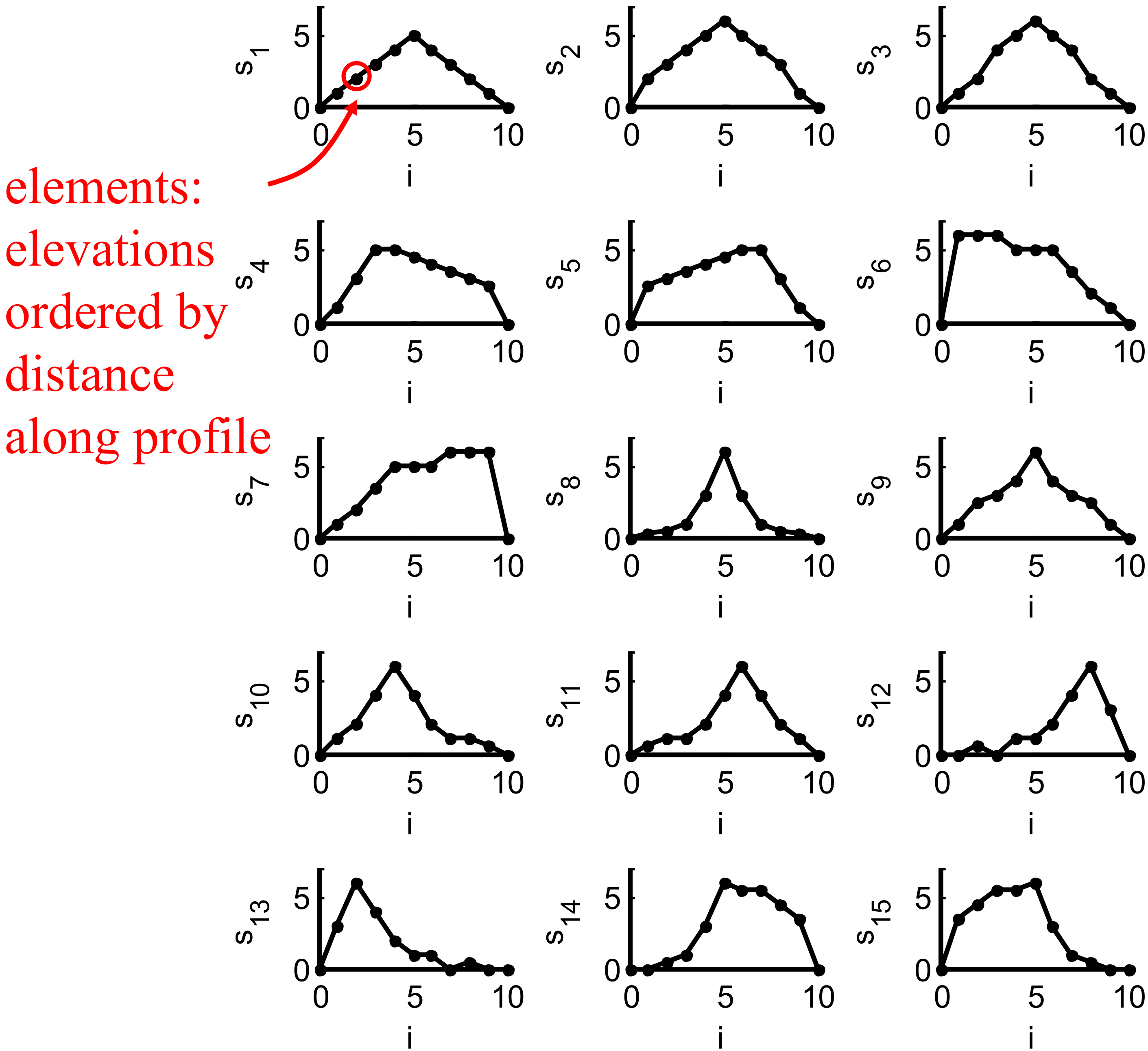
this problem has space but not time

$$s(x_j, i) = \Sigma_{k=1}^{p}\, c_{ki}\, f^{(k)}(x_j)$$

factors are spatial
patterns that add
together to make
mountain profiles

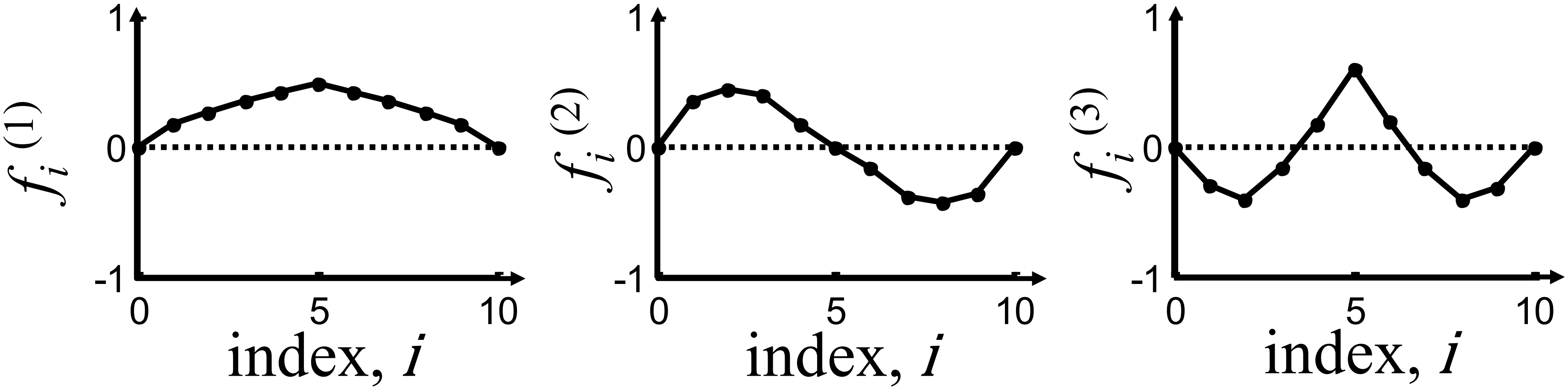


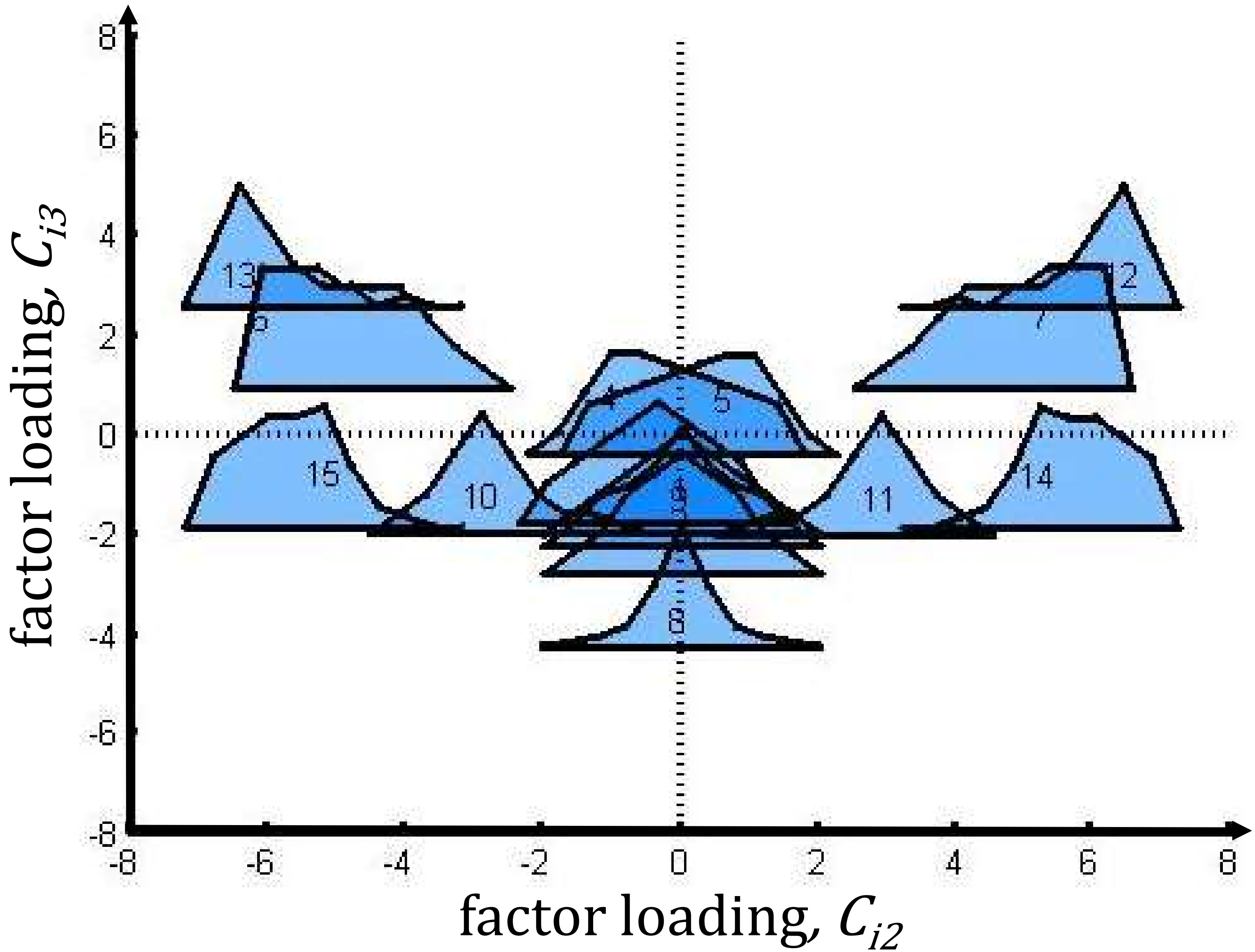
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12
s13
s14
s15
0
5
10
i


elements:
elevations
ordered by
distance
along profile
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12
s13
s14
s15
0
5
10
i

