3-WAY PCA
AS A POWERFUL
DATA DISPLAY METHOD

R. Leardi

Department of Pharmaceutical and Food Chemistry and Technology, Via Brigata Salerno (Ponte), 16147 Genoa, Italy.

DATA SET VENICE

(M.L. Tercier-Waeber[1], B. Gianni[2], G. Ferrari[3])
[1]*Department of Inorganic and Analytical Chemistry, University of Geneva, Switzerland.*
[2]*Venice Water Authority-Consorzio Venezia Nuova, Venice, Italy.*
[3]*Magistrato alle Acque-Antipollution Section, Venice, Italy.*

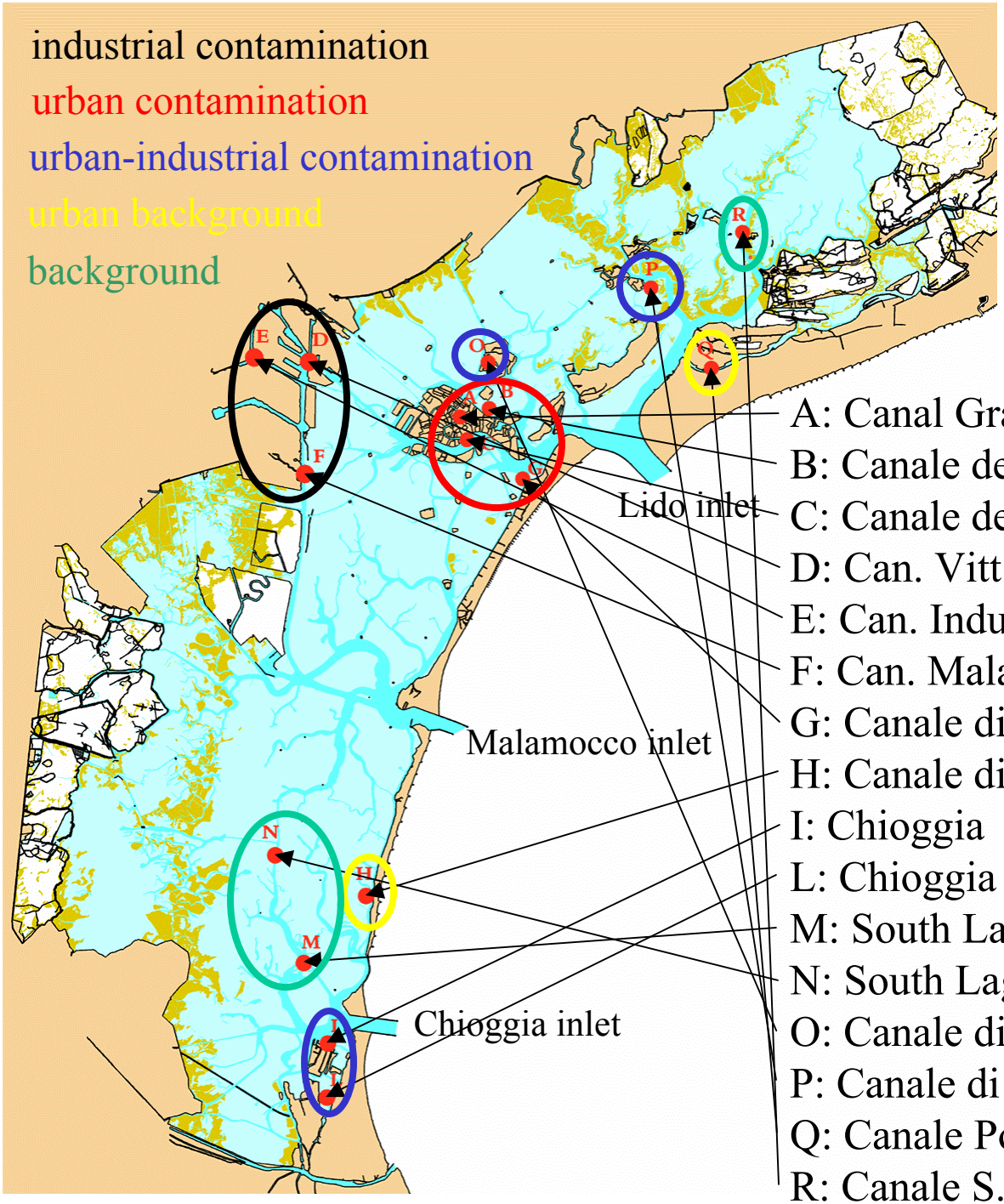
13 variables:

1) pH
2) NO_2^-
3) NO_3^-
4) NH_4^+
5) PO_4^{3-}
6) Dissolved organic P
7) Redox E
8) Cd (total)
9) Pb (total)
10) Cu (total)
11) Cd (dynamic)
12) Pb (dynamic)
13) Cu (dynamic)

12 samplings:

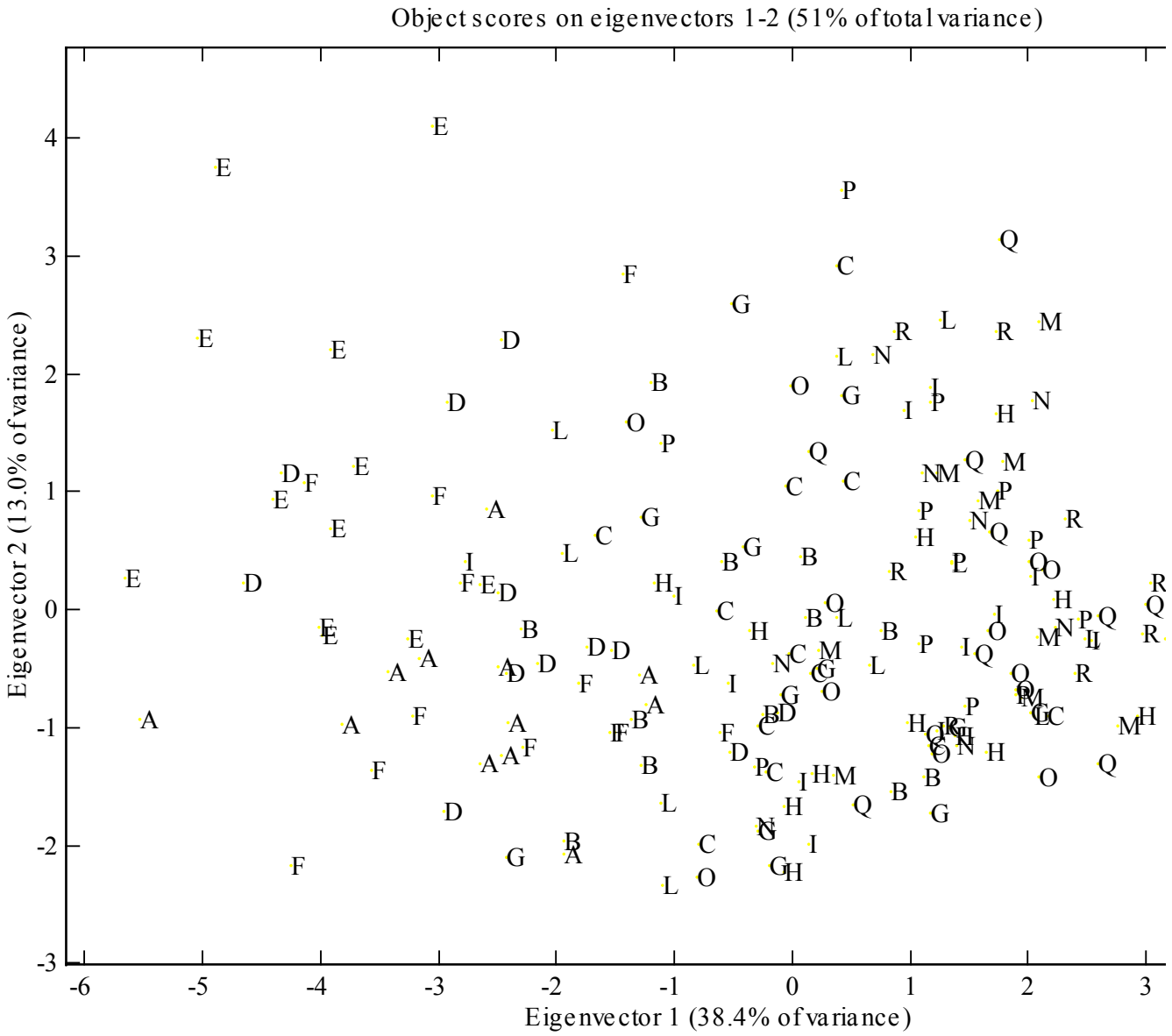
Samplings have been performed once per month, from January to December 2001, at the quadrature of the tide, in the 50 cm superficial water layer of each station.

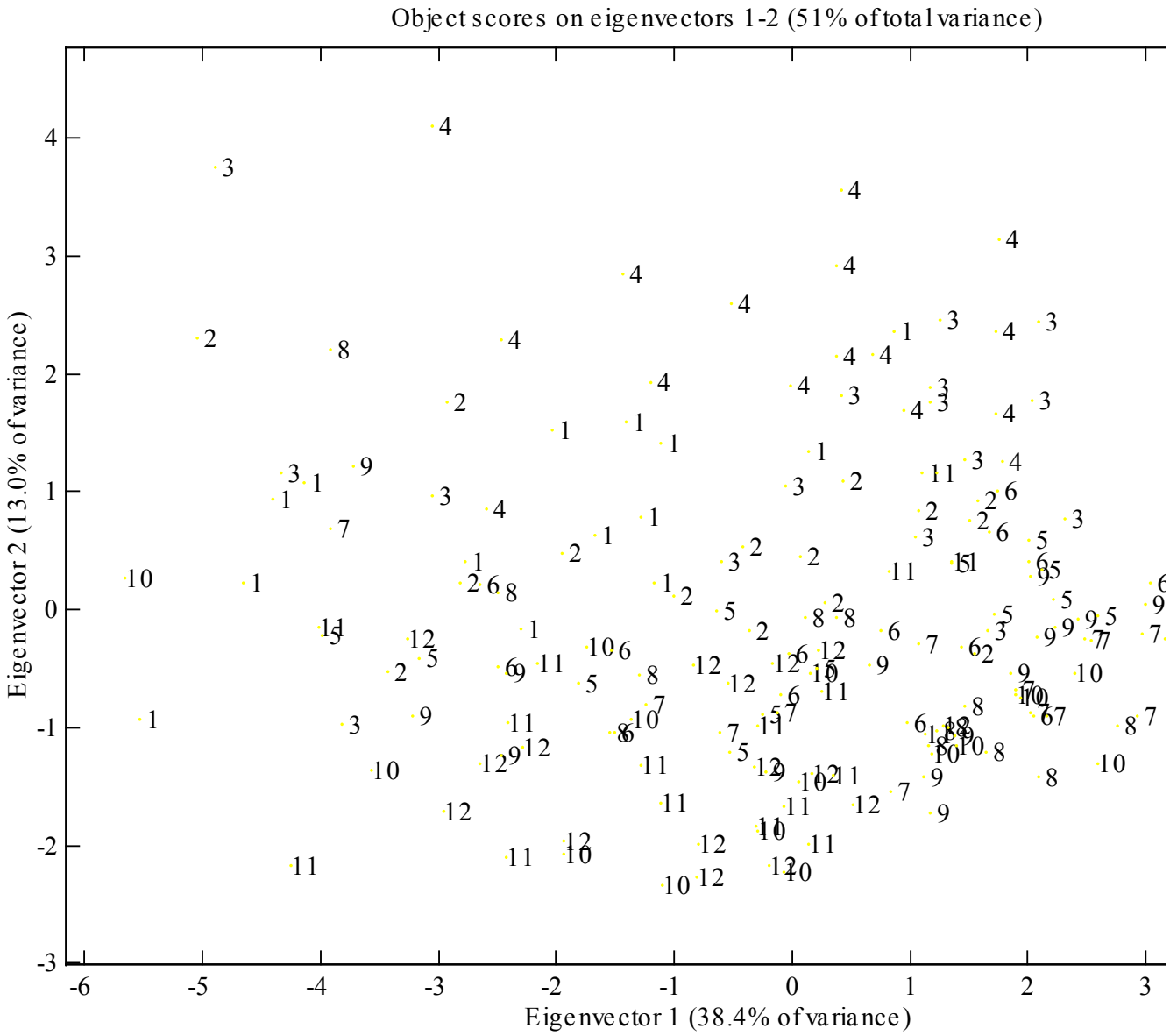
Dynamic fraction: free metal ion and small labile complexes with size of few nanometers. It is the most easily bioavailable and therefore the most toxic fraction.



