Categorical Functional Data Analysis applied to Temporal Dominance of Sensations data

Caroline Peltier, Michel Visalli, Pascal Schlich, Hervé Cardot

Centre des Sciences du Goût et de l'Alimentation

AGRO SUP DIJON · cnrs · INRAE · uB · UBFC Université Bourgogne Franche-Comté

ChemoSens

www.chemosens.fr
chemosens@inrae.fr

CERTIFIED ISO 9001

IBiSA Infrastructures en Biologie Santé et Agronomie

Categorical Functional Data Analysis applied to Temporal Dominance of Sensations data

Caroline Peltier, Michel Visalli, Pascal Schlich, Hervé Cardot
caroline.peltier@inrae.fr

PROBE Platform for profiling properties of food and biobased products

LLOYD'S REGISTER QUALITY ASSURANCE ISO9001

IBiSA

Outline

- Presentation of Temporal Dominance of Sensations (TDS) and Categorical Functional Data Analysis (CFDA)

- Illustration of the application of CFDA on a simulated dataset

- Illustration on real TDS dataset

Center of Taste and Feeding Behavior

A problematic:
What happens during a tasting?

Two approaches to understand

Chemical methods: measuring directly in the mouth

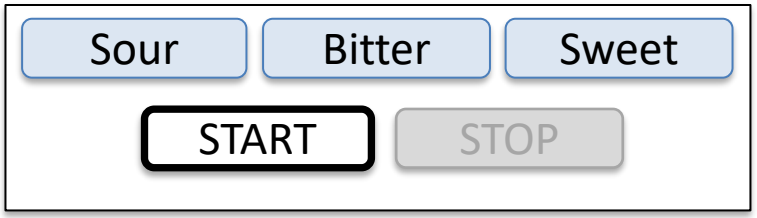
Sensory methods: asking the subjects what they perceive

Static or **temporal** data…

ChemoSens

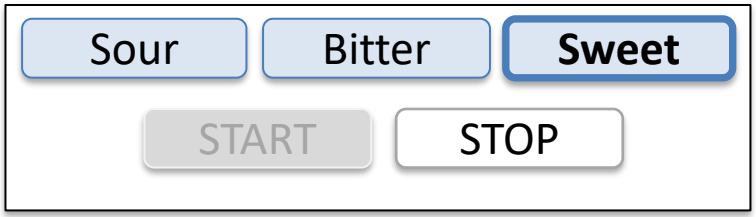
Temporal Dominance of Sensations

Temporal Dominance of Sensations (TDS): protocol of sensory analysis that consists in asking the subject to declare dynamically the dominant sensation (one among a list of descriptors)



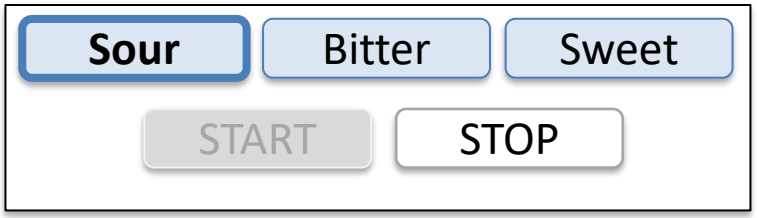
Temporal Dominance of Sensations

Temporal Dominance of Sensations (TDS): protocol of sensory analysis that consists in asking the subject to declare dynamically the dominant sensation (one among a list of descriptors)



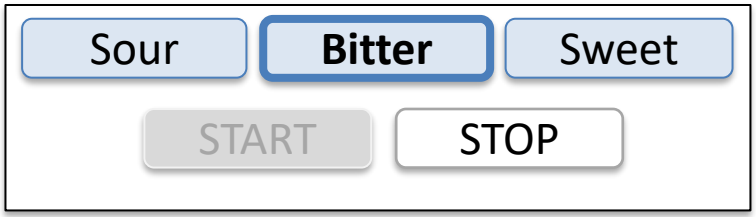
Temporal Dominance of Sensations

Temporal Dominance of Sensations (TDS): protocol of sensory analysis that consists in asking the subject to declare dynamically the dominant sensation (one among a list of descriptors)



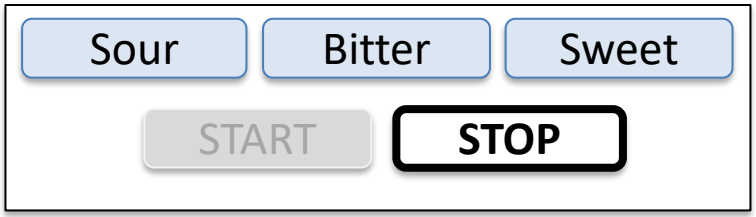
Temporal Dominance of Sensations

Temporal Dominance of Sensations (TDS): protocol of sensory analysis that consists in asking the subject to declare dynamically the dominant sensation (one among a list of descriptors)



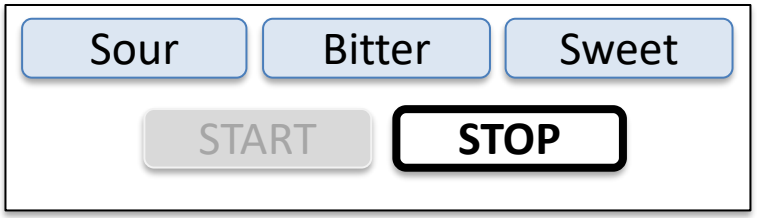
Temporal Dominance of Sensations

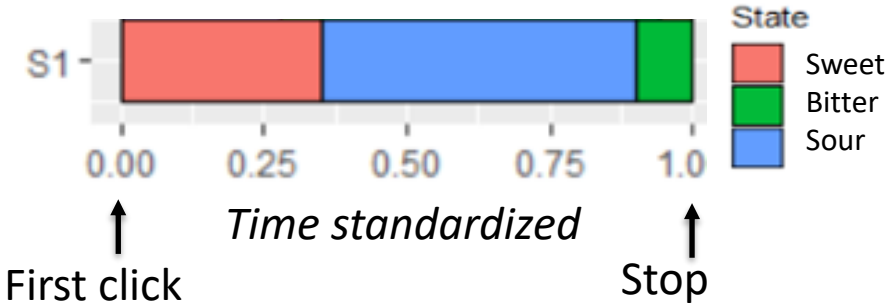
Temporal Dominance of Sensations (TDS): protocol of sensory analysis that consists in asking the subject to declare dynamically the dominant sensation (one among a list of descriptors)



Temporal Dominance of Sensations

Temporal Dominance of Sensations (TDS): protocol of sensory analysis that consists in asking the subject to declare dynamically the dominant sensation (one among a list of descriptors)





These are categorical functional data… a method adapted for this data is categorical functional data analysis (CFDA)

History of Categorical Functional Data Analysis

CFDA as « analyse harmonique qualitative »

DEVILLE J.C. et SAPORTA G. (1979) Analyse harmonique qualitative. In Data Analysis and Informatics (E. DIDAY et a l . E d i t o r s) , 375-389, North-Holland, Amsterdam.