$\lambda_1 = 38$ $\lambda_2 = 13$ $\lambda_3 = 7$

EOF 1 EOF 2 EOF 3



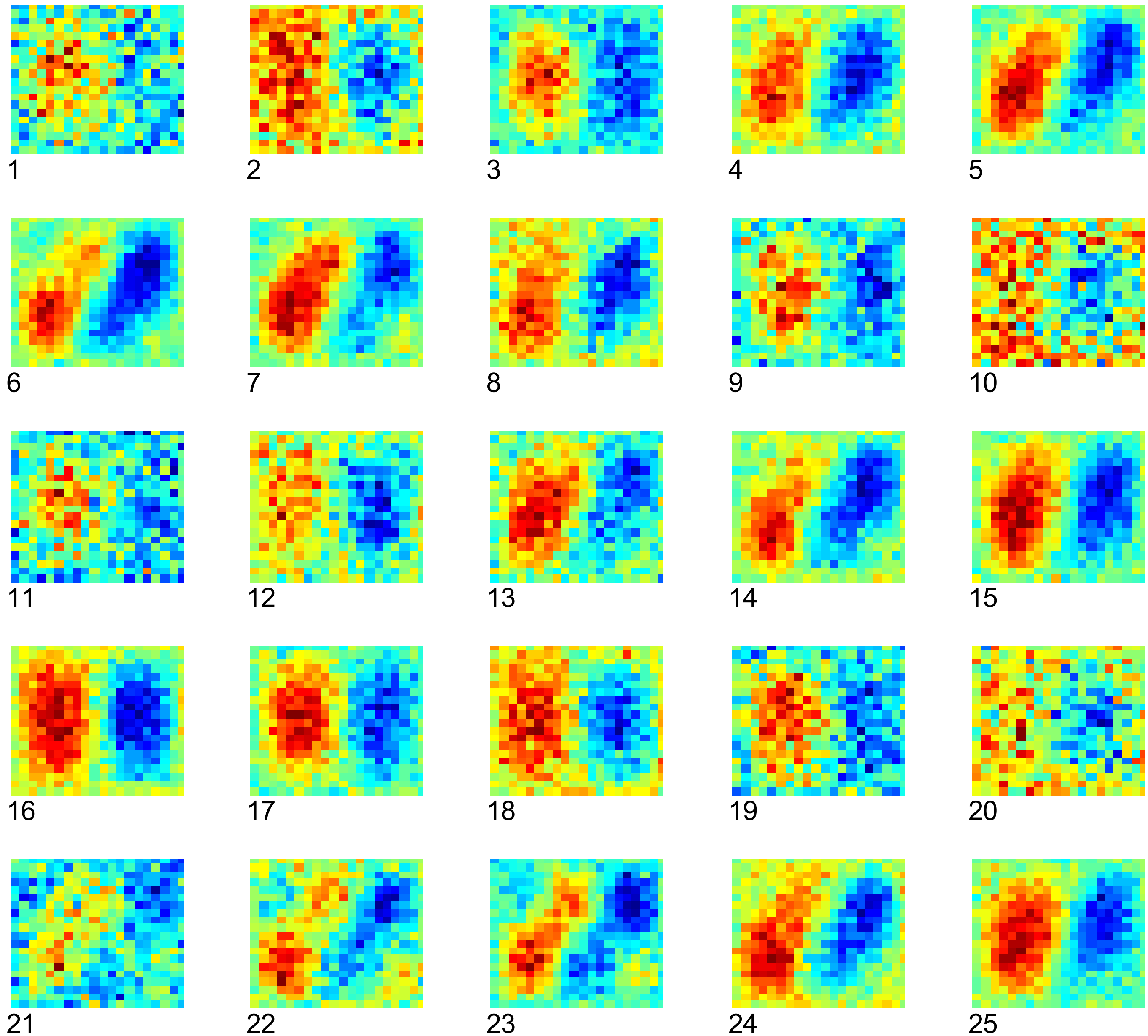


factor loading, C_{i3}
factor loading, C_{i2}
8
6
4
2
0
-2
-4
-6
-8
-8
-6
-4
-2
0
2
4
6
8
13
12
7
5
15
10
11
14
8

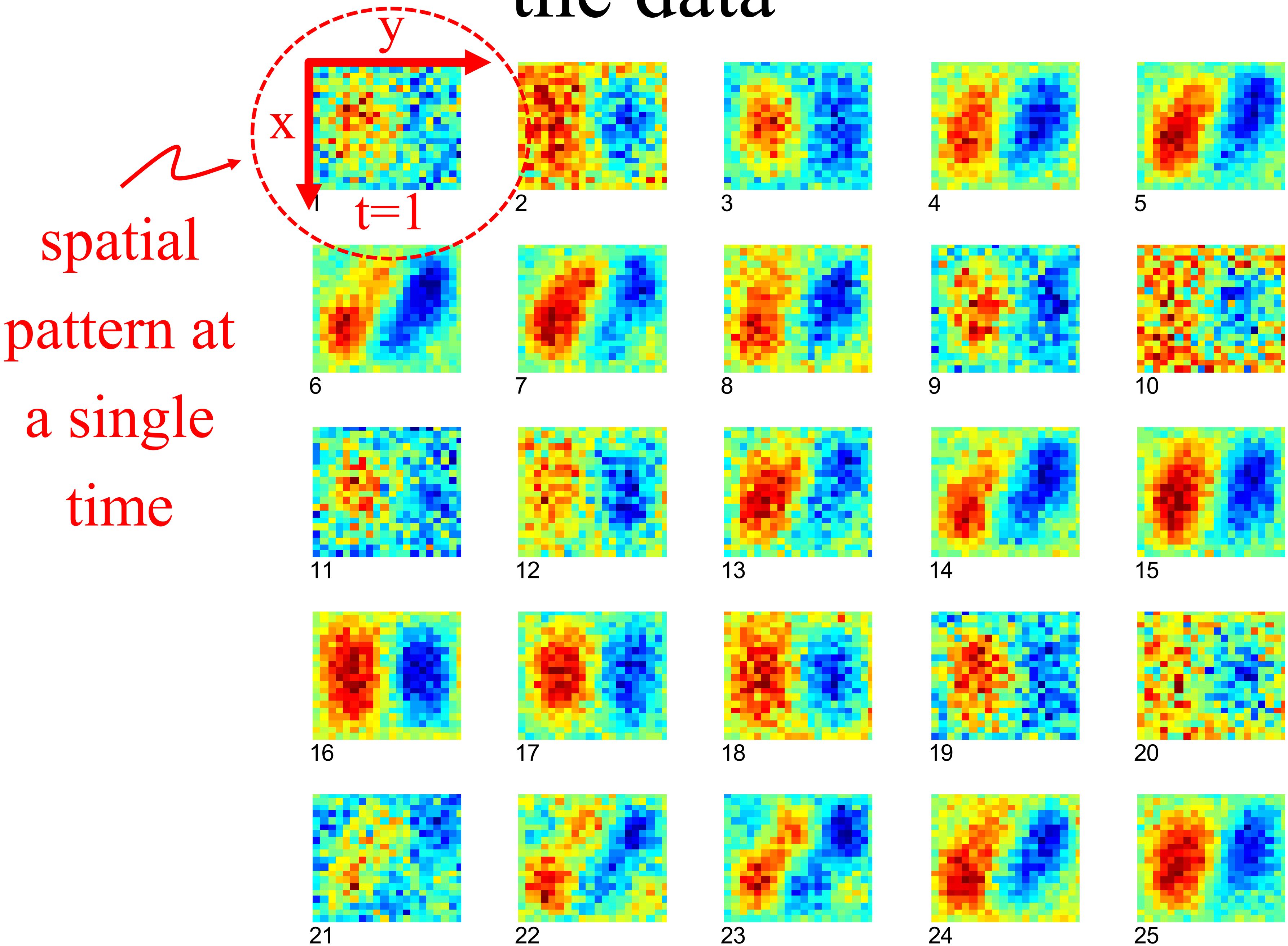
example 2

spatial-temporal patterns
(synthetic data)