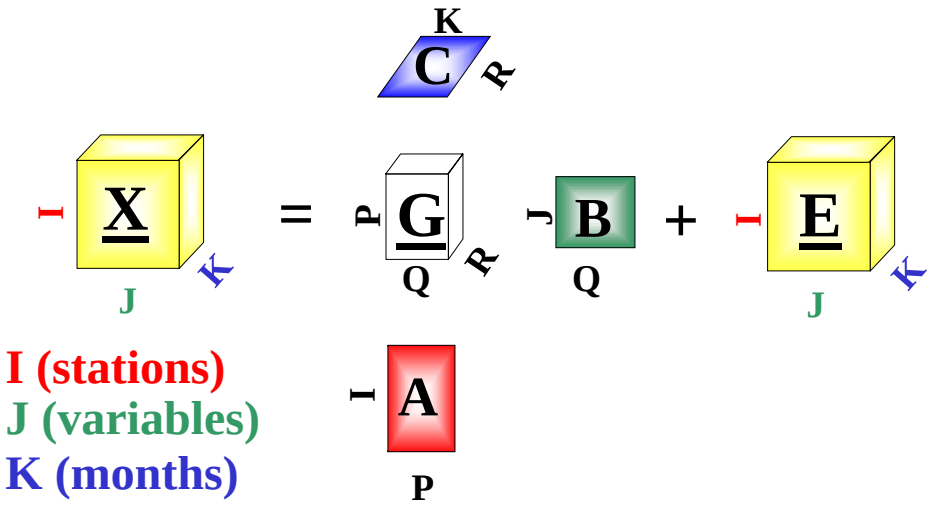
16 sampling stations:

A: Canal Grande
B: Canale della Giudecca
C: Canale delle Fondamenta Nuove
D: Can. Vitt. Emanuele (Porto Marghera)
E: Can. Industriale Ovest (P. Marghera)
F: Can. Malamocco-Marghera (P. Marghera)
G: Canale di S. Maria Elisabetta (Lido)
H: Canale di Pellestrina
I: Chioggia
L: Chioggia
M: South Lagoon (reference)
N: South Lagoon (reference)
O: Canale di Sacca Serenella (Murano)
P: Canale di Burano
Q: Canale Pordelio (Ca' Savio)
R: Canale S. Felice (reference)


Object scores on eigenvectors 1-2 (51% of total variance)
Eigenvector 2 (13.0% of variance)
Eigenvector 1 (38.4% of variance)
4
3
2
1
0
-1
-2
-3
-6
-5
-4
-3
-2
-1
0
1
2
3


Object scores on eigenvectors 1-2 (51% of total variance)
Eigenvector 2 (13.0% of variance)
Eigenvector 1 (38.4% of variance)
4
3
2
1
0
-1
-2
-3
-6
-5
-4
-3
-2
-1
0
1
2
3

Three-way principal component analysis: Tucker3 model



Study of pollution profile

Mode 1: 16 stations

Mode 2: 13 variables

Mode 3: 12 months

$$x_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip} b_{jp} c_{kr} g_{pqr} + e_{ijk}$$

a_{ip}, b_{jq}, c_{kr} = elements of the loading matrices A, B and C of order IxP, JxQ, KxR resp.

g_{pqr} = element (p,q,r) of the PxQxR core array G;

the core array describes the relationship among the three loading matrices

e_{ijk} = error term for the element x_{ijk}; element of the IxJxK array E

THREE-WAY PCA RESULTS

2 components per each mode
40.3% of the total variance explained

Core matrix:

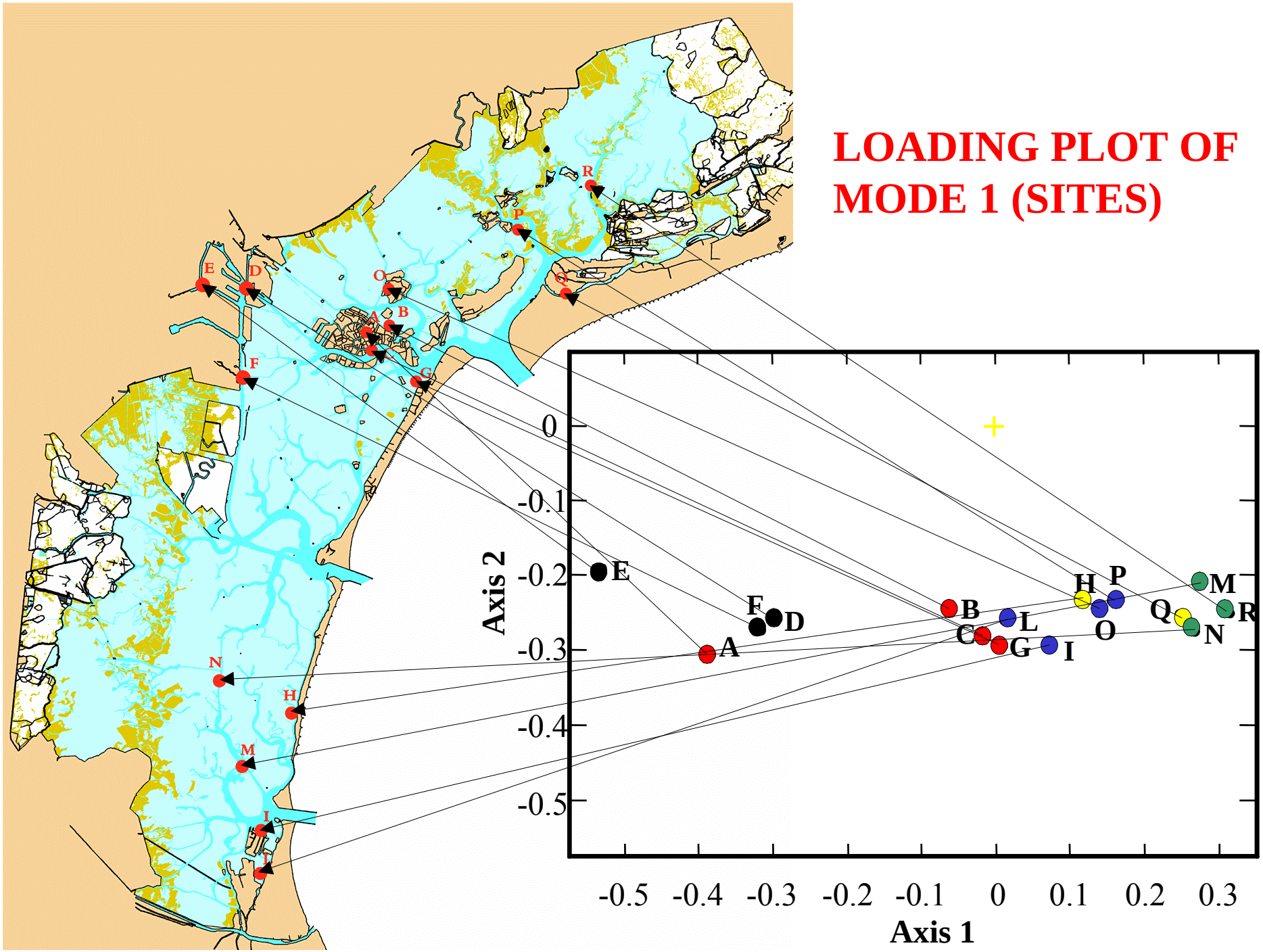
c111	c121	c112	c122
c211	c221	c212	c222
26.75	-3.89	-0.55	0.31
-0.26	0.93	7.42	-14.62

% explained variance:

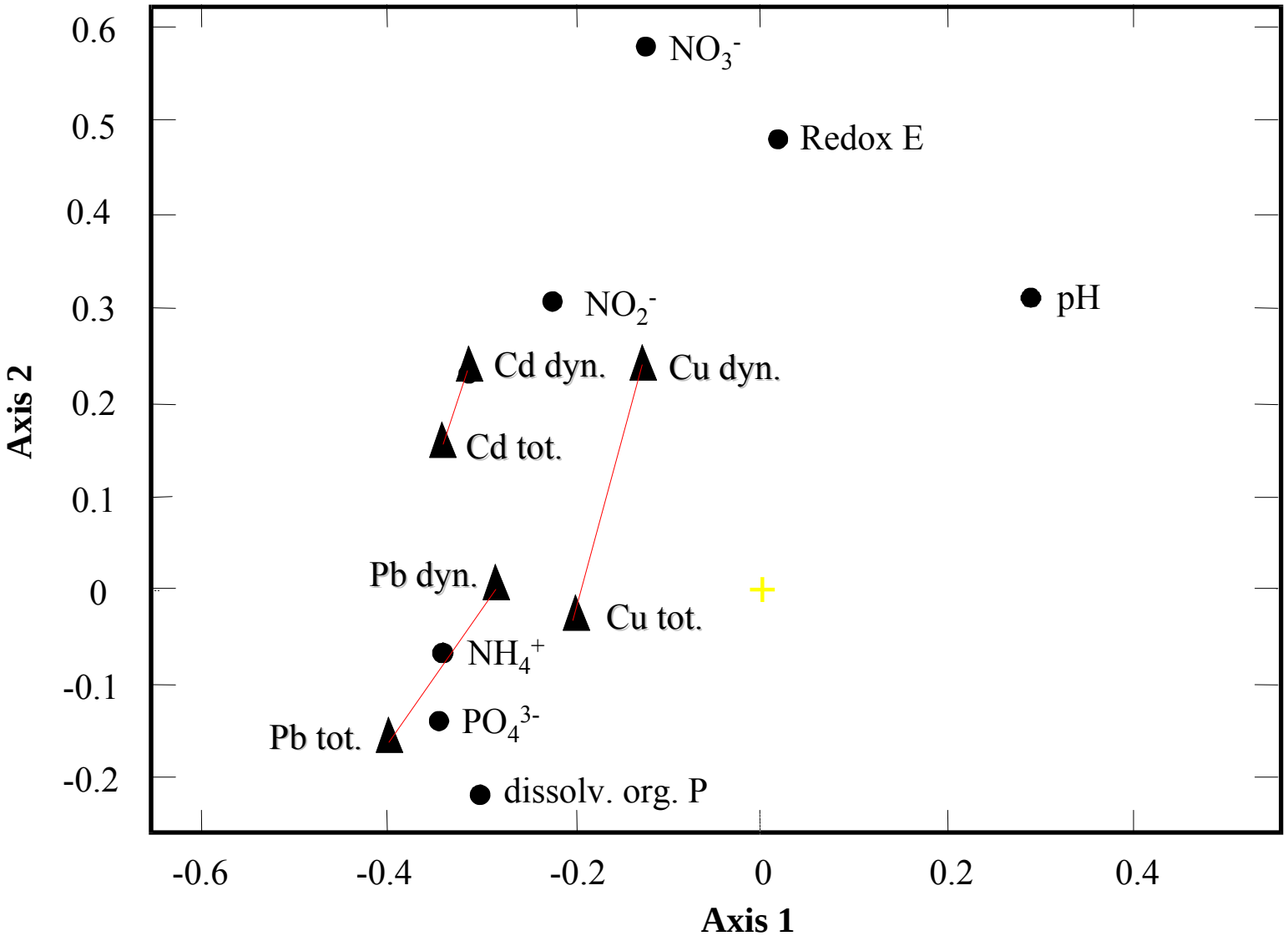
28.8	0.6	0.0	0.0
0.0	0.0	2.2	8.6

Since the core matrix is almost totally superdiagonal, the three loading plots (samples, variables and conditions) can be interpreted jointly.

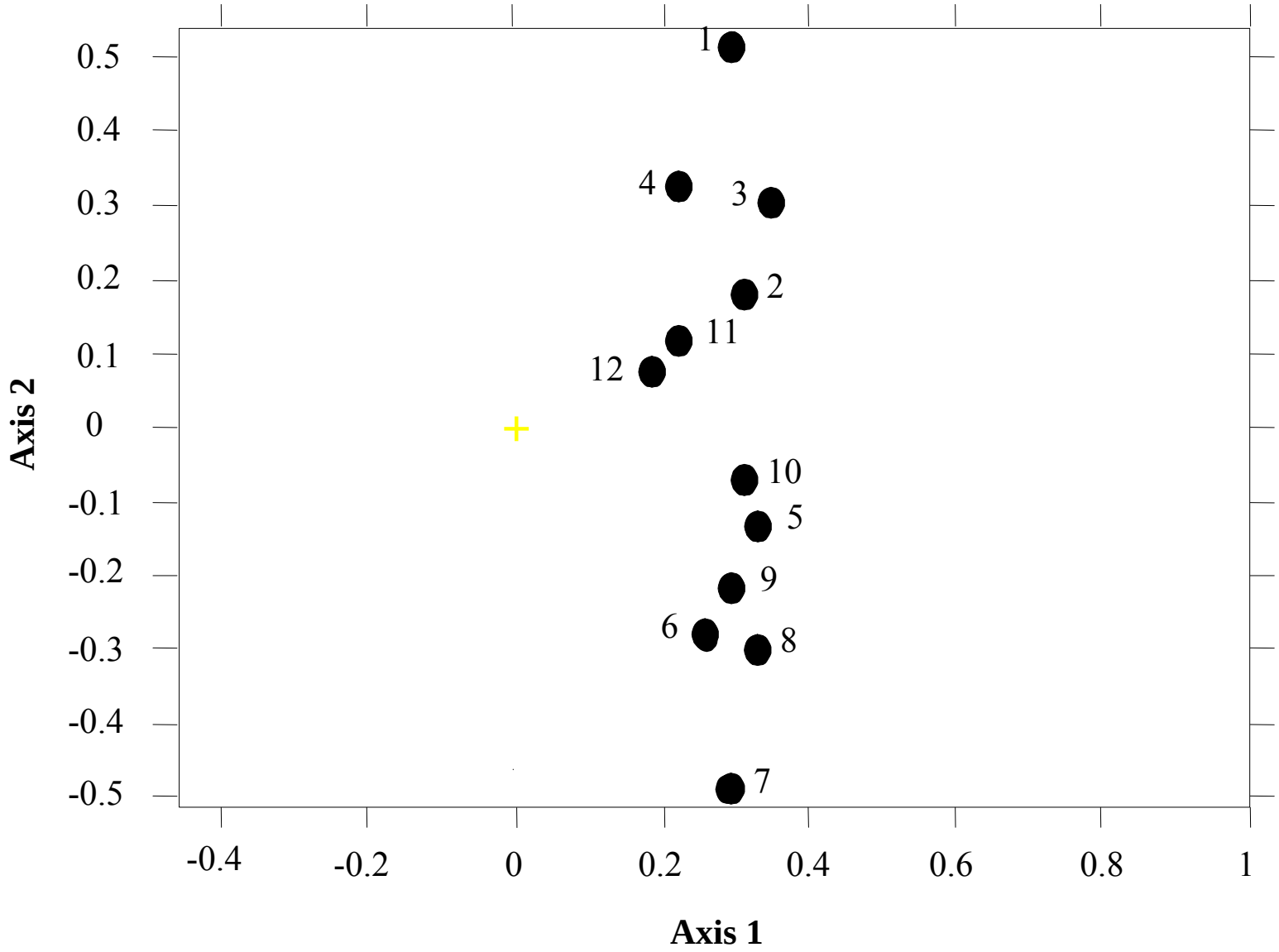
LOADING PLOT OF MODE 1 (SITES)

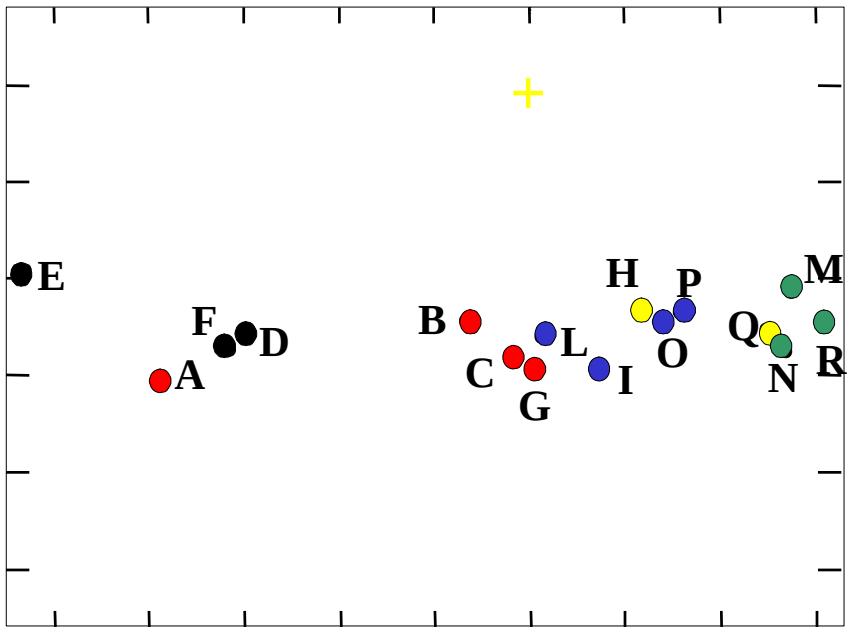


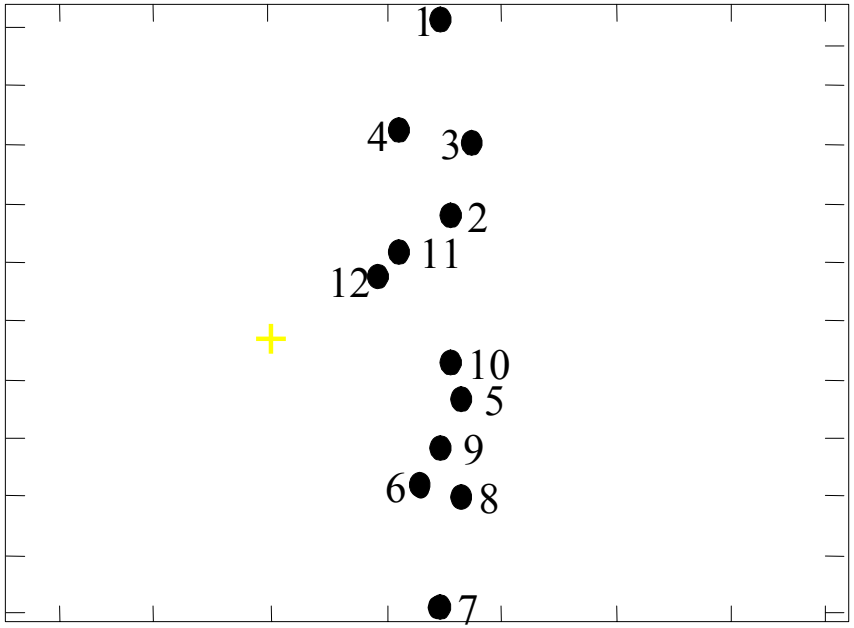
LOADING PLOT OF MODE 2 (VARIABLES)

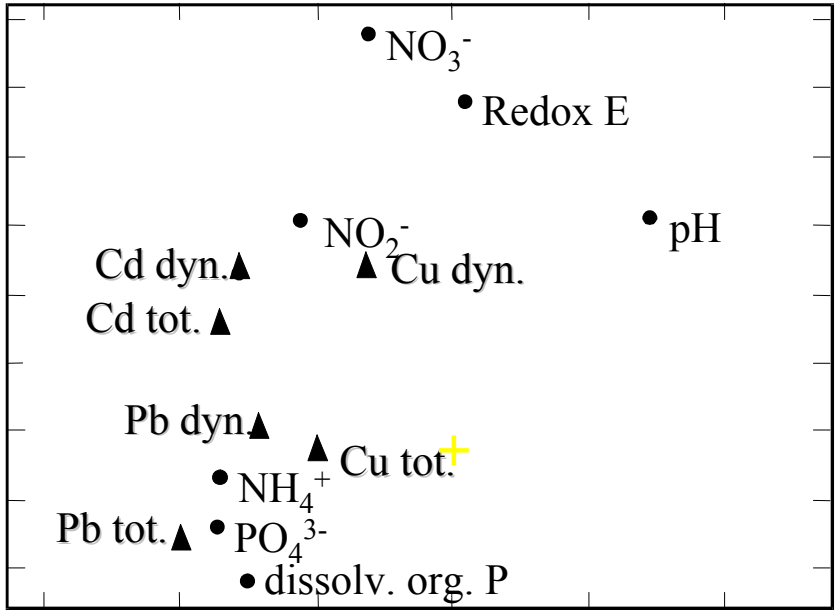


LOADING PLOT OF MODE 3 (MONTHS)