1979

1981

GILBERT SAPORTA

Méthodes exploratoires d'analyse de données temporelles

Cahiers du Bureau universitaire de recherche opérationnelle. Série Recherche, tome 37-38 (1981), p. 7-194

<http://www.numdam.org/item?id=BURO_1981__37-38__7_0>

1982

Analyses de données chronologiques qualitatives : comment analyser des calendriers?

par Jean-Claude DEVILLE*

J.-C. Deville

1996

L'analyse harmonique qualitative, une synthèse de la théorie

Gilbert Saporta

G. Saporta

Didactic synthesis of CFDA

C. Preda

2021

mathematics MDPI

Article

Categorical Functional Data Analysis. The cfda R Package

Cristian Preda [1,2,3,*], Quentin Grimonprez [4] and Vincent Vandewalle [3,5]

ChemoSens

CFDA: an extension of Correspondance Analysis

Correspondence Analysis: a usual statistical analysis allowing two qualitative variables X: $\Omega \to S_1$ and Y : $\Omega \to S_2$ to be conjointly analyzed.

It can also be seen as optimal encoding finding numerical values associated to each modality by finding $\varphi_1 : S_1 \to \mathbb{R}$ and $\varphi_2 : S_2 \to \mathbb{R}$ minimizing:

$$E\left(\left(\varphi_1(X) - \varphi_2(Y)\right)^2\right)$$

Under constraints $E\left(\varphi_1(X)\right) = E\left(\varphi_2(Y)\right) = 0$ and $E(\varphi_1(X)^2) = E(\varphi_2(Y)^2) = 1$

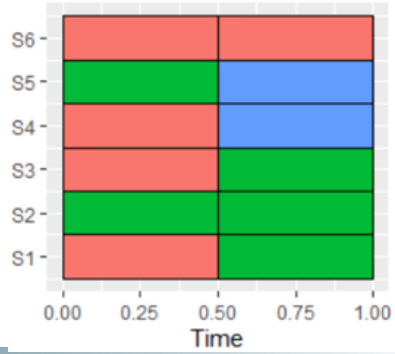
Example of encoding with previous data considering
X: state chosen during period 1 and Y: state chosen during period 2

Period. 1

Modality	Numerical Encoding
	NA
	-0.35
	0.7

Period. 2

Modality	Numerical Encoding
	-0.35
	0.7
	0



ChemoSens

CFDA in brief

X_t: $\Omega \to S = \{x_1, \dots, x_S\}$: $(X_t)_{t \in T}$ is a random qualitative process = qualitative functional data

Idea of CFDA: Finding optimal encodings of the states evolving over time

$$\varphi : S \times T \to \mathbb{R}$$
$$(x, t) \to \varphi(x, t),$$

such as the encoded process $\varphi(X_t, t)$ minimizes

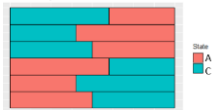
$$\iint_{T \times T} E((\varphi(X_t, t) - \varphi(X_s, s))^2) dt\, ds$$

u. c. $\forall t, E\big(\varphi(X_t, t)\big) = 0; Var\big(\varphi(X_t, t)\big) = 1$

$\varphi(X_t, .)$ can be approximated by projection on a basis of functions

In practice…

Data

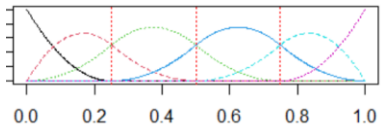


id	state	time
P2S1	A	0.0
P2S1	C	0.5
P2S1	C	1.0
P2S2	A	0.0
P2S2	C	0.4
P2S2	C	1.0
P2S3	A	0.0
P2S3	C	0.6
P2S3	C	1.0
P3S1	C	0.0
P3S1	A	0.5
P3S1	A	1.0
P3S2	C	0.0
P3S2	A	0.4
P3S2	A	1.0
P3S3	C	0.0
P3S3	A	0.6
P3S3	A	1.0