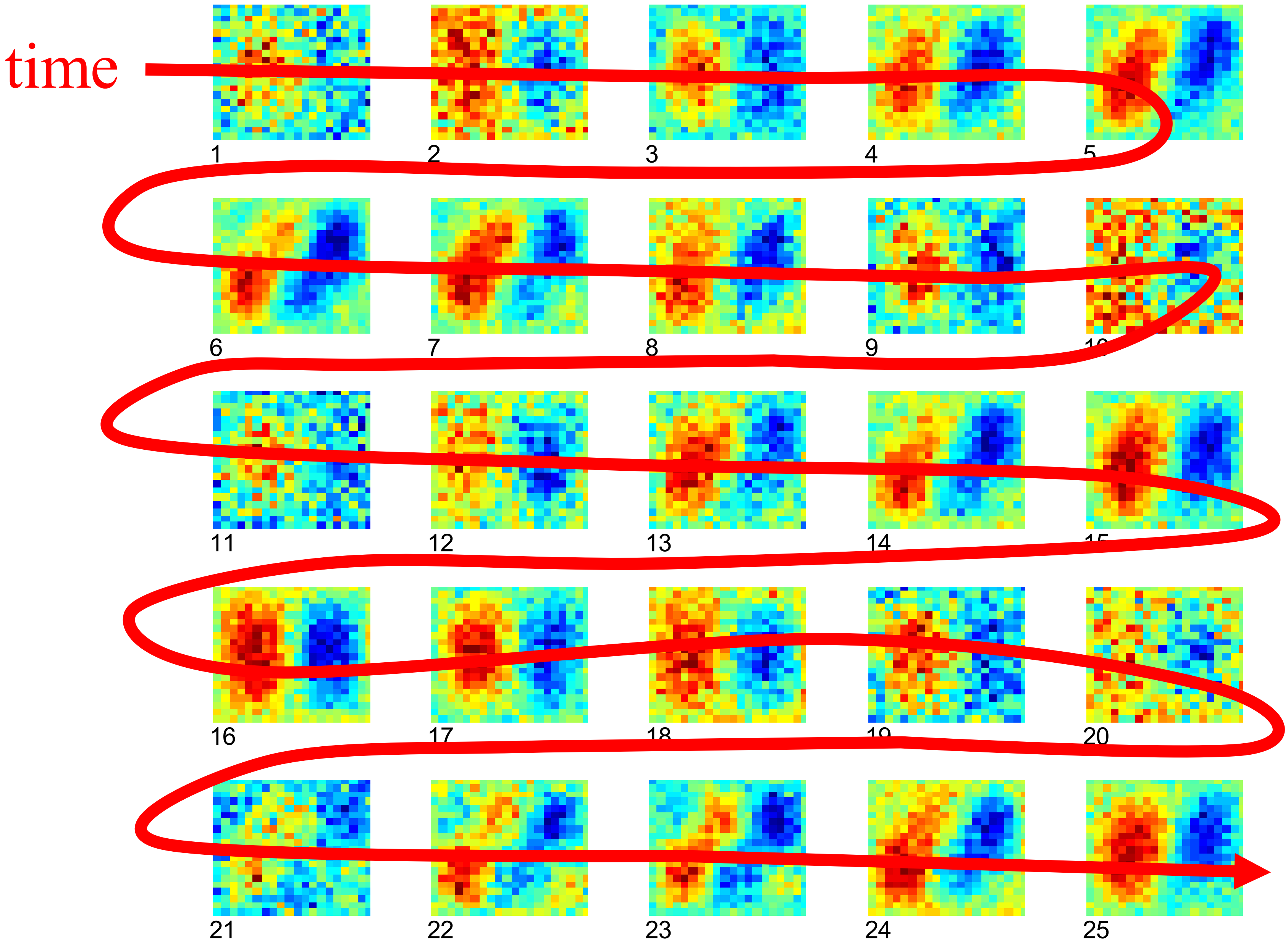
the data



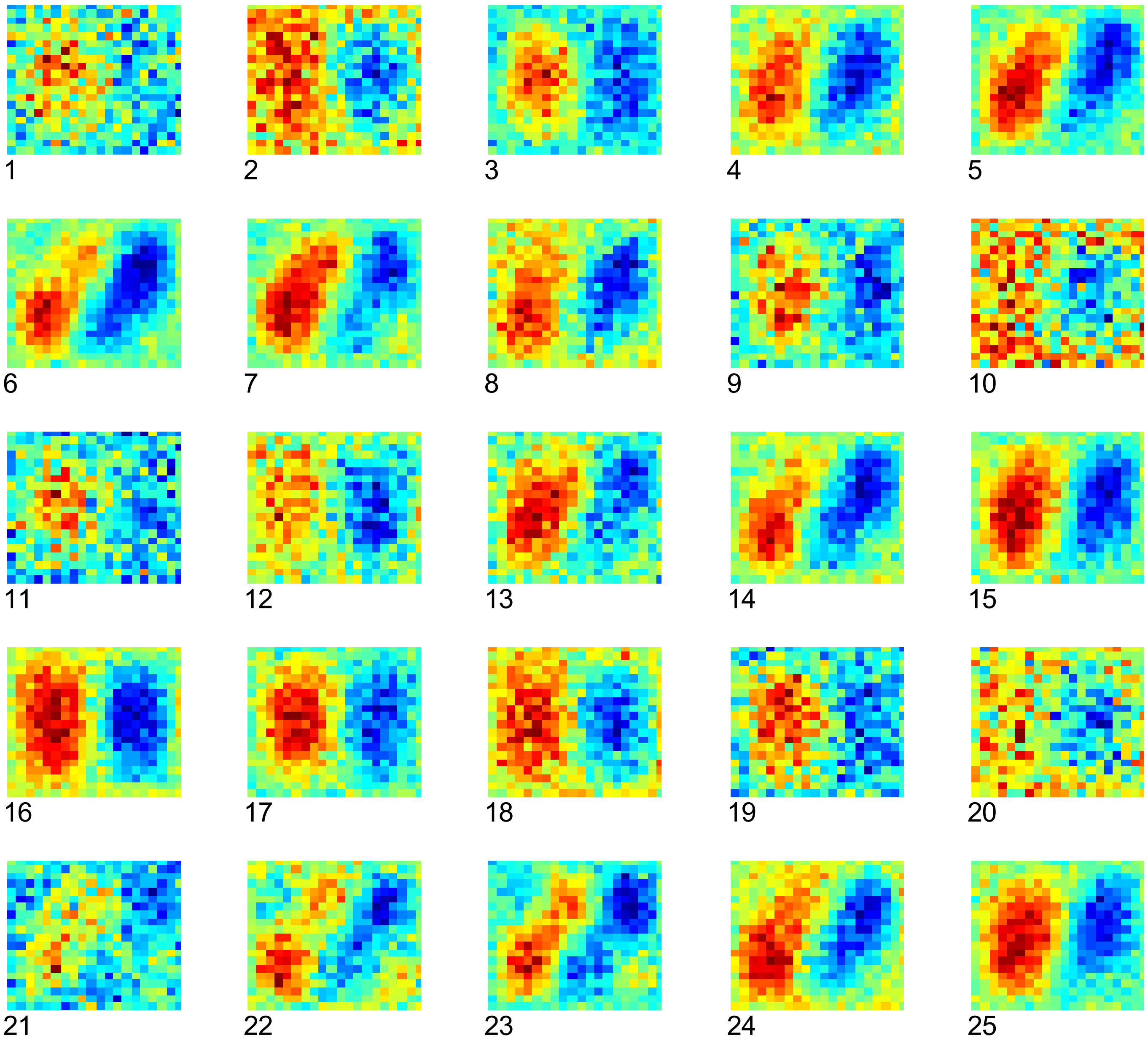
the data


y
x
t=1
spatial pattern at a single time
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