E
F
D
A
B
C
G
L
I
H
P
O
Q
N
M
R


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


NO3-
Redox E
NO2-
pH
Cd dyn.
Cu dyn.
Cd tot.
Pb dyn.
Cu tot.
NH4+
Pb tot.
PO43-
dissolv. org. P

DATA SET CARS

Objects (cars):

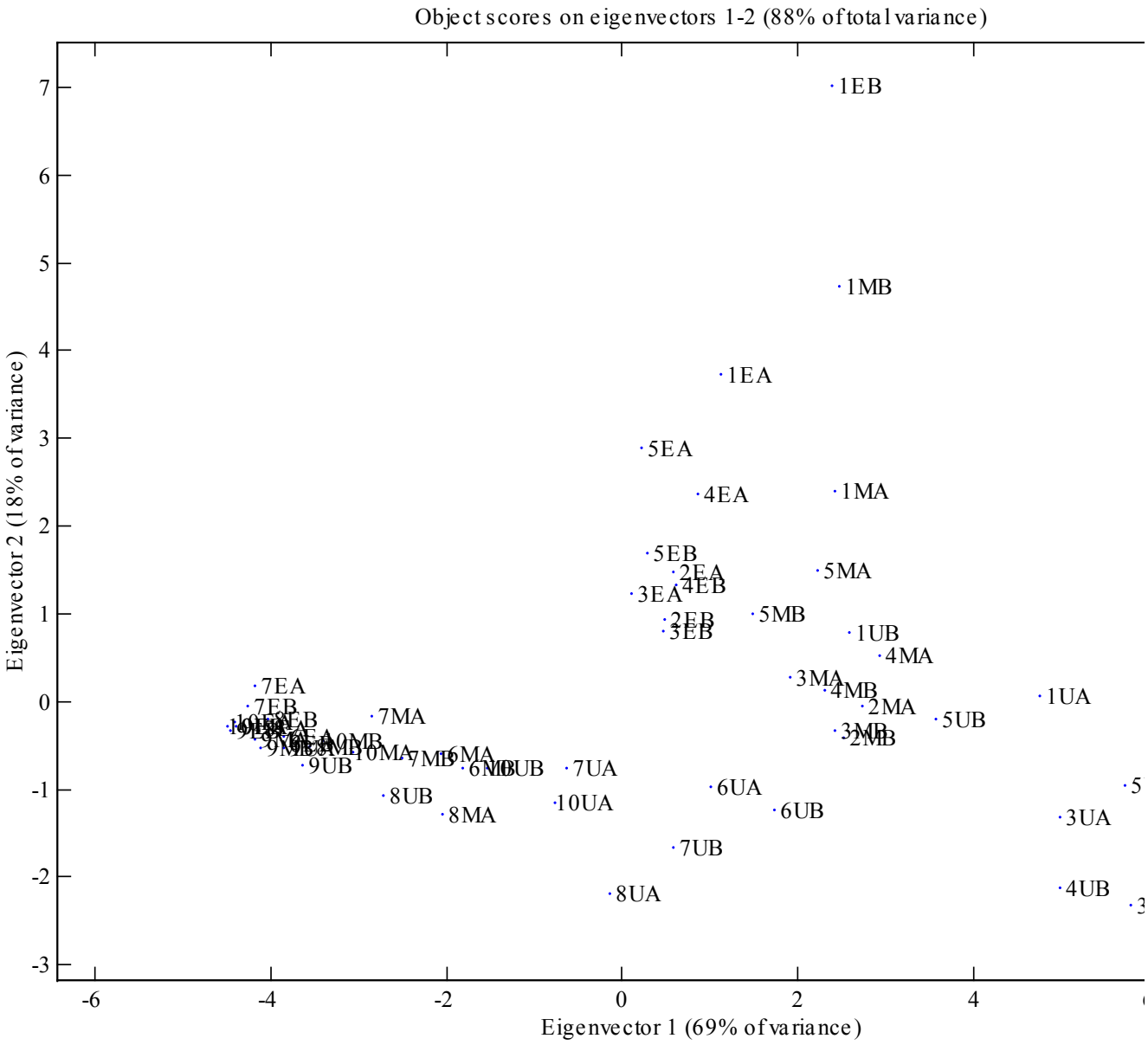
1) Fiat Tempra 1.6
2) Fiat Uno 45 Fire
3) Fiat Uno 60
4) Panda Ecobox
5) Fiat Tipo 1.4
6) VW Polo Kat
7) Alfa Romeo 33 1.7 K
8) Fiat Uno 1000 K
9) Fiat Panda 1000 K
10)Fiat Tipo 1.4 K