Basis of functions

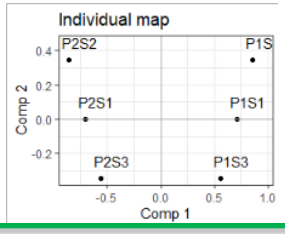
order=3

n_bases=6

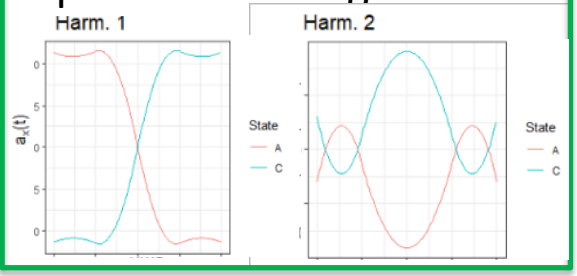


CFDA

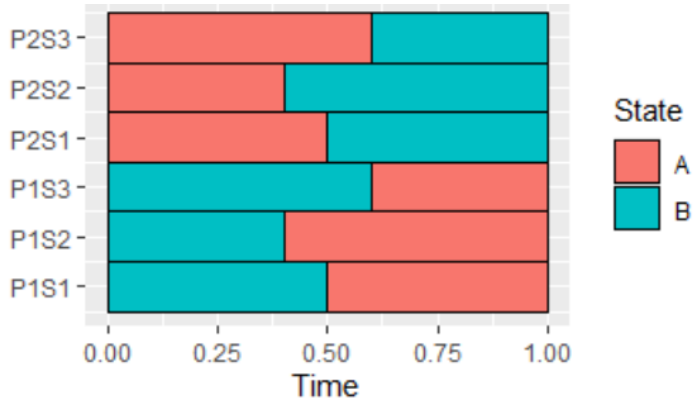
Coordinates of individuals



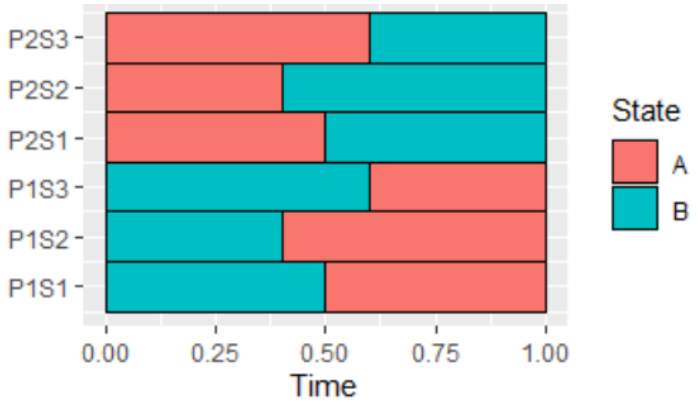
Optimal encodings of states

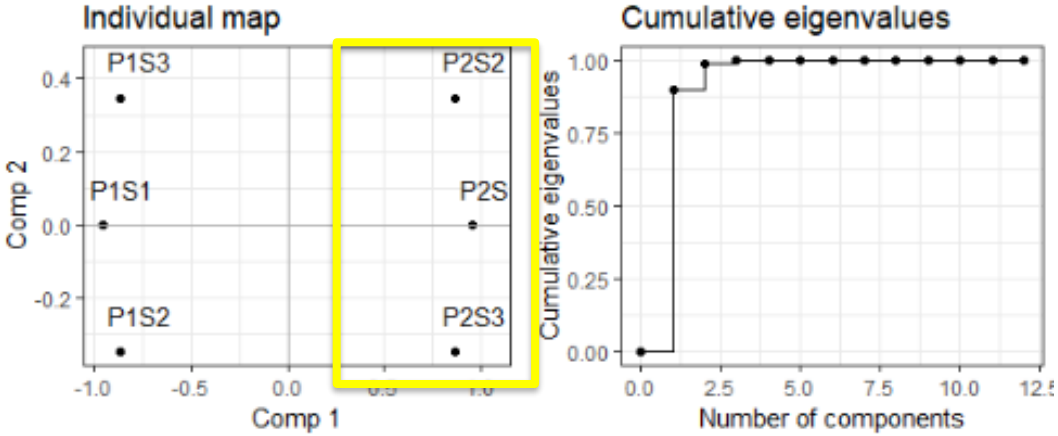


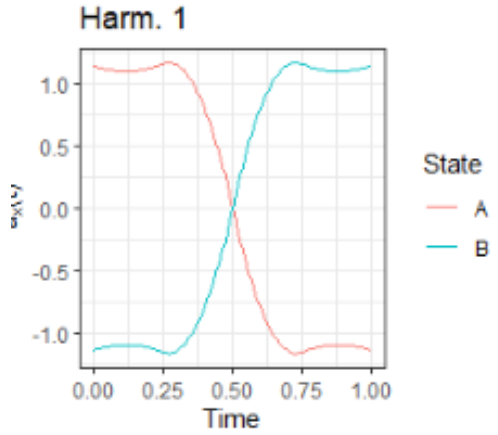
Simulation

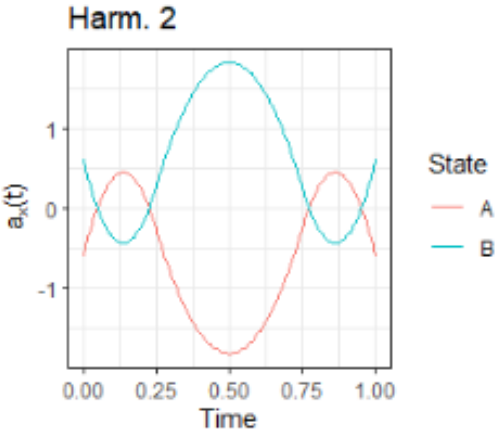

P2S3
P2S2
P2S1
P1S3
P1S2
P1S1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

Simulation

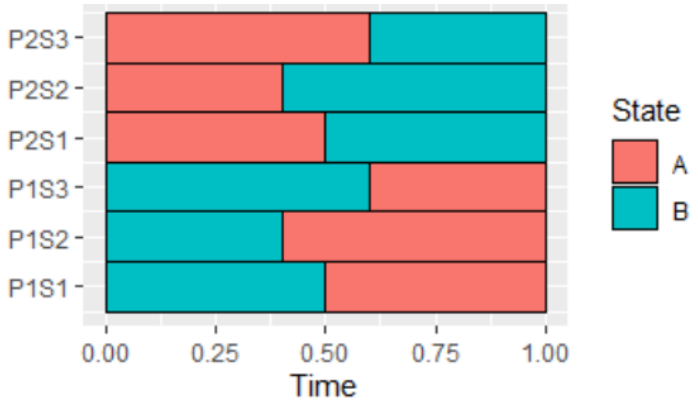

P2S3
P2S2
P2S1
P1S3
P1S2
P1S1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