the data



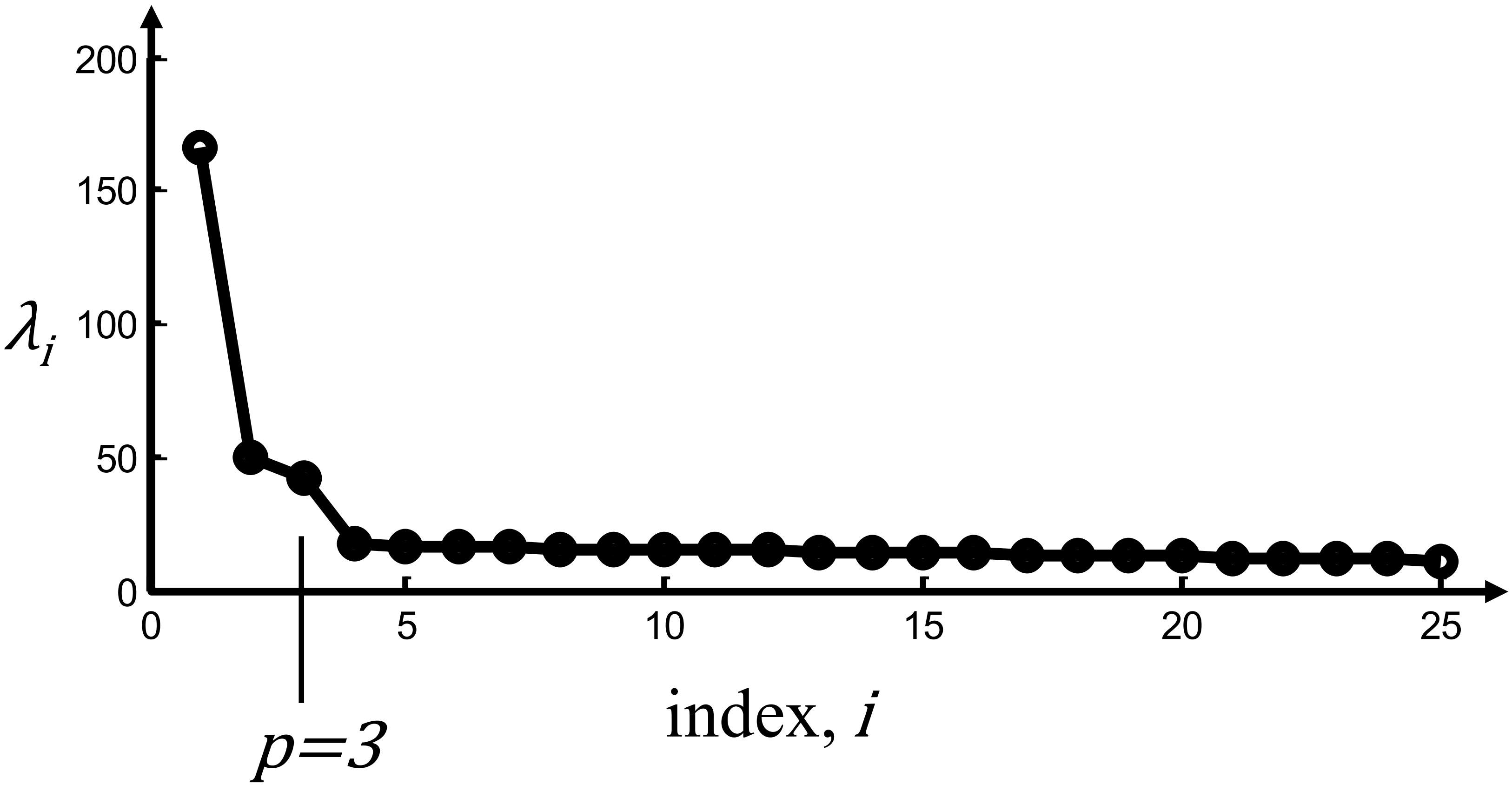
the data



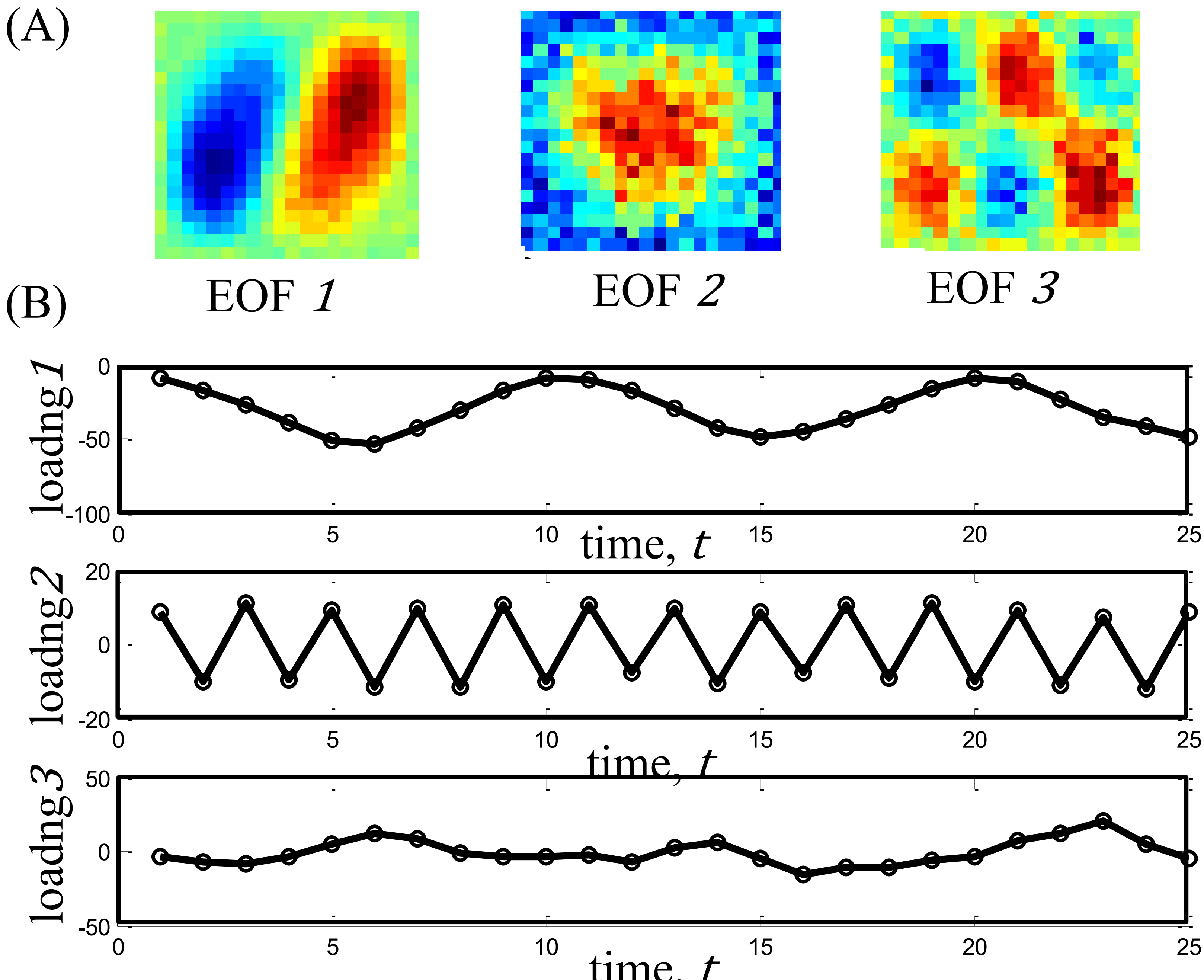
need to unfold each 2D image into vector

$$\begin{pmatrix} 4 & 6 & 1 \\ 9 & 1 & 2 \\ 1 & 3 & 6 \end{pmatrix} \longrightarrow \begin{bmatrix} 4 \\ 6 \\ 1 \\ 9 \\ 1 \\ 2 \\ 1 \\ 3 \\ 6 \end{bmatrix}$$