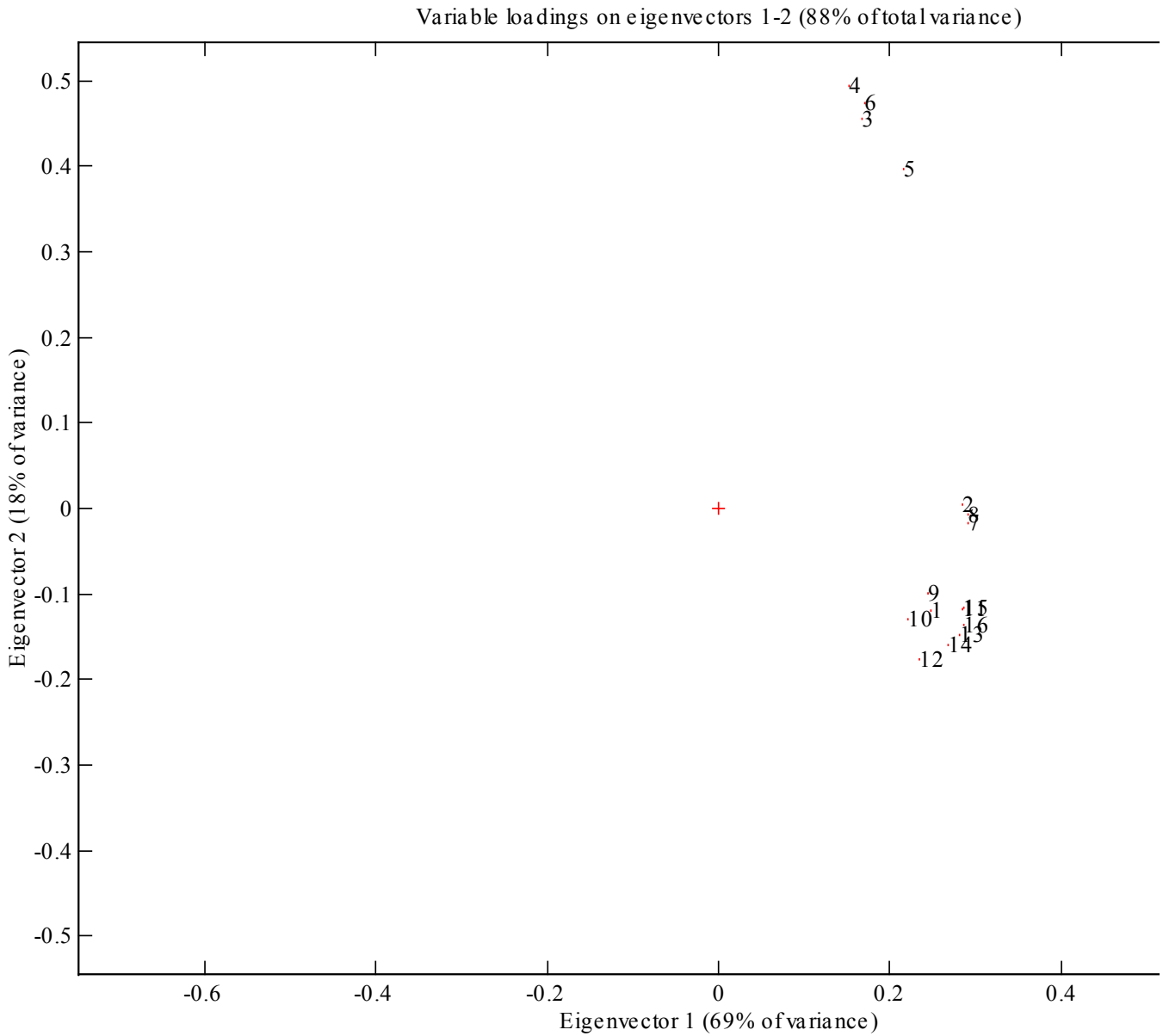
Variables

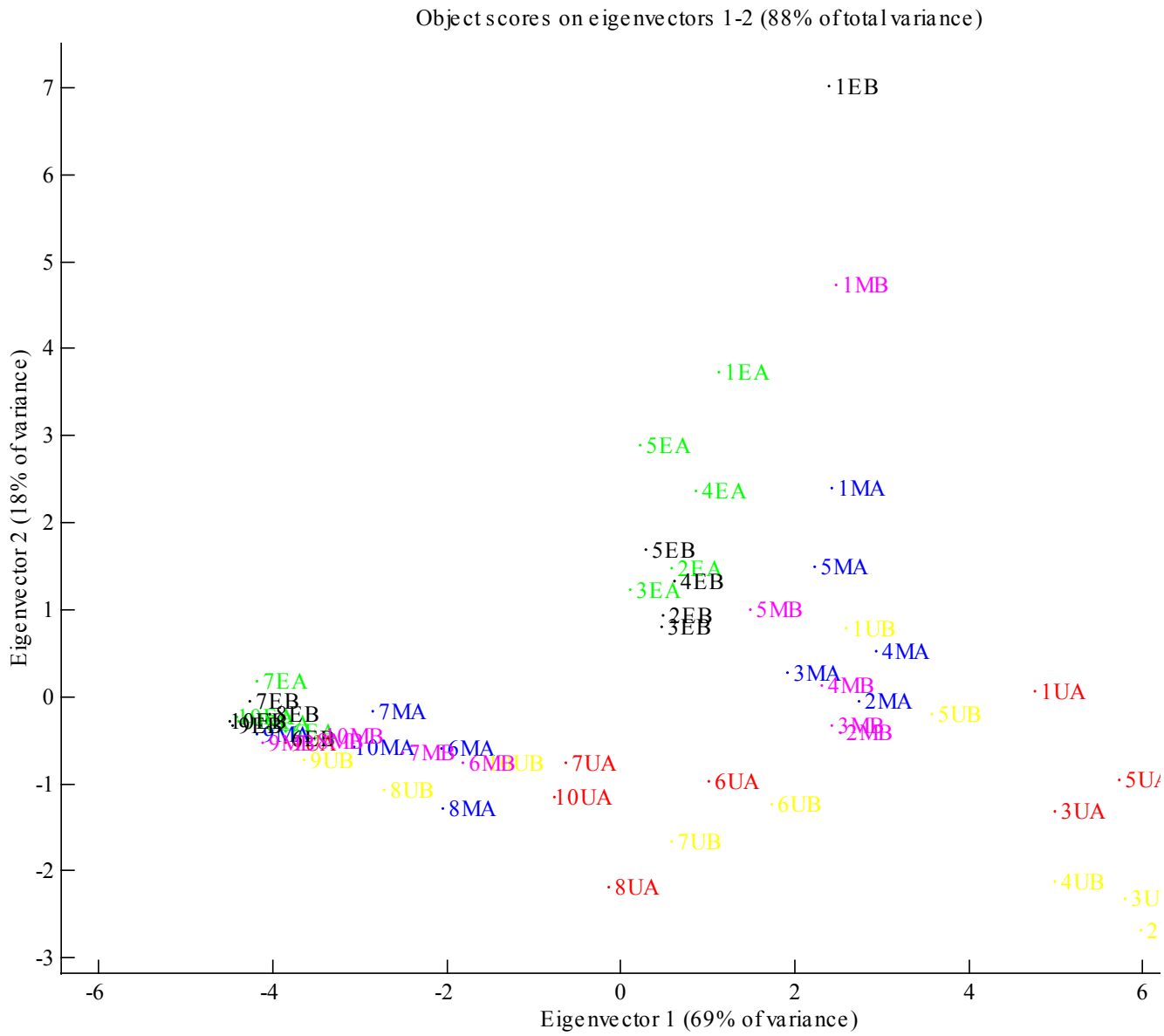
1) CO
2) Total hydrocarbons
3) NOx
4) Formic aldehyde
5) Acetic aldehyde
6) Total aldehydes
7) Ethylene
8) Propylene
9) Acetylene
10) 1,3-butadyene
11) Benzene
12) Ethylbenzene
13) p,m-xylene
14) o-xylene
15) Toluene
16) Total aromatic comp.

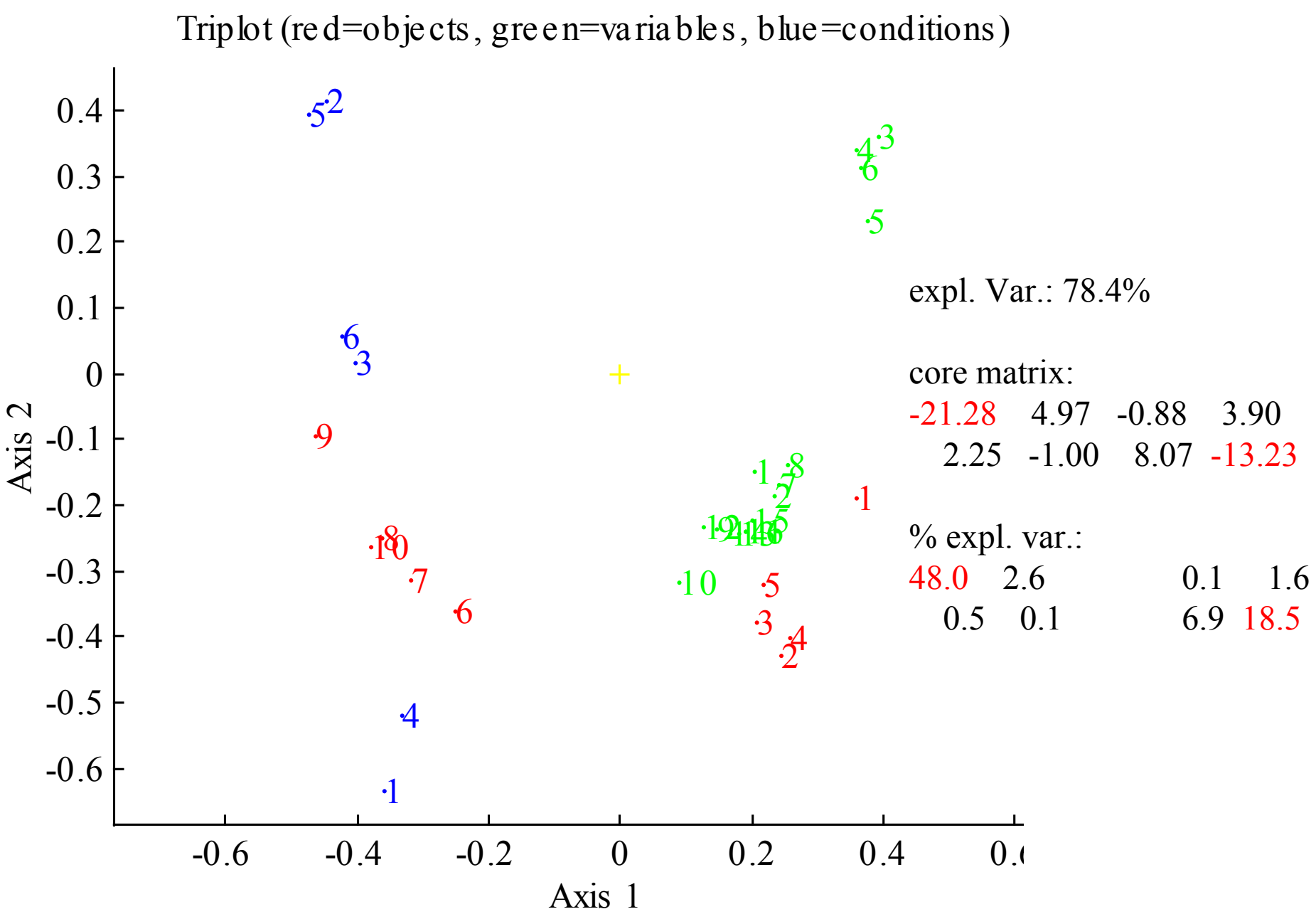
Conditions
(cycles and gasoline)

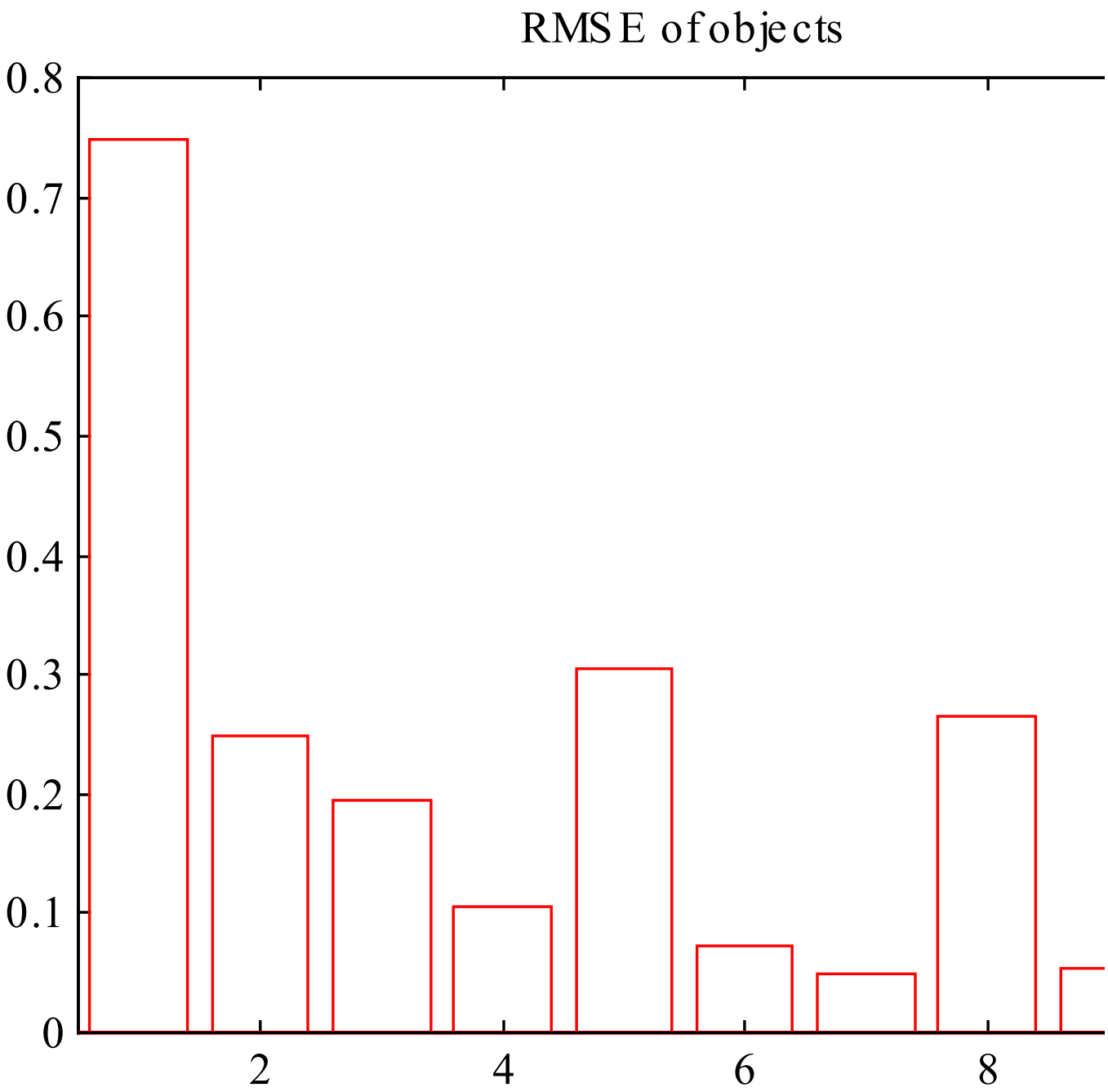
1) Urban, A
2) Extra-urban, A
3) Mixed, A
4) Urban, B
5) Extra-urban, B
6) Mixed, B


Object scores on eigenvectors 1-2 (88% of total variance)
Eigenvector 2 (18% of variance)
Eigenvector 1 (69% of variance)
1EB
1MB
1EA
5EA
4EA
1MA
5EB
2EA
5MA
4EB
3EA
5MB
2EB
3EB
1UB
4MA
3MA
4MB
1UA
7EA
7EB
2MA
5UB
7MA
3MB
2MB
9UB
7MB
6MA
7UA
6UA
8UB
10UA
6UB
3UA
8MA
7UB
8UA
4UB
7
6
5
4
3
2
1
0
-1
-2
-3
-6
-4
-2
0
2
4


Variable loadings on eigenvectors 1-2 (88% of total variance)
Eigenvector 2 (18% of variance)
Eigenvector 1 (69% of variance)
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.4
-0.2
0
0.2
0.4
4
6
3
5
2
8
7
9
1
10
12
14
13
16