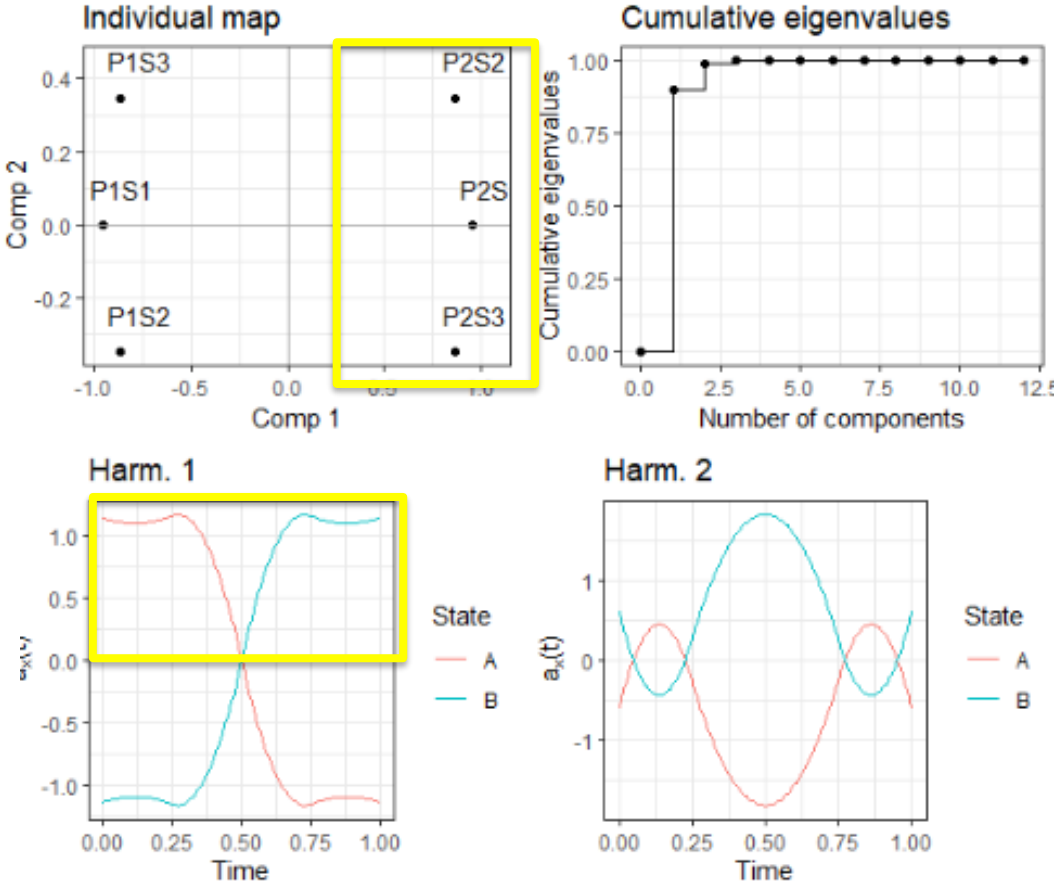

Individual map
P1S3
P2S2
P1S1
P2S
P1S2
P2S3
Comp 2
Comp 1
Cumulative eigenvalues
Cumulative eigenvalues
Number of components


Harm. 1
Time
State
A
B

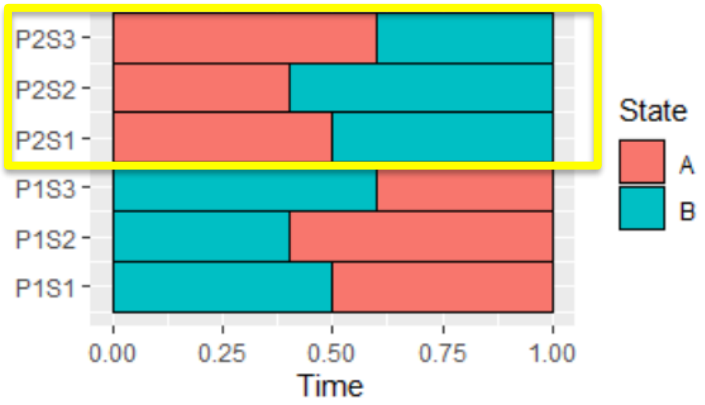

Harm. 2
Time
State
A
B

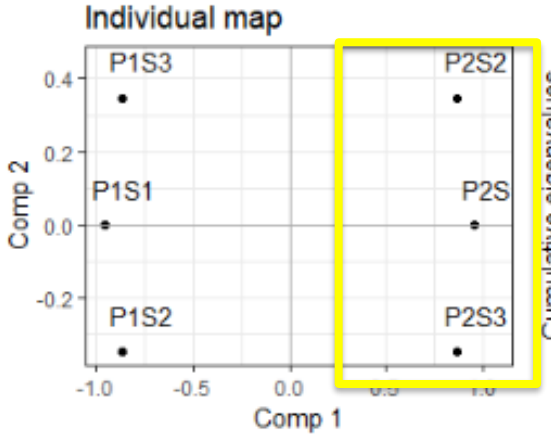
Simulation

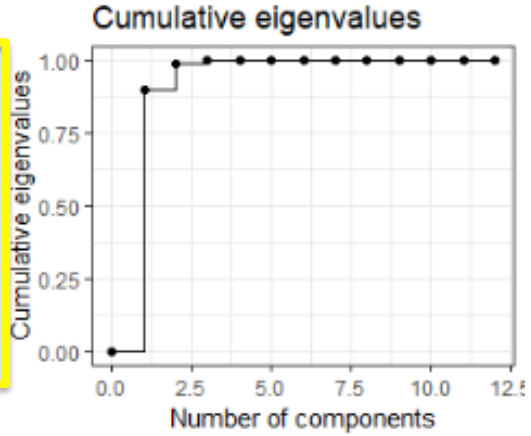

P2S3
P2S2
P2S1
P1S3
P1S2
P1S1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

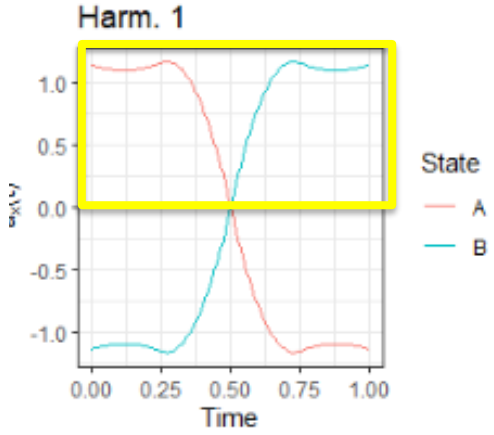

Individual map
P1S3
P2S2
P1S1
P2S
P1S2
P2S3
Comp 2
Comp 1
Cumulative eigenvalues
Cumulative eigenvalues
Number of components
Harm. 1
Time
State
A
B
Harm. 2
Time
State
A
B