$$\boldsymbol{s}$$



200
150
100
50
0
λ_i
0
5
10
15
20
25
index, i
p=3



(A)
EOF 1
EOF 2
EOF 3
(B)
loadng1
time, t
loadng2
time, t
loadng3
time, t

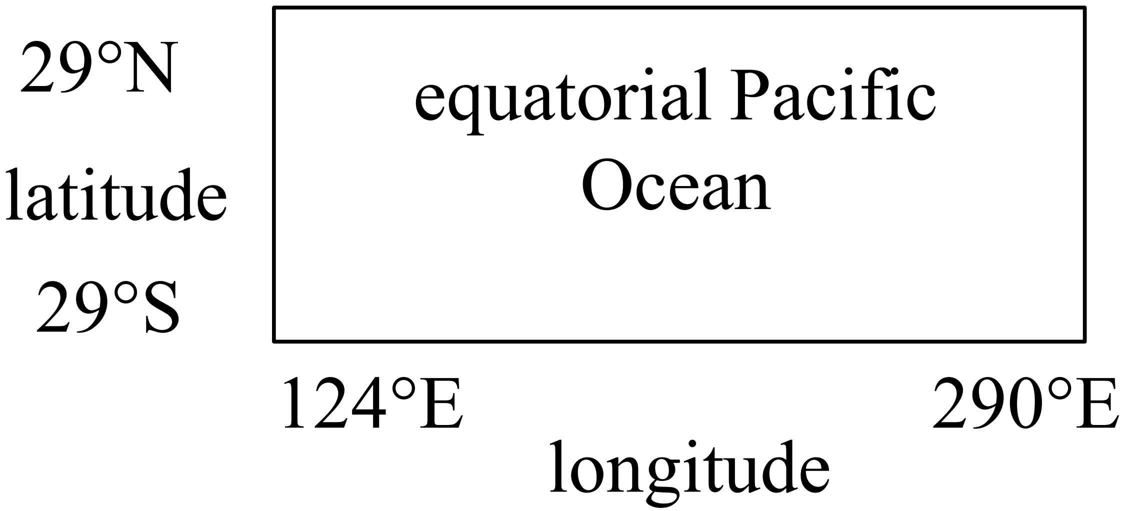
example 3

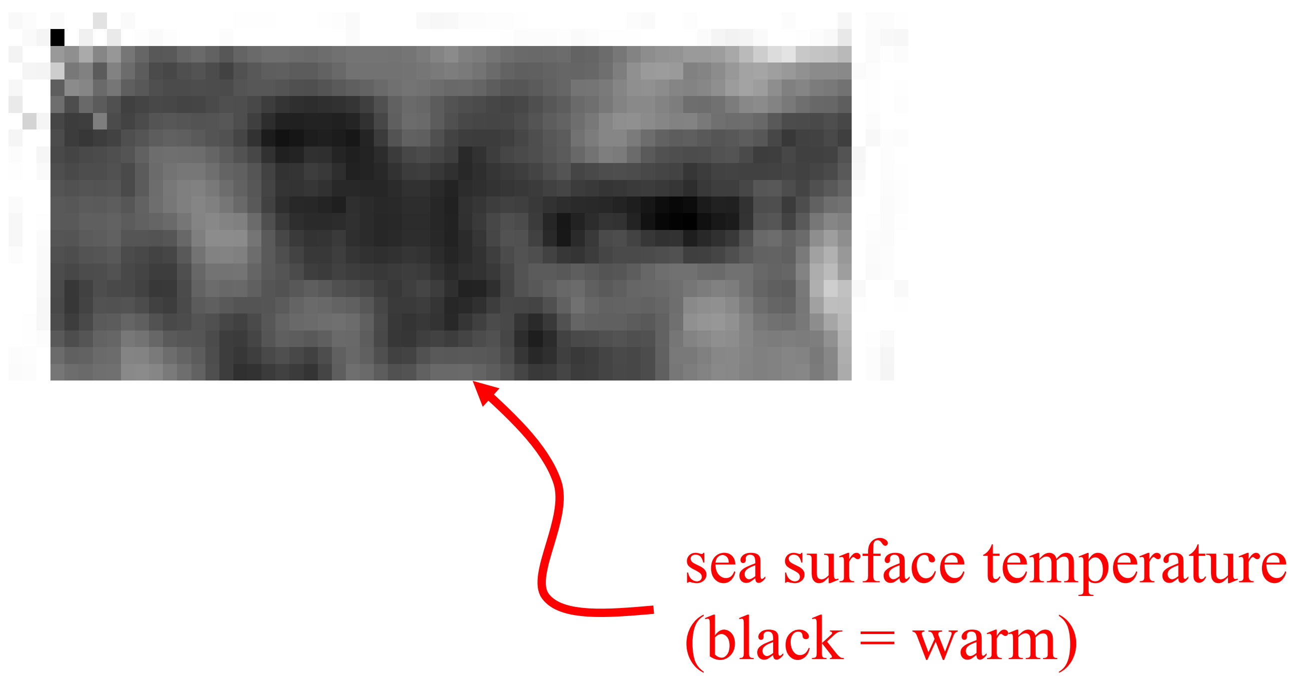
spatial-temporal patterns

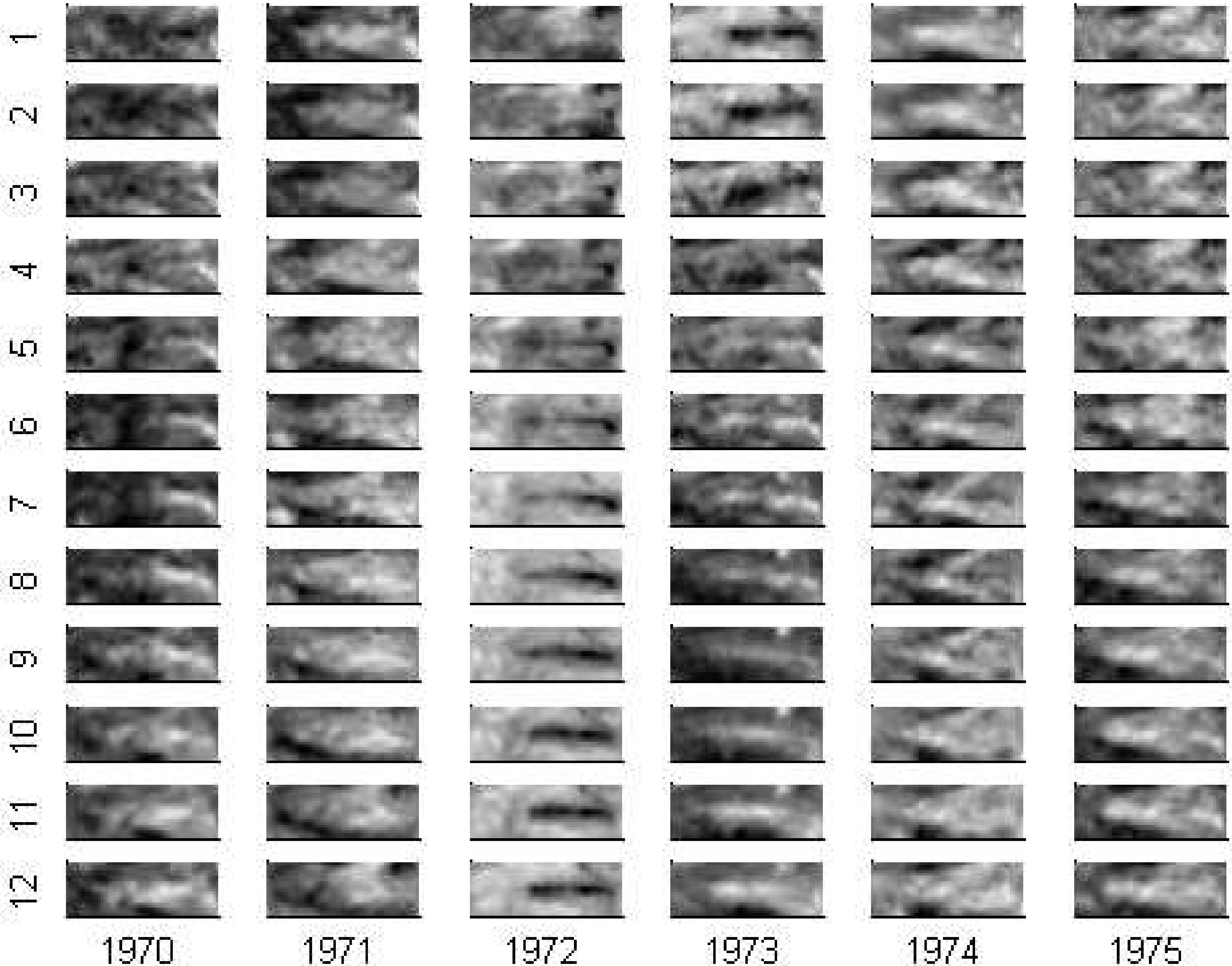
(actual data)

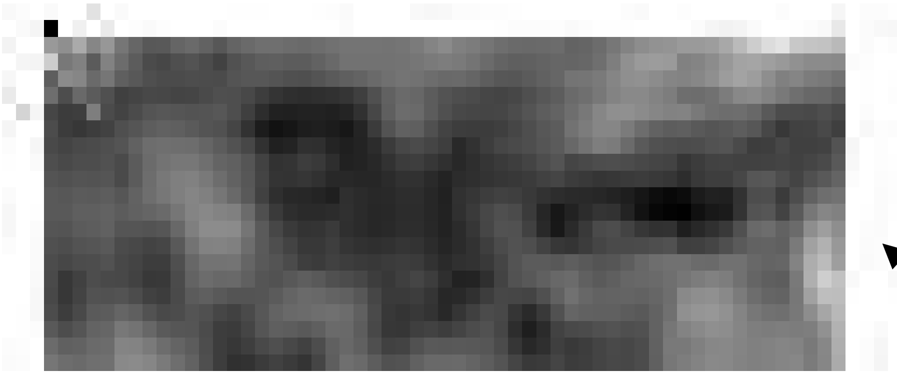
sea surface temperature in the Pacific Ocean

CAC Sea Surface Temperature






1
2
3
4
5
6
7
8
9
10
11
12
1970
1971
1972
1973
1974
1975



the image is 30 by 84 pixels in size, or 2520 pixels total

to use `svd()`, the image must be unwrapped into a vector of length 2520

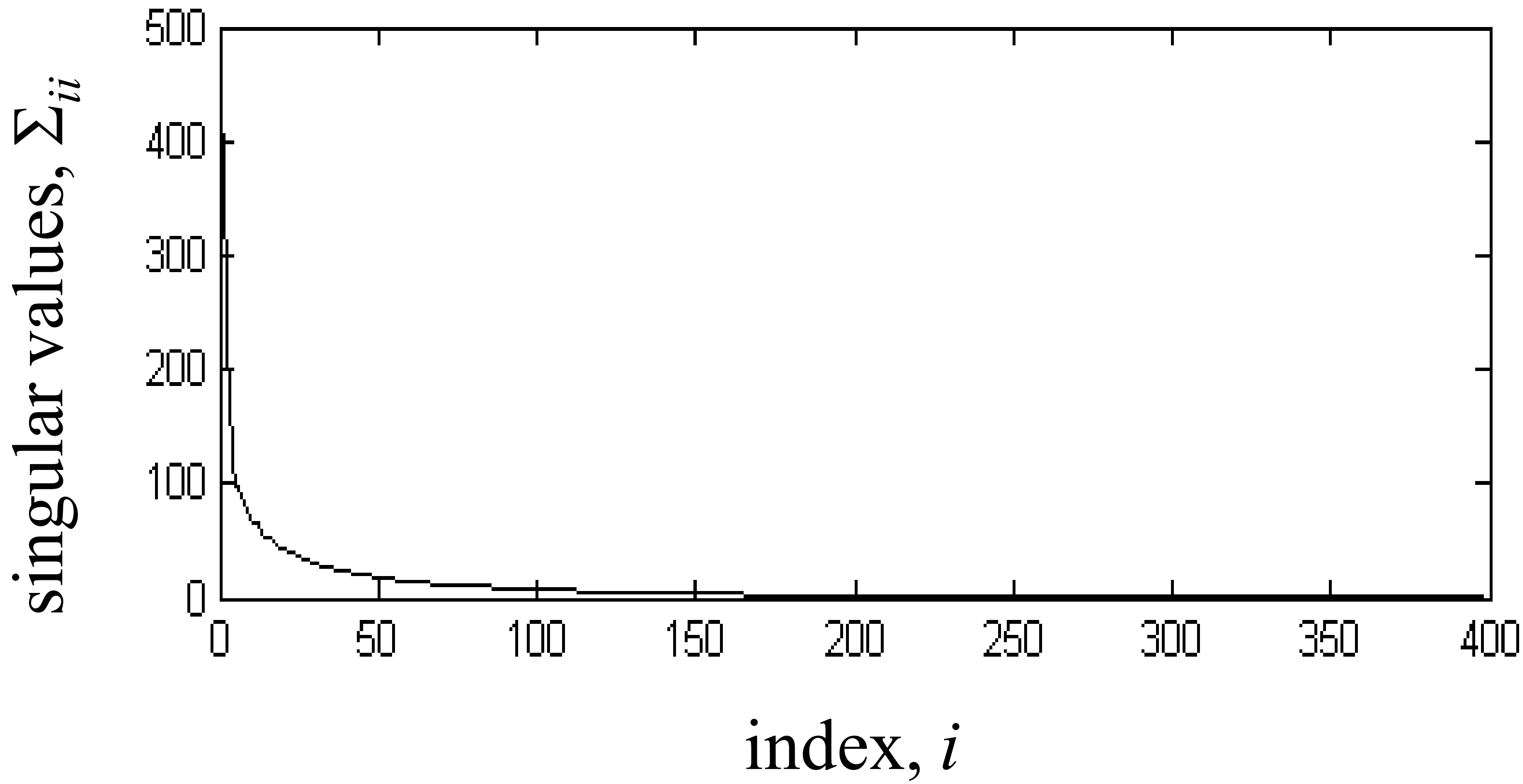
2520 positions in the equatorial Pacific ocean

399 times

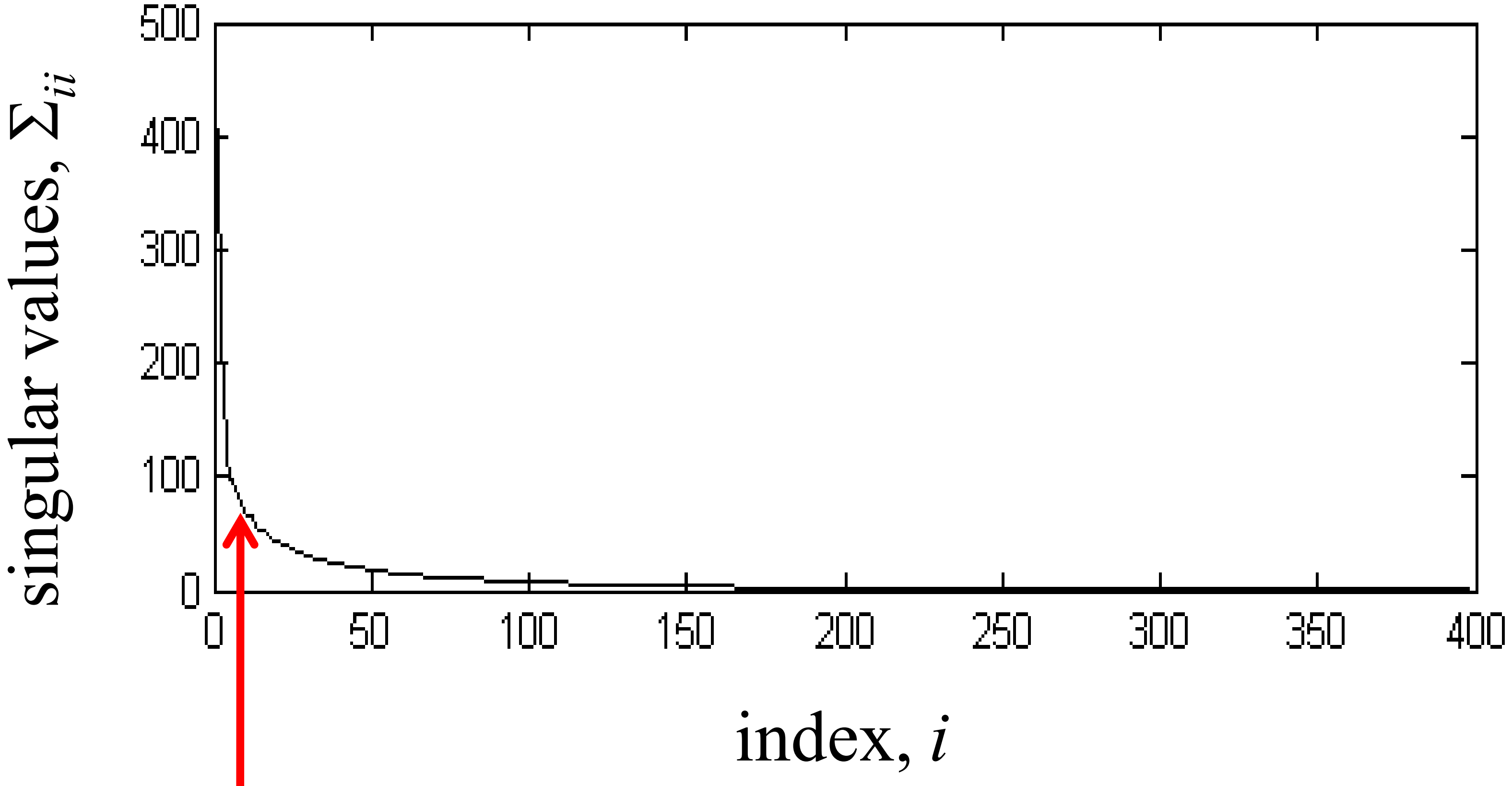
$$\mathbf{s} = \begin{bmatrix} \text{element 1 in sample 1} & \text{element 2 in sample 1} & \cdots & \text{element } M \text{ in sample 1} \\ \cdots & \cdots & \cdots & \cdots \\ \text{element 1 in sample } N & \text{element 2 in sample } N & \cdots & \text{element } M \text{ in sample } N \end{bmatrix}$$