Object scores on eigenvectors 1-2 (88% of total variance)
Eigenvector 2 (18% of variance)
Eigenvector 1 (69% of variance)
7
6
5
4
3
2
1
0
-1
-2
-3
-6
-4
-2
0
2
4
6
1EB
1MB
1EA
5EA
4EA
1MA
5EB
2EA
4EB
5MA
3EA
5MB
2EB
3EB
1UB
4MA
3MA
4MB
2MA
1UA
5UB
3MB
2MB
7EA
7EB
7MA
10MA
7MB
6MA
6MB
9UB
7UA
6UA
8UB
10UA
6UB
8MA
3UA
7UB
8UA
4UB
3U
2
5U


Triplot (red=objects, green=variables, blue=conditions)
Axis 2
Axis 1
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.6
-0.4
-0.2
0
0.2
0.4
expl. Var.: 78.4%
core matrix:
-21.28 4.97 -0.88 3.90
2.25 -1.00 8.07 -13.23
% expl. var.:
48.0 2.6 0.1 1.6
0.5 0.1 6.9 18.5


RMSE of objects
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2
4
6
8

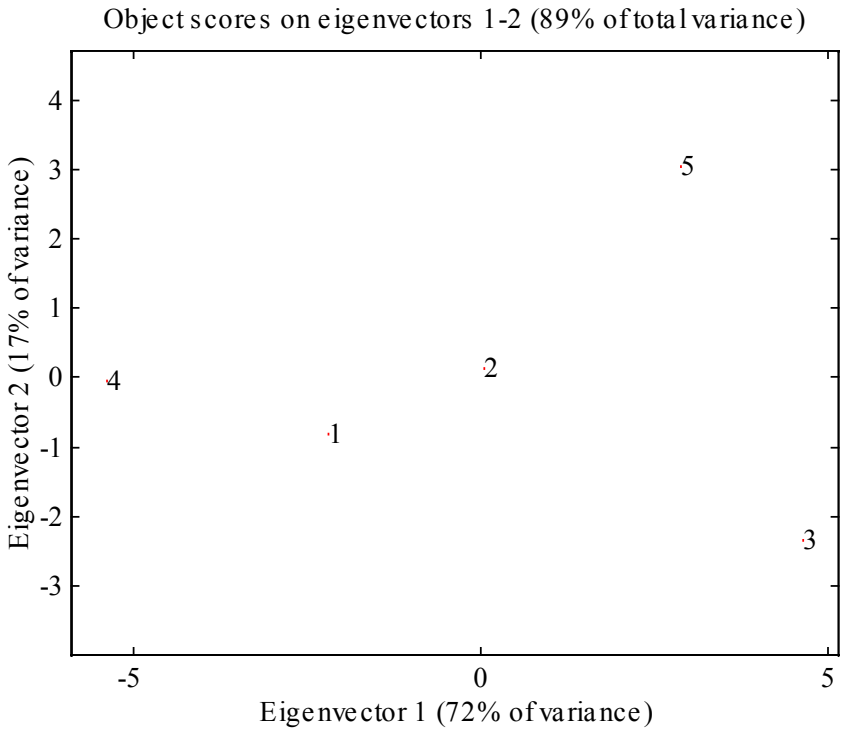
DATA SET PANEL TEST

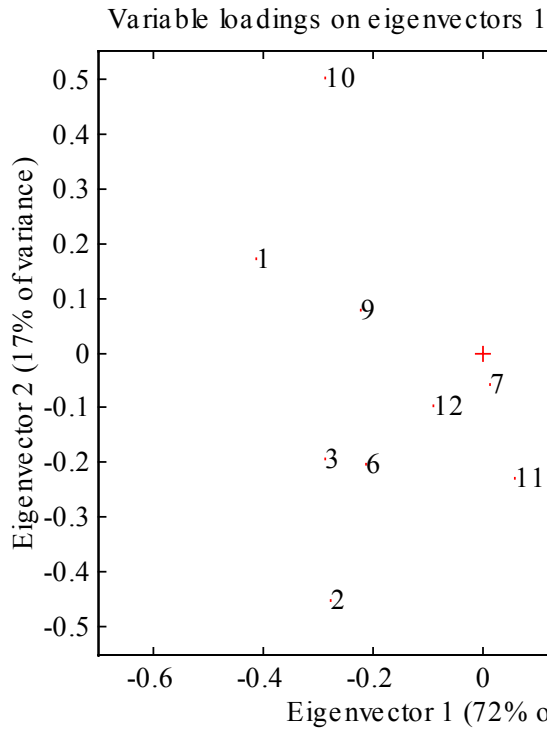
			RONDEUR	RANCIO	CASSIS	BOIS	LEGERETE	VIEUX	NETTETE	VIEUX CHAI	PAIN D'EPICE	LONGUEUR	PIQUANT	AMERTUME
JUGES	PRODUITS		1	2	3	4	5	6	7	8	9	10	11	12
1	1		5	7	8	4	2	8	8	3	7	8	5	7
1	2		6	4	2	5	3	6	9	4	5	4	2	3
1	3		2	4	3	8	6	3	8	8	1	2	7	5
1	4		9	8	5	2	1	5	9	2	5	6	7	8
1	5		3	2	2	4	9	4	9	1	8	6	3	3
2	1		8	7	1	2	1	8	7	4	5	6	1	1
2	2		6	6	3	6	7	6	7	3	4	7	1	3
2	3		6	5	3	6	5	5	7	5	3	3	3	3
2	4		8	8	1	2	2	7	7	3	5	7	1	1
2	5		3	3	3	6	7	5	7	3	1	5	3	3
3	1		8	6	5	6	5	7	7	6	10	9	0	0
3	2		5	4	0	6	9	7	10	7	9	8	2	0
3	3		4	6	2	8	8	8	7	8	8	4	2	2
3	4		9	6	6	5	6	6	7	4	10	7	0	0
3	5		8	5	3	6	9	7	9	6	9	9	0	0
4	1		4	6	6	4	6	3	5	3	2	4	1	0
4	2		8	8	2	5	2	7	7	4	6	3	0	0
4	3		2	8	5	8	8	5	7	5	5	2	1	3
4	4		7	4	8	3	4	3	7	1	2	6	0	5
4	5		5	2	3	6	6	1	6	2	2	3	0	0
5	1		4	5	3	4	2	4	5	5	4	2	0	1
5	2		6	7	5	8	4	8	6	7	6	7	0	6
5	3		4	5	3	6	7	6	8	8	3	4	0	3
5	4		7	9	8	4	1	10	6	1	9	9	0	8
5	5		6	3	1	4	9	2	6	6	1	8	0	6
6	1		6	5	2	3	4	5	10	2	1			
6	2		4	4	1	5	7	4	8	1	3			
6	3		2	2	0	7	7	2	10	5	1			
6	4		8	6	3	3	5	8	10	2	5			
6	5		4	3	1	8	8	5	7	4	4			

WHAT PEOPLE USUALLY DO:
LOOK AT AVERAGES

		RONDEUR	RANCIO	CASSIS	BOIS	LEGERETE	VIEUX	NETTETE	VIEUX CHAI	PAIN D'EPICE	LONGUEUR	PIQUANT	AMERTUME
		1	2	3	4	5	6	7	8	9	10	11	12
1		5.8	6.0	4.2	3.8	3.3	5.8	7.0	3.8	4.8	5.8	1.4	1.8
2		5.8	5.5	2.2	5.8	5.3	6.3	7.8	4.3	5.5	5.8	1.0	2.4
3		3.3	5.0	2.7	7.2	6.8	4.8	7.8	6.5	3.5	3.0	2.6	3.2
4		8.0	6.8	5.2	3.2	3.2	6.5	7.7	2.2	6.0	7.0	1.6	4.4
5		4.8	3.0	2.2	5.7	8.0	4.0	7.3	3.7	4.2	6.2	1.2	2.4

ONE STEP FORWARD:
PCA ON THE AVERAGE SCORES





Since it is based on average scores, an "hypothetical judge" is looked at, and there is no idea about the "experimental error" of the different judges and of the different attributes

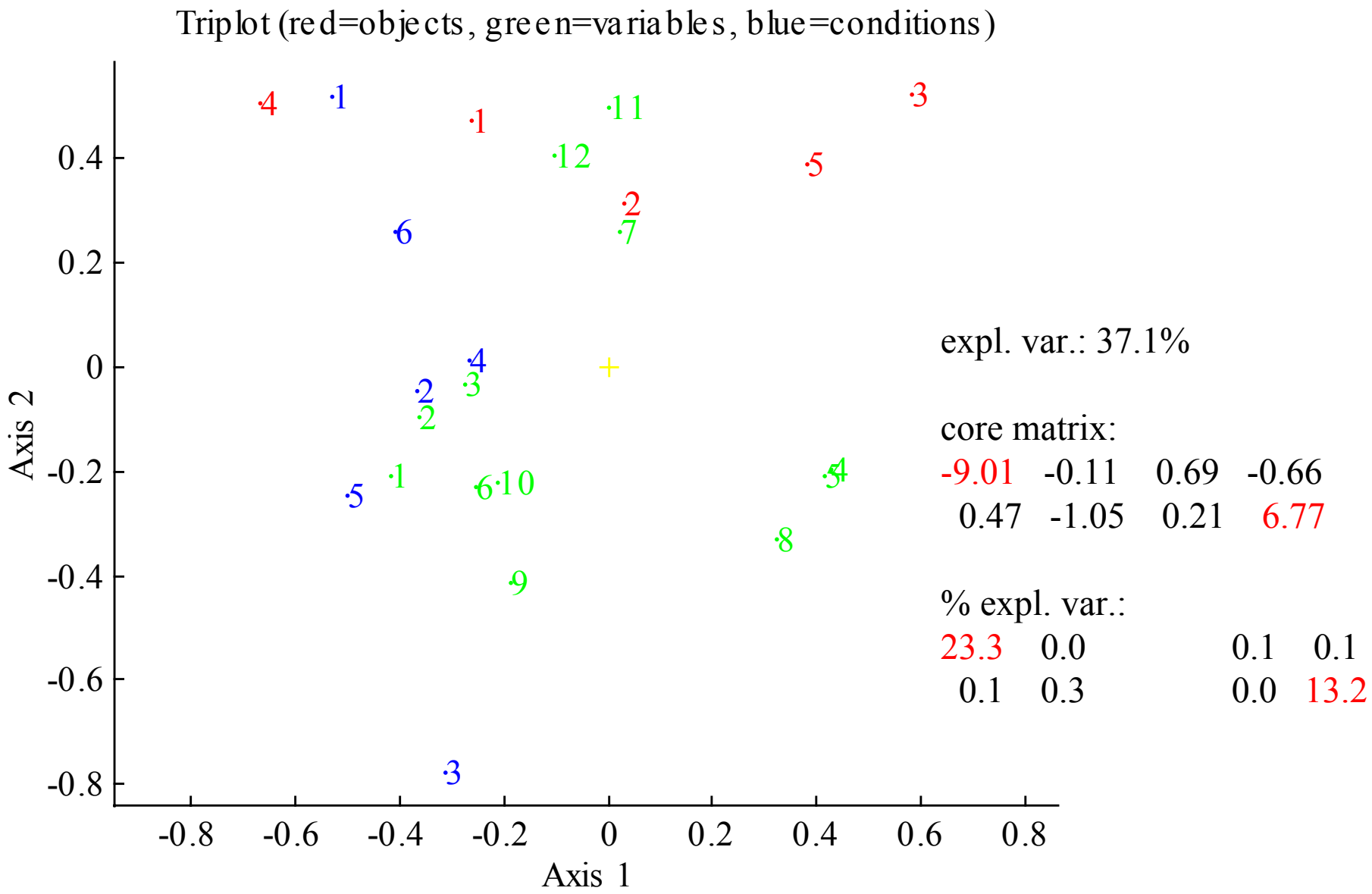
The missing data are not taken into account, and therefore the averages can be biased depending on which data are missing

THREE-WAY PCA

It takes into account also the effect of the judges: this means that the way of scoring of each of them is taken into account, not just the “average”.