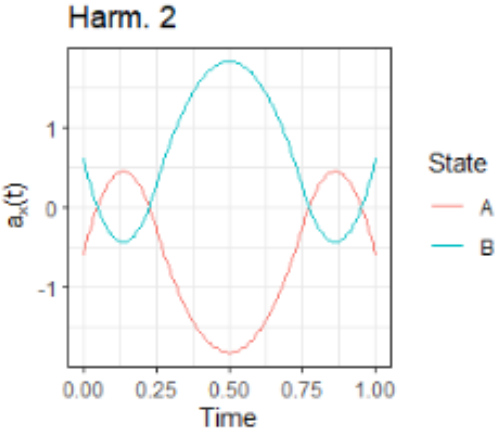
Simulation


P2S3
P2S2
P2S1
P1S3
P1S2
P1S1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

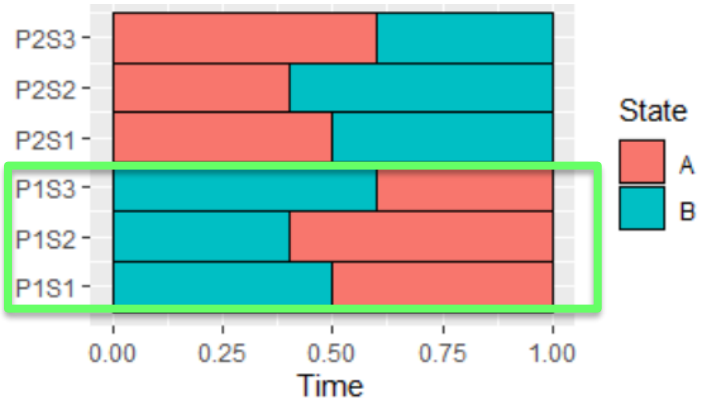

Individual map
P1S3
P2S2
P1S1
P2S
P1S2
P2S3
0.4
0.2
0.0
-0.2
Comp 2
-1.0
-0.5
0.0
Comp 1

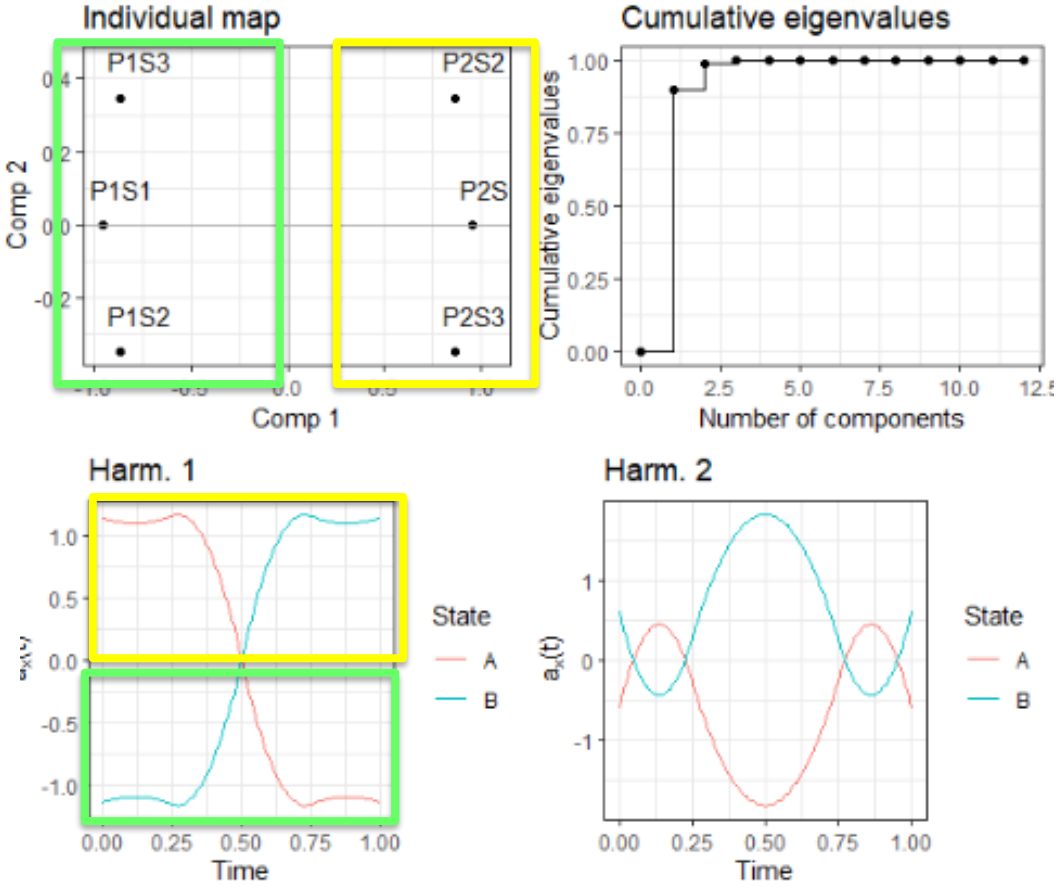

Cumulative eigenvalues
1.00
0.75
0.50
0.25
0.00
Cumulative eigenvalues
0.0
2.5
5.0
7.5
10.0
Number of components


Harm. 1
1.0
0.5
0.0
-0.5
-1.0
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