“element” means temperature

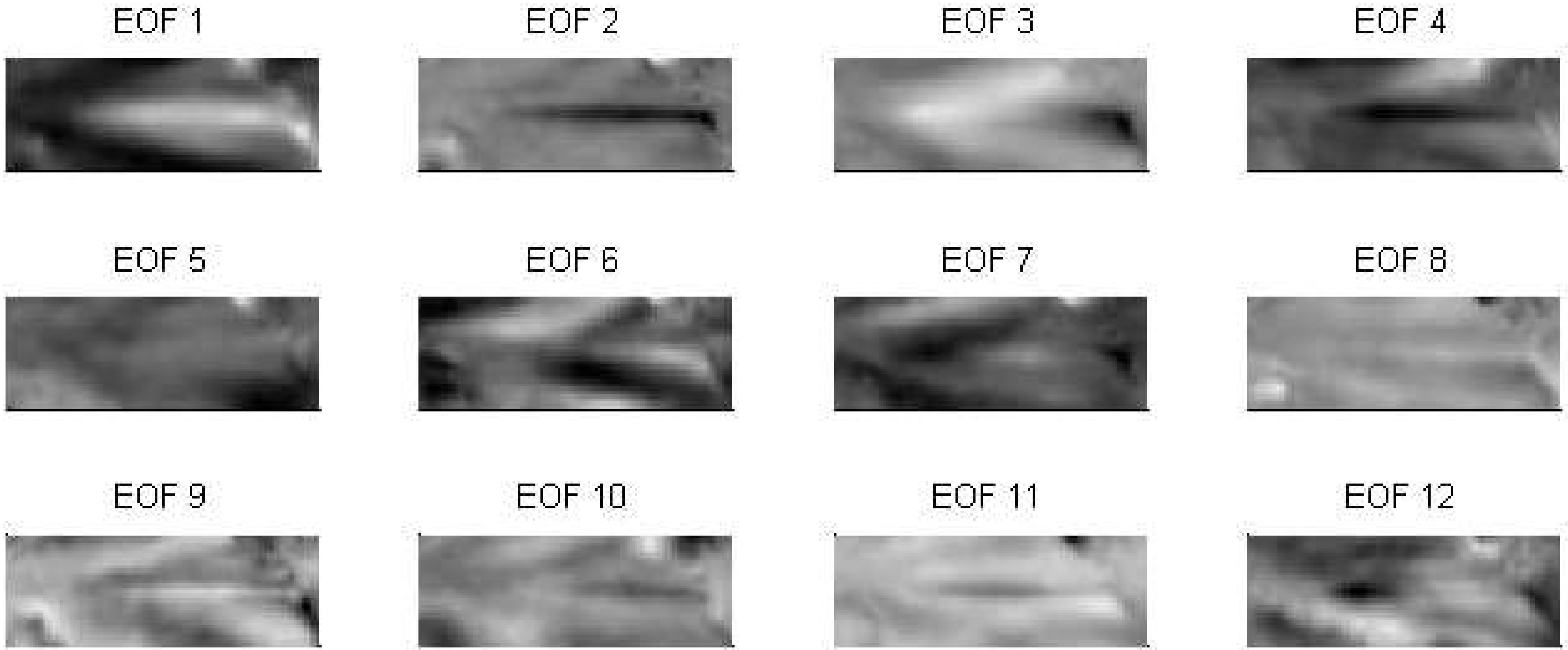
singular values

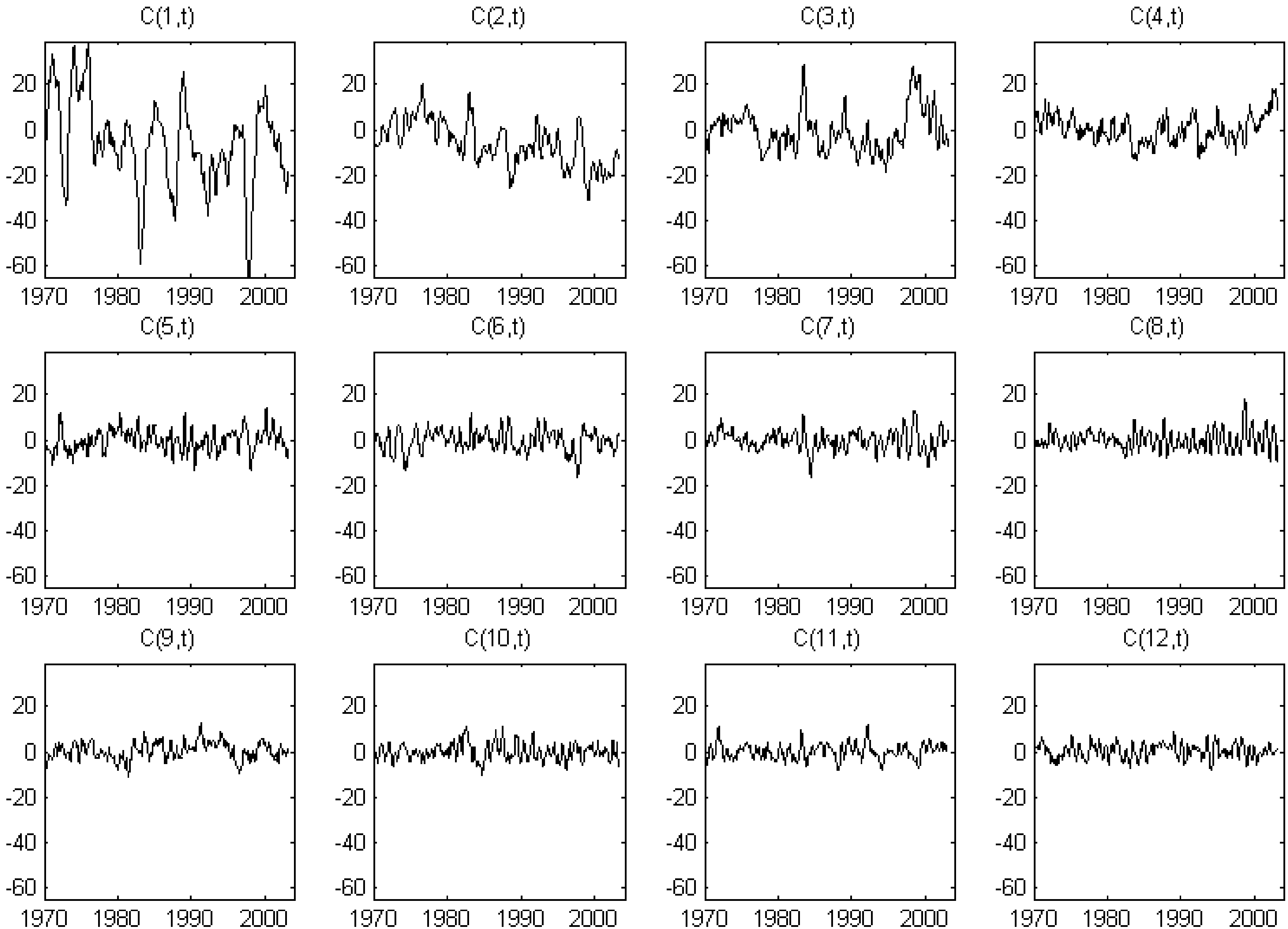


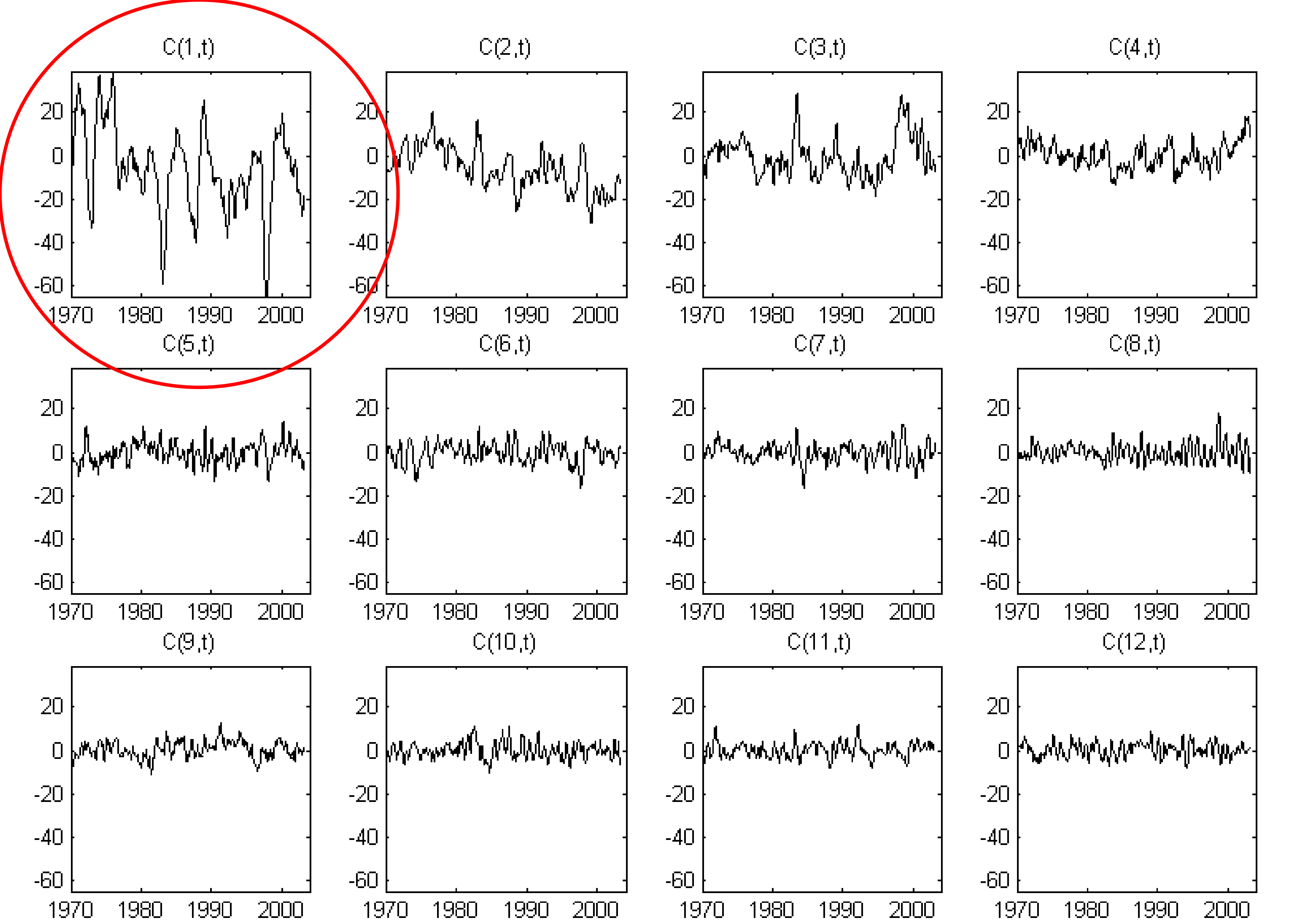
singular values



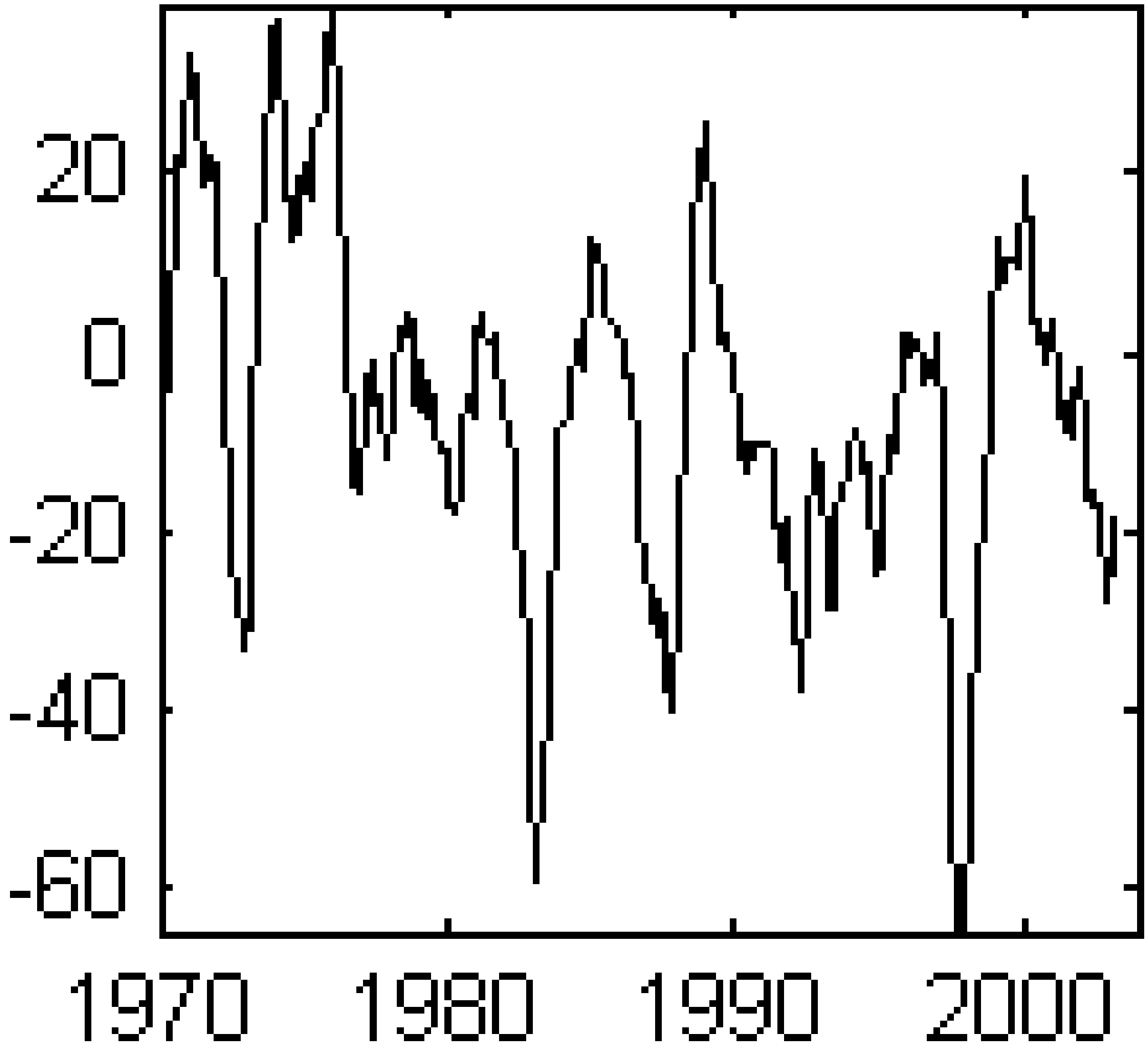
no clear cutoff for P, but the first 12 singular values are considerably larger than the rest