It is possible to “reconstruct” the missing data.

JUGES	PRODUITS	RONDEUR	RANCIO	CASSIS	BOIS	LEGERETE	VIEUX	NETTETE	VIEUX CHAI	PAIN D'EPICE	LONGUEUR	PIQUANT	AMERTUME
		1	2	3	4	5	6	7	8	9	10	11	12
1	1	5	7	8	4	2	8	8	3	7	8	5	7
1	2	6	4	2	5	3	6	9	4	5	4	2	3
1	3	2	4	3	8	6	3	8	8	1	2	7	5
1	4	9	8	5	2	1	5	9	2	5	6	7	8
1	5	3	2	2	4	9	4	9	1	8	6	3	3
2	1	8	7	1	2	1	8	7	4	5	6	1	1
2	2	6	6	3	6	7	6	7	3	4	7	1	3
2	3	6	5	3	6	5	5	7	5	3	3	3	3
2	4	8	8	1	2	2	7	7	3	5	7	1	1
2	5	3	3	3	6	7	5	7	3	1	5	3	3
3	1	8	6	5	6	5	7	7	6	10	9	0	0
3	2	5	4	0	6	9	7	10	7	9	8	2	0
3	3	4	6	2	8	8	8	7	8	8	4	2	2
3	4	9	6	6	5	6	6	7	4	10	7	0	0
3	5	8	5	3	6	9	7	9	6	9	9	0	0
4	1	4	6	6	4	6	3	5	3	2	4	1	0
4	2	8	8	2	5	2	7	7	4	6	3	0	0
4	3	2	8	5	8	8	5	7	5	5	2	1	3
4	4	7	4	8	3	4	3	7	1	2	6	0	5
4	5	5	2	3	6	6	1	6	2	2	3	0	0
5	1	4	5	3	4	2	4	5	5	4	2	0	1
5	2	6	7	5	8	4	8	6	7	6	7	0	6
5	3	4	5	3	6	7	6	8	8	3	4	0	3
5	4	7	9	8	4	1	10	6	1	9	9	0	8
5	5	6	3	1	4	9	2	6	6	1	8	0	6
6	1	6	5	2	3	4	5	10	2	1	6	2	3
6	2	4	4	1	5	7	4	8	1	3	6	2	3
6	3	2	2	0	7	7	2	10	5	1	6	1	3
6	4	8	6	3	3	5	8	10	2	5	5	2	3
6	5	4	3	1	8	8	5	7	4	4	6	1	3


Triplot (red=objects, green=variables, blue=conditions)
Axis 2
Axis 1
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
expl. var.: 37.1%
core matrix:
-9.01 -0.11 0.69 -0.66
0.47 -1.05 0.21 6.77
% expl. var.:
23.3 0.0 0.1 0.1
0.1 0.3 0.0 13.2

DATA SET STRAWBERRIES

(C. Patz, *Research Center Geisenheim, Department of Wine Analysis and Beverage Research, Germany*)

12 cultivars:

A) Andana
B) Arena
C) 88009/o2v2
D) 88009/o3v3
E) Cijosee
F) Cirano
G) Elsanta
H) Honeoye
I) Kimberly
J) Lambada
K) Pavana
L) Vima Zanta

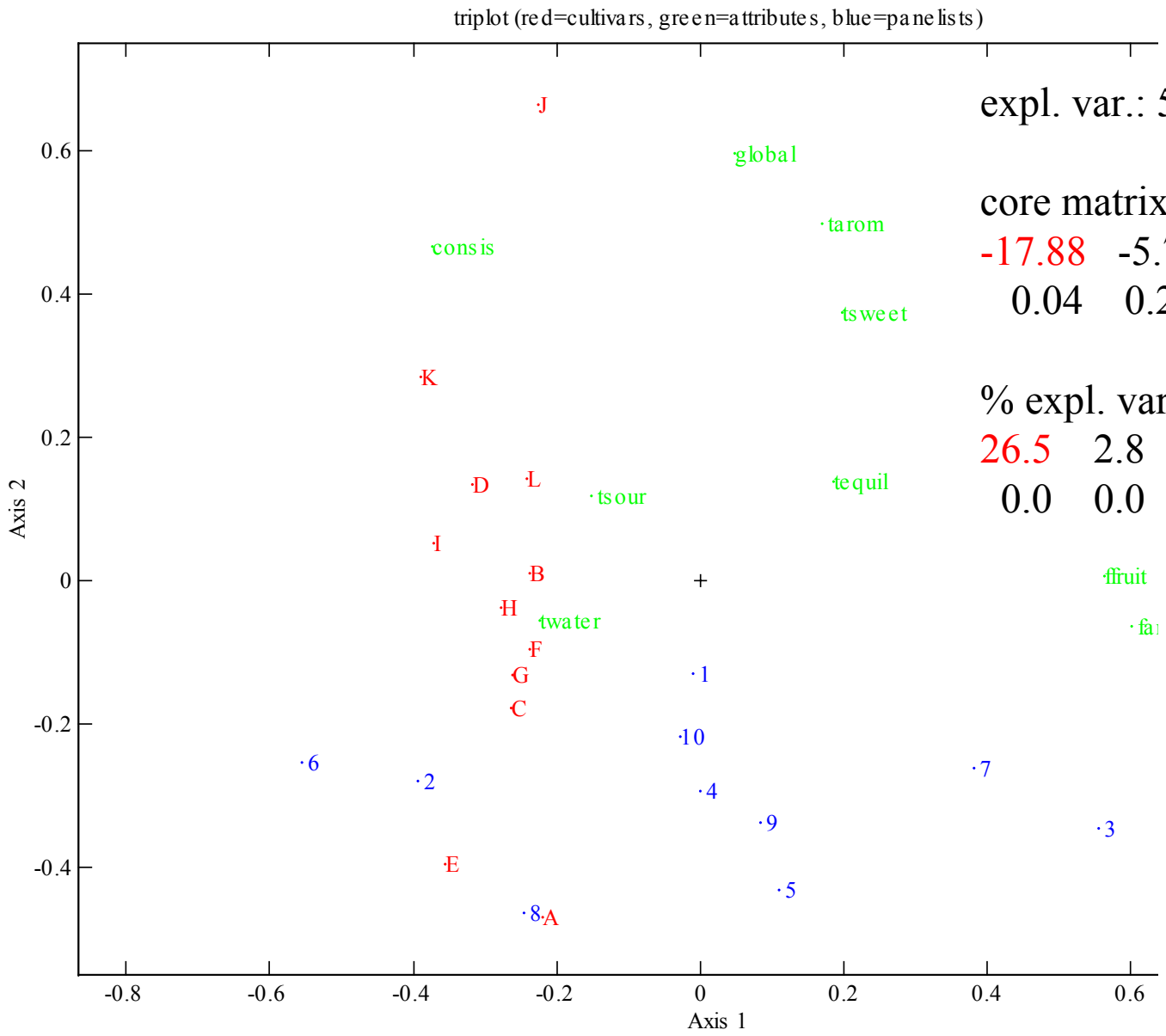
9 attributes:

a) aromatic (flavour)
b) fruity (flavour)
c) sweet (taste)
d) sour (taste)
e) sweet/sour equilibrium
f) aromatic (taste)
g) watery (taste)
h) consistency
i) global score

10 panelists:

Panelist 01
...
Panelist 10

(with several missing data)



expl. var.: 53.3%

core matrix:

-17.88	-5.77	0.41	-0.10
0.04	0.26	-6.49	-15.77

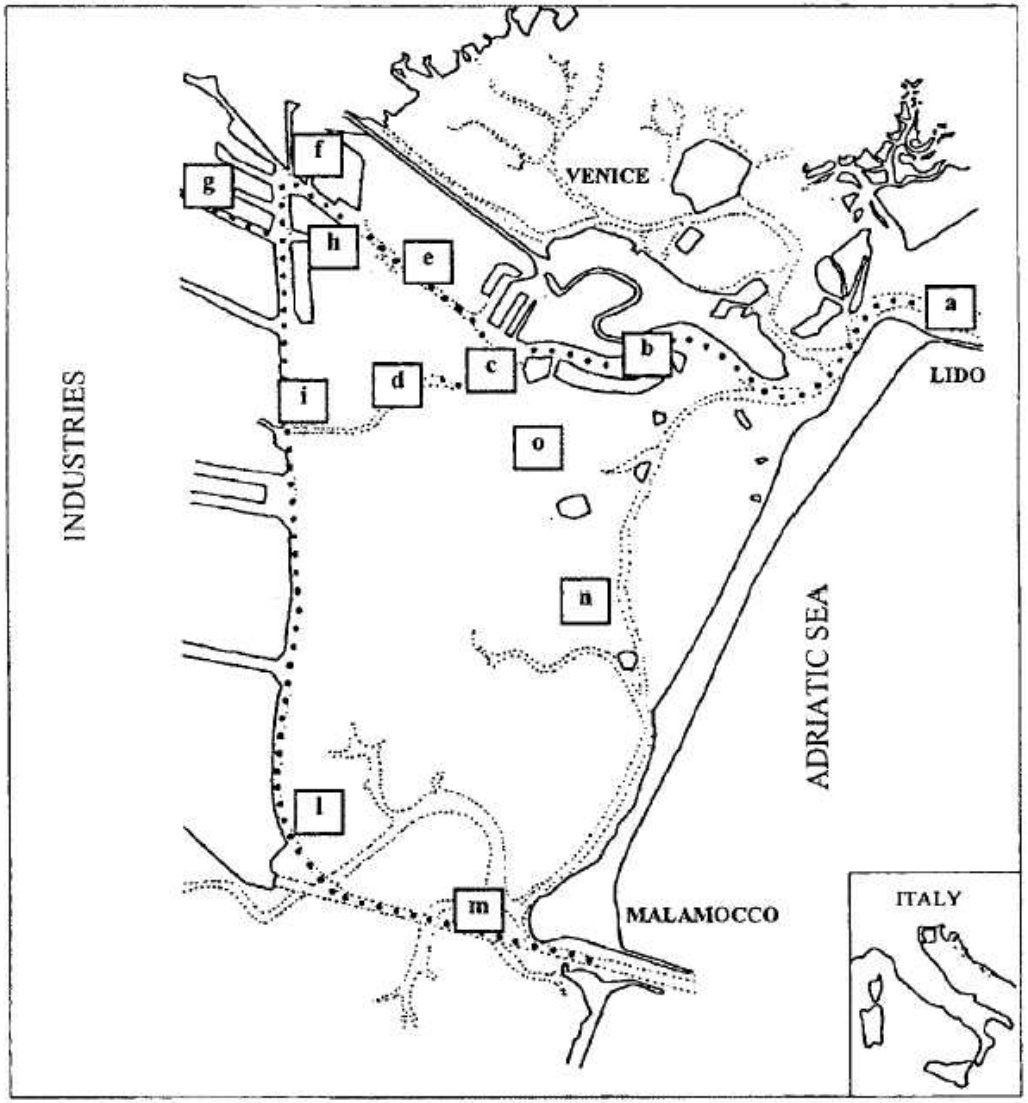
% expl. var.:

26.5	2.8	0.0	0.0
0.0	0.0	3.5	20.6

DATA SET VENICE (II)

(L. Alberotanza, *Istituto per lo Studio della Dinamica delle Grandi Masse, Venezia, Italy*)

Objects: 13 sampling sites



Variables:

1) chlorophyll-α
2) total suspended matter
3) water transparency
4) fluorescence
5) turbidity
6) suspended solids
7) NH_4^+
8) NO_3^-
9) P
10) COD
11) BOD_5

Samplings:

1) May ‘87
2) June ‘87
...
44) December ‘90

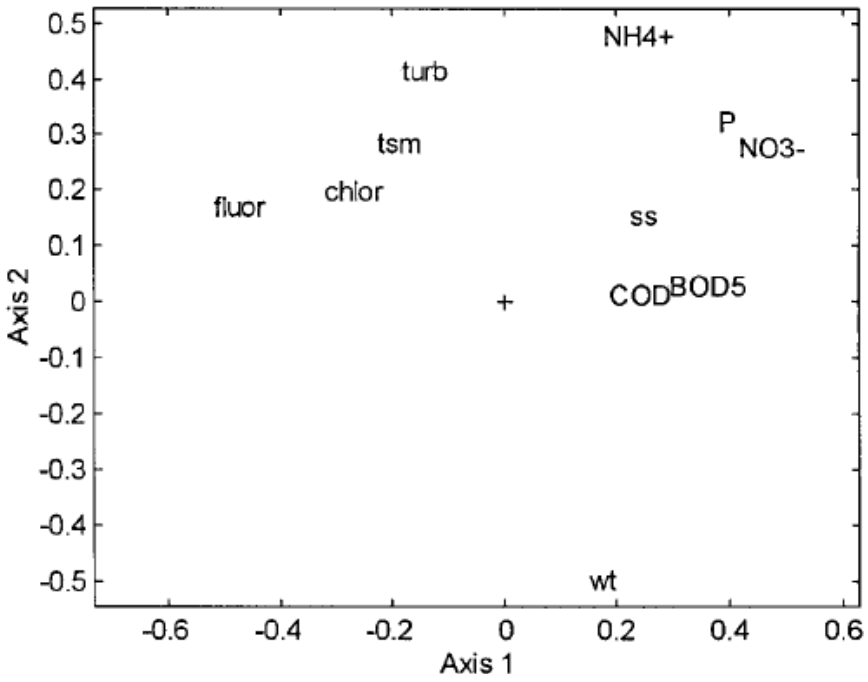
explained variance: 34.6%

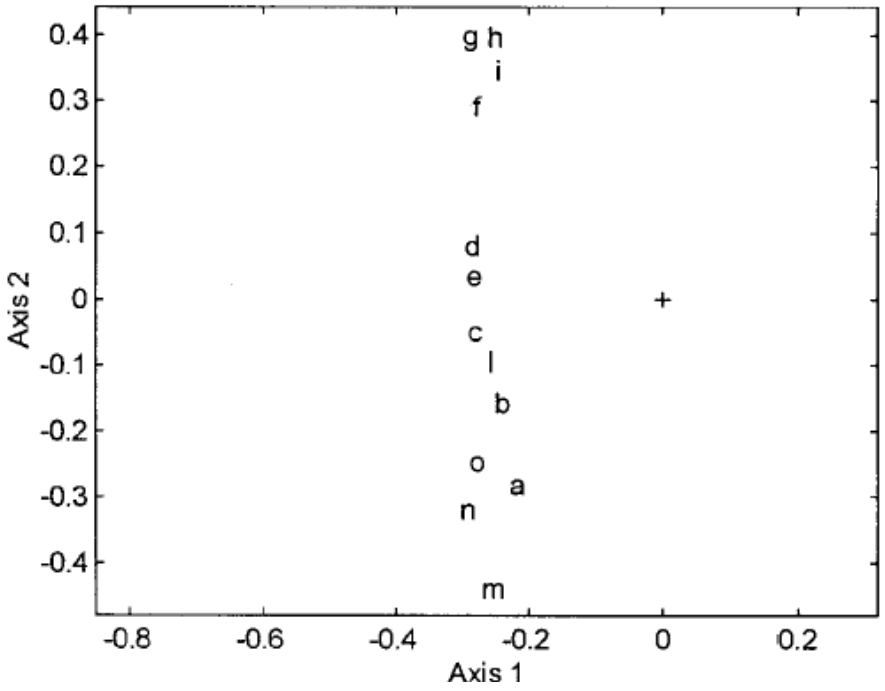
core matrix:

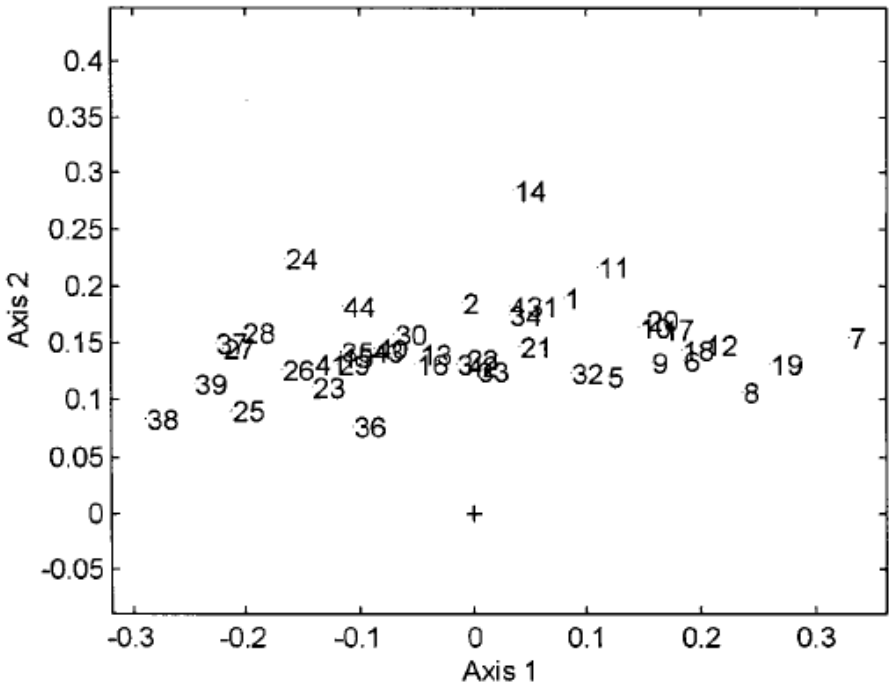
-34.94	-1.99	1.86	-1.96
-1.39	2.12	-2.84	-30.48

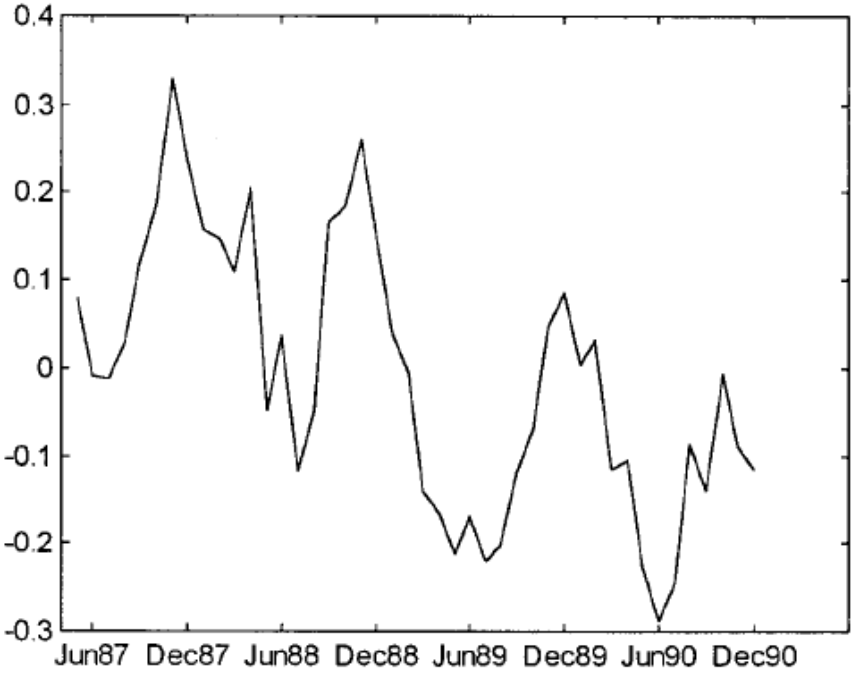
% explained variance:

19.4	0.1	0.1	0.1
0.0	0.1	0.1	14.8


NH4+
turb
P
NO3-
tsm
chlor
fluor
ss
COD
BOD5
wt
Axis 1
Axis 2


g h
i
f
d
e
c
l
b
o
a
n
m
Axis 1
Axis 2


Axis 1
Axis 2


Jun87
Dec87
Jun88
Dec88
Jun89
Dec89
Jun90
Dec90

FIELD OF APPLICATION	OBJECTS	VARIABLES	CONDITIONS
environmental analysis	air or water samples (different locations)	chemico-physical analyses	time
environmental analysis	water samples (different locations)	chemico-physical analyses	depth
panel tests	food products (wines, oils, …)	attributes	assessors
food chemistry	foods (cheeses, spirits, …)	chemical composition	aging
food chemistry	foods (oils, wines, …)	chemical composition	crops
sport medicine	athletes	blood analyses	time after effort
process monitoring	batches	chemical analyses	time