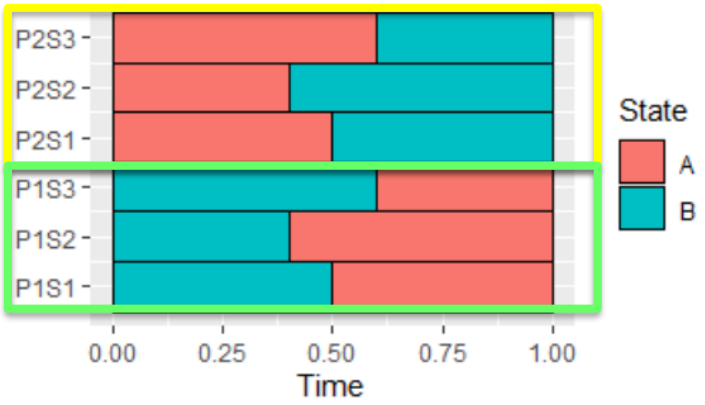

Harm. 2
1
0
-1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

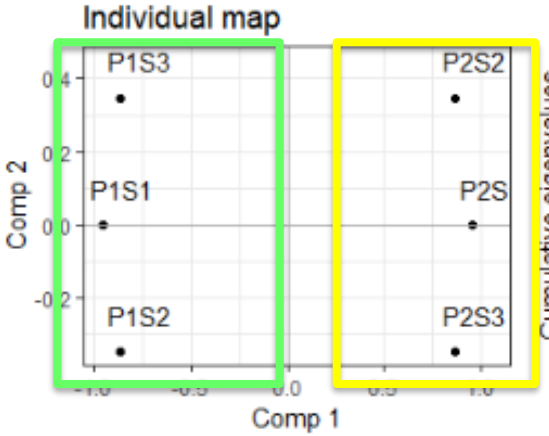
Simulation

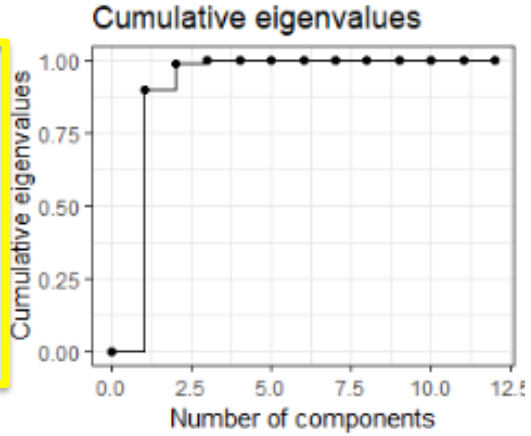

P2S3
P2S2
P2S1
P1S3
P1S2
P1S1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

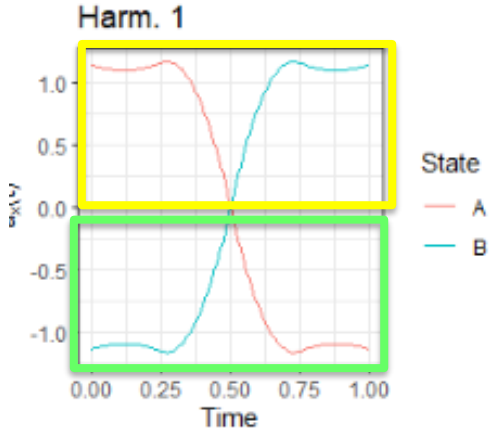

Individual map
P1S3
P2S2
P1S1
P2S
P1S2
P2S3
Comp 2
Comp 1
Cumulative eigenvalues
Cumulative eigenvalues
Number of components
Harm. 1
Time
State
A
B
Harm. 2
Time
State
A
B

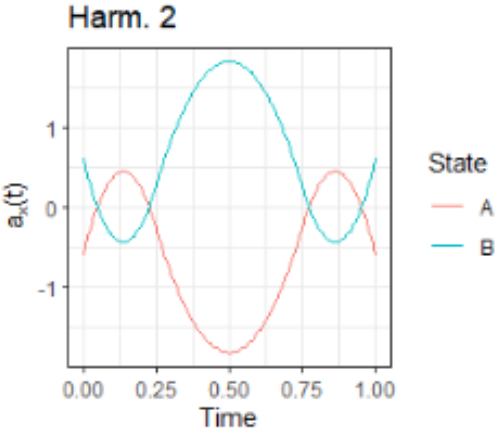
Simulation


P2S3
P2S2
P2S1
P1S3
P1S2
P1S1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

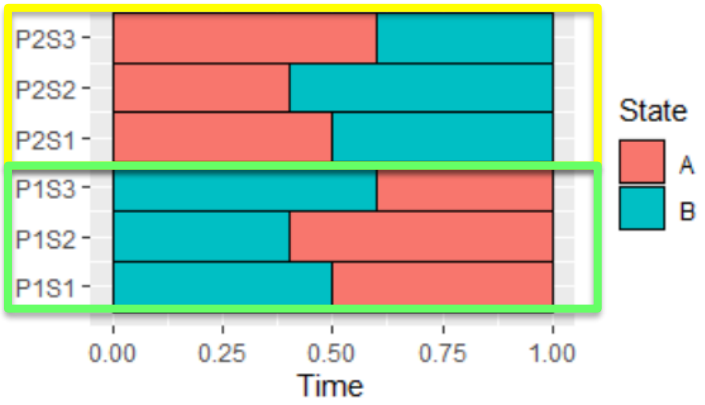

Individual map
P1S3
P2S2
P1S1
P2S
P1S2
P2S3
Comp 2
Comp 1

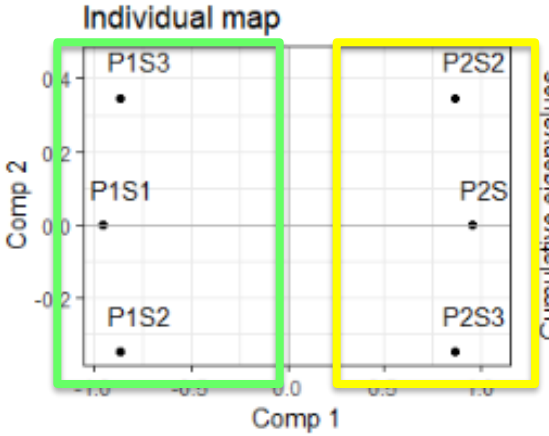

Cumulative eigenvalues
Cumulative eigenvalues
1.00
0.75
0.50
0.25
0.00
0.0
2.5
5.0
7.5
10.0
Number of components