EOF 1
EOF 2
EOF 3
EOF 4
EOF 5
EOF 6
EOF 7
EOF 8
EOF 9
EOF 10
EOF 11
EOF 12

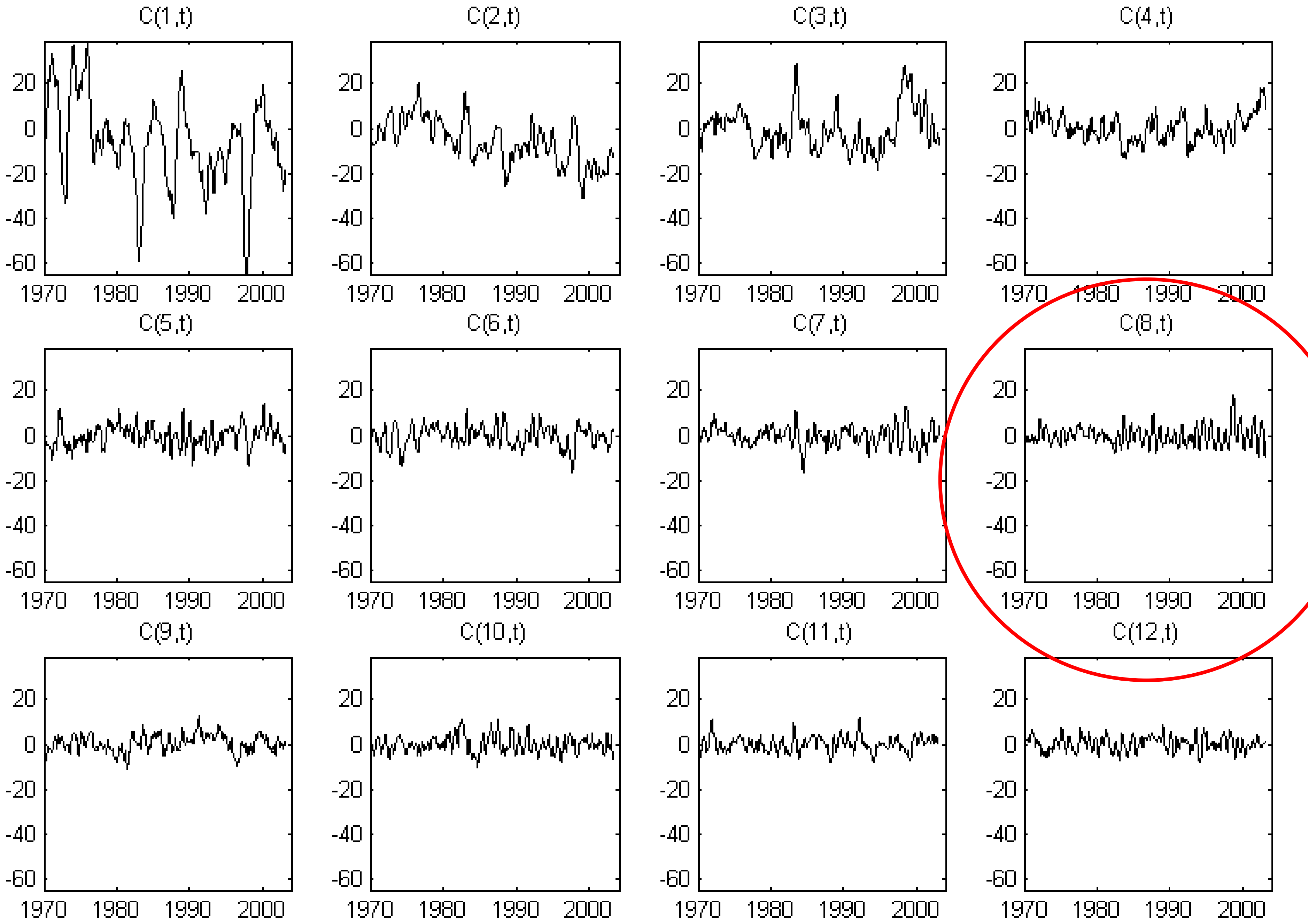

C(1,t)
C(2,t)
C(3,t)
C(4,t)
C(5,t)
C(6,t)
C(7,t)
C(8,t)
C(9,t)
C(10,t)
C(11,t)
C(12,t)
20
0
-20
-40
-60
1970
1980
1990
2000