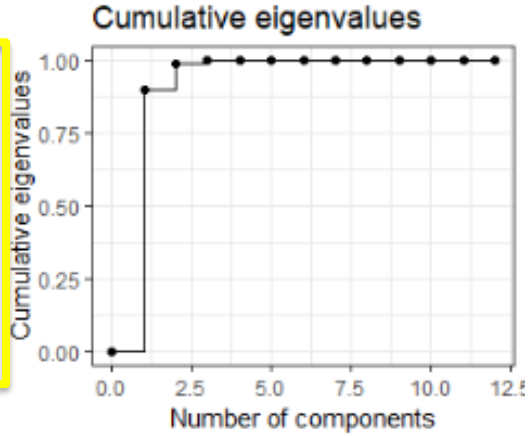

Harm. 1
1.0
0.5
0.0
-0.5
-1.0
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

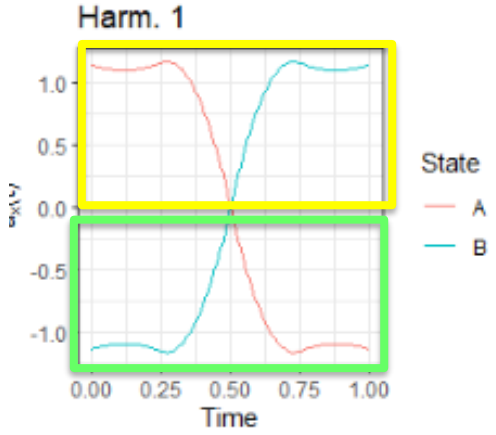

Harm. 2
1
0
-1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

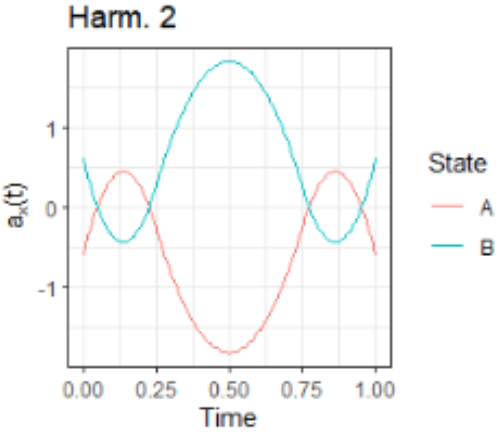
Simulation


P2S3
P2S2
P2S1
P1S3
P1S2
P1S1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

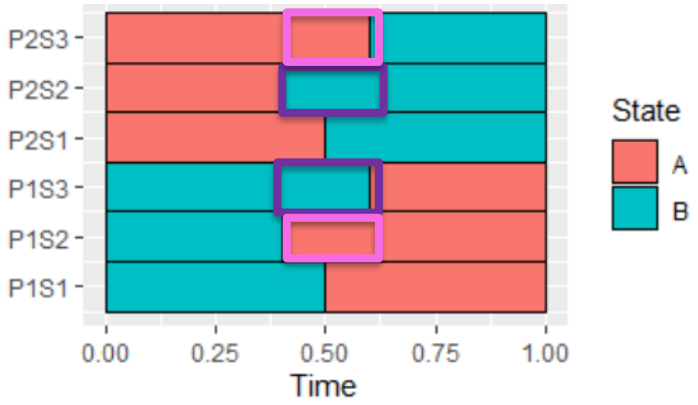

Individual map
P1S3
P2S2
P1S1
P2S
P1S2
P2S3
Comp 2
Comp 1

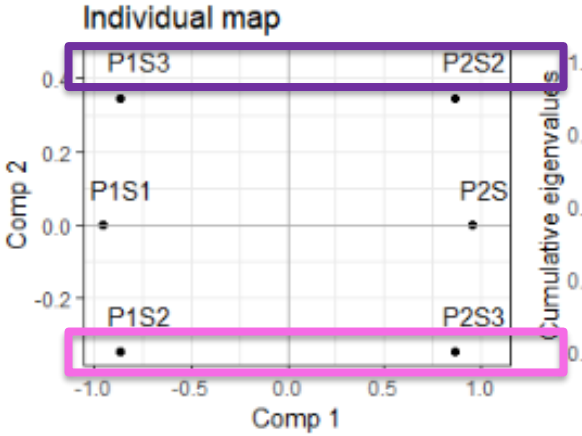

Cumulative eigenvalues
Cumulative eigenvalues
1.00
0.75
0.50
0.25
0.00
0.0
2.5
5.0
7.5
10.0
Number of components

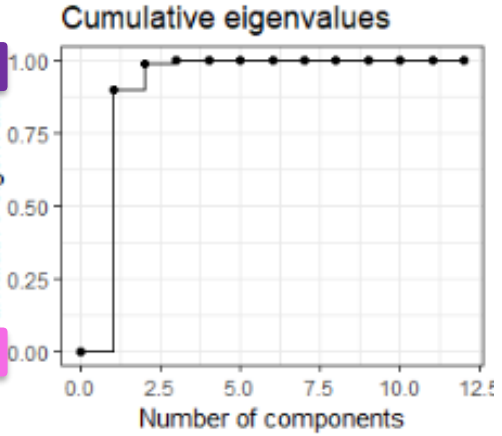

Harm. 1
1.0
0.5
0.0
-0.5
-1.0
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

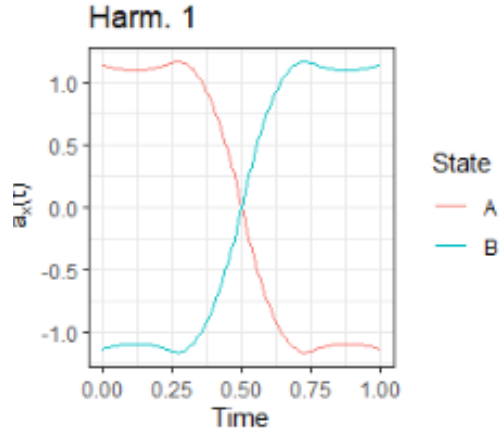

Harm. 2
1
0
-1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B

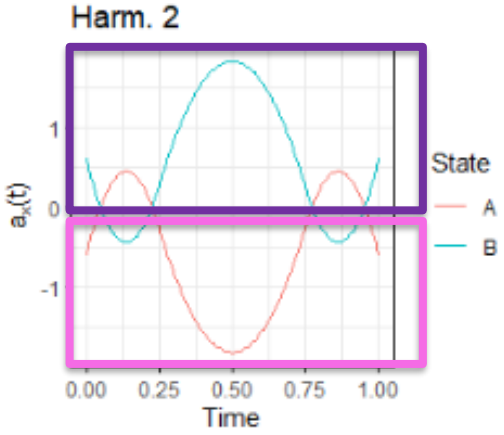
Simulation


P2S3
P2S2
P2S1
P1S3
P1S2
P1S1
0.00
0.25
0.50
0.75
1.00
Time
State
A
B


Individual map
P1S3
P2S2
P1S1
P2S
P1S2
P2S3
Comp 2
Comp 1
0.4
0.2
0.0
-0.2
-1.0
-0.5
0.0
0.5
1.0


Cumulative eigenvalues
Cumulative eigenvalues
Number of components
1.00
0.75
0.50
0.25
0.00
0.0
2.5
5.0
7.5
10.0

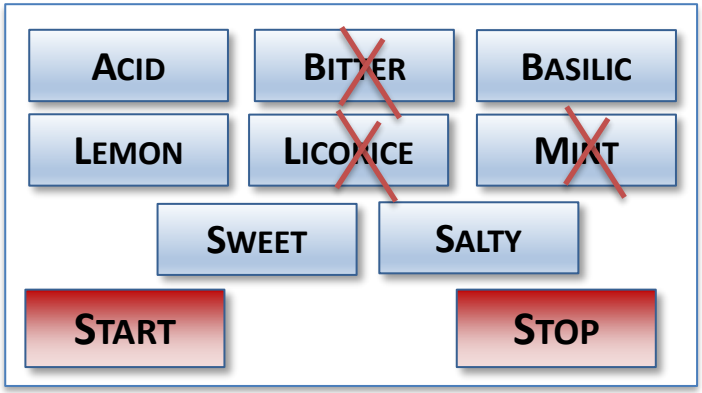

Harm. 1
$a_x(t)$
Time
1.0
0.5
0.0
-0.5
-1.0
0.00
0.25
0.50
0.75
1.00
State
A
B


Harm. 2
$a_x(t)$
Time
1
0
-1
0.00
0.25
0.50
0.75
1.00
State
A
B

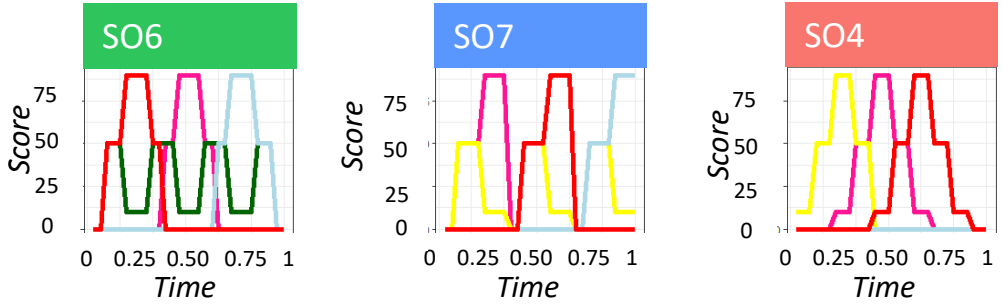
A real dataset with gustometer

A gustometer delivers controlled solutions in the mouth

TDS screen



Study on a dataset with three signals evaluated by 50 subjects on 8 descriptors[1,2]

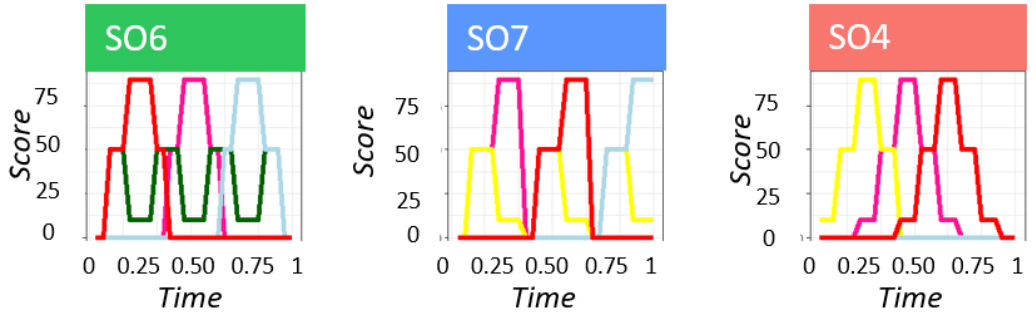


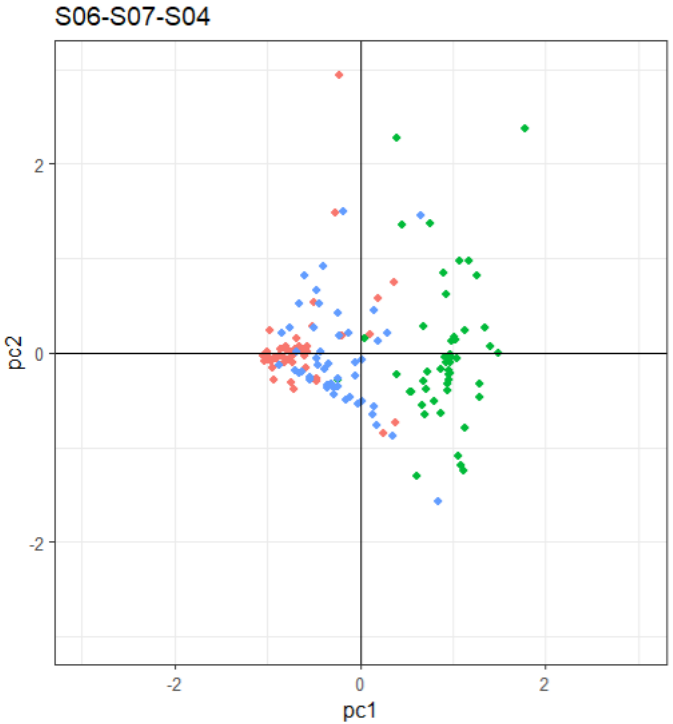
[1]Beno, Nicolle, Visalli (2023) A dataset of consumer perceptions of gustometer-controlled stimuli measured with three temporal sensory evaluation methods. Data in Brief