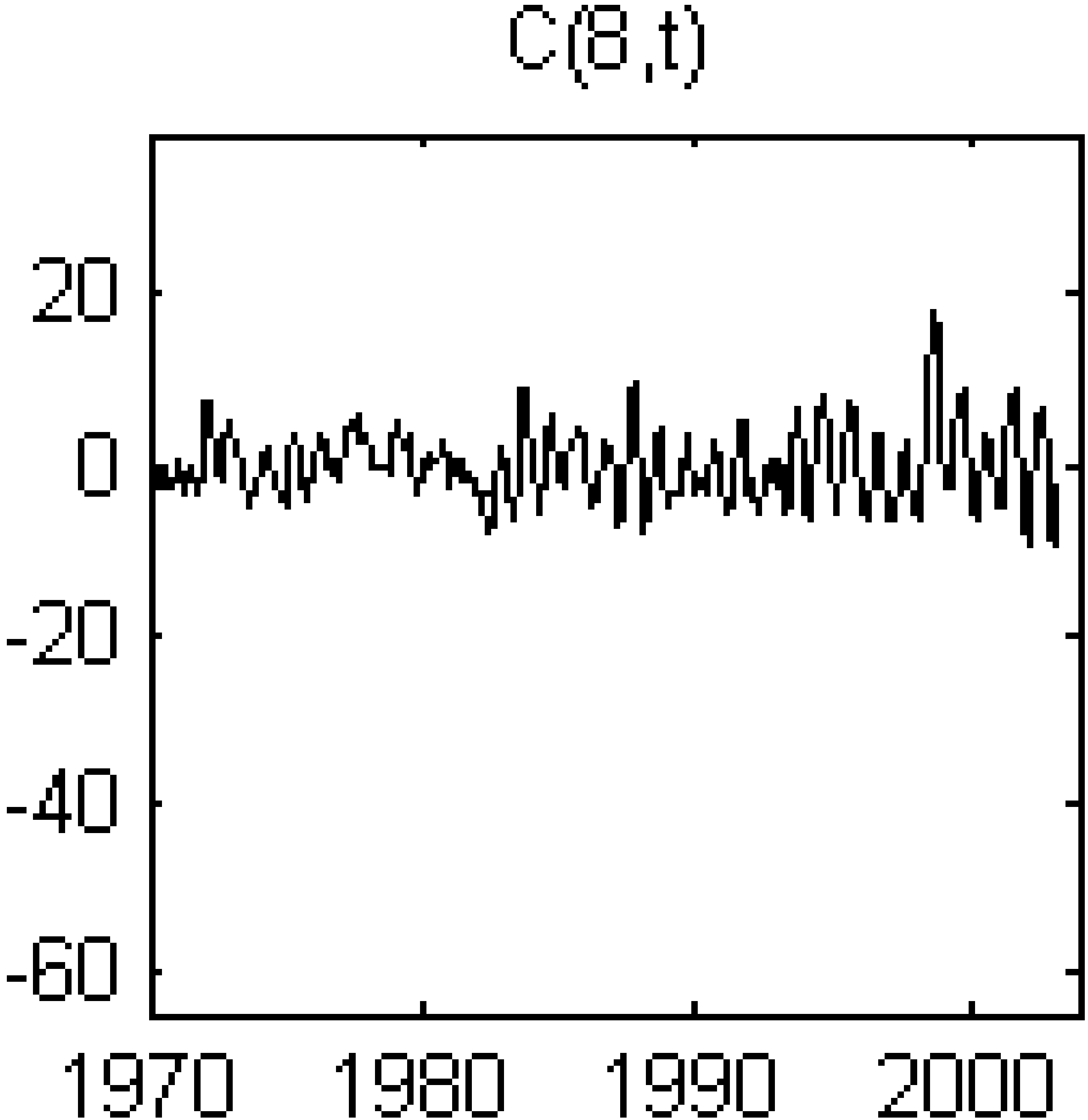

C(1,t)
C(2,t)
C(3,t)
C(4,t)
C(5,t)
C(6,t)
C(7,t)
C(8,t)
C(9,t)
C(10,t)
C(11,t)
C(12,t)
20
0
-20
-40
-60
1970
1980
1990
2000

C(1,t)



C(5,t)


C(1,t)
C(2,t)
C(3,t)
C(4,t)
C(5,t)
C(6,t)
C(7,t)
C(8,t)
C(9,t)
C(10,t)
C(11,t)
C(12,t)
20
0
-20
-40
-60
1970
1980
1990
2000

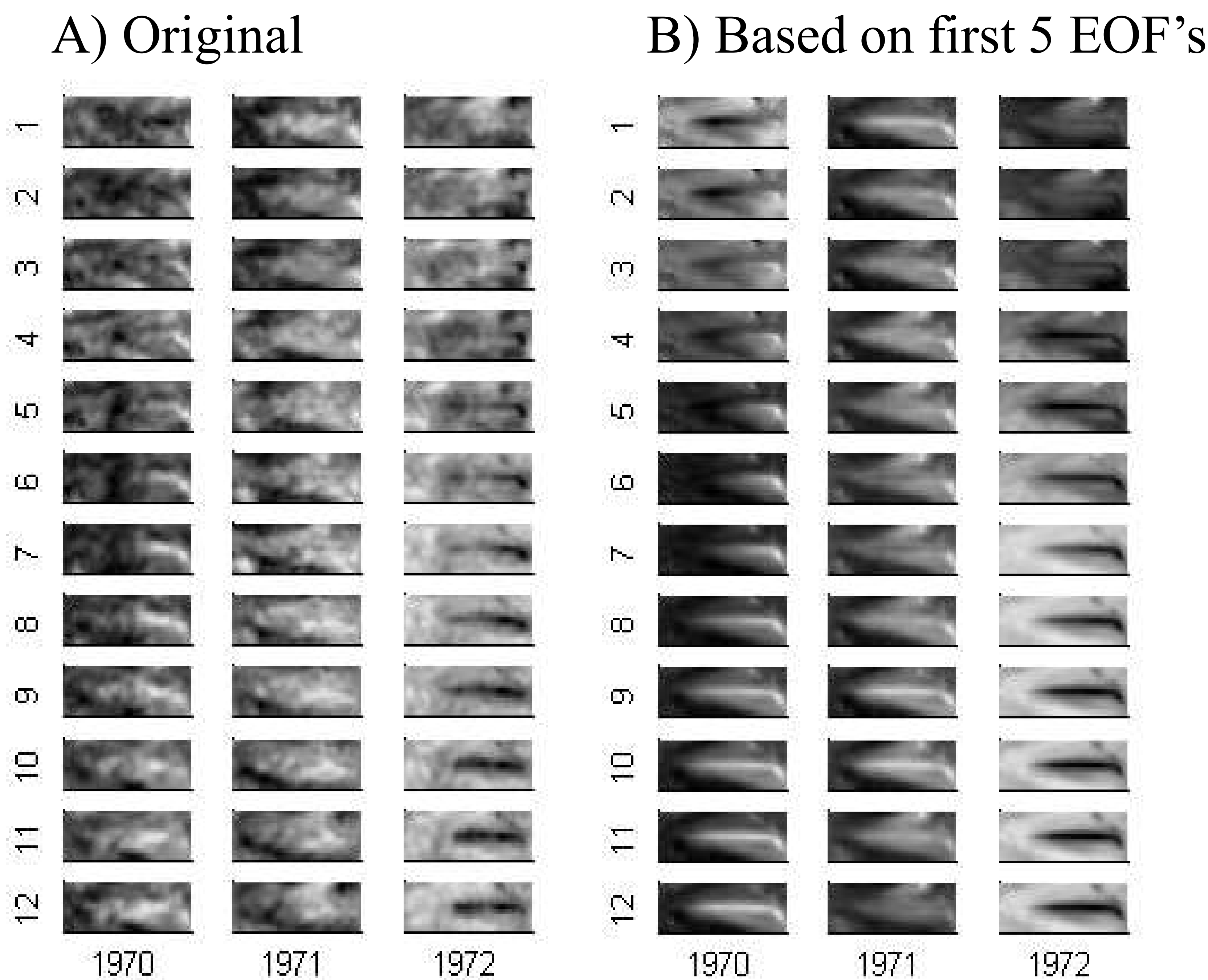

C(8,t)
20
0
-20
-40
-60
1970
1980
1990
2000

using SVD to approximate data

$\mathbf{S}=\mathbf{C}_{\mathrm{M}}\mathbf{F}_{\mathrm{M}}$ With M EOF’s, the data is fit exactly

$\mathbf{S}=\mathbf{C}_{\mathrm{P}}\mathbf{F}_{\mathrm{P}}$ With P chosen to exclude only zero singular values, the data is fit exactly

$\mathbf{S}\approx\mathbf{C}_{\mathrm{P'}}\mathbf{F}_{\mathrm{P'}}$ With P’<P, small non-zero singular values are excluded too, and the data is fit only approximately


A) Original
B) Based on first 5 EOF's
1
2
3
4
5
6
7
8
9
10
11
12
1970
1971
1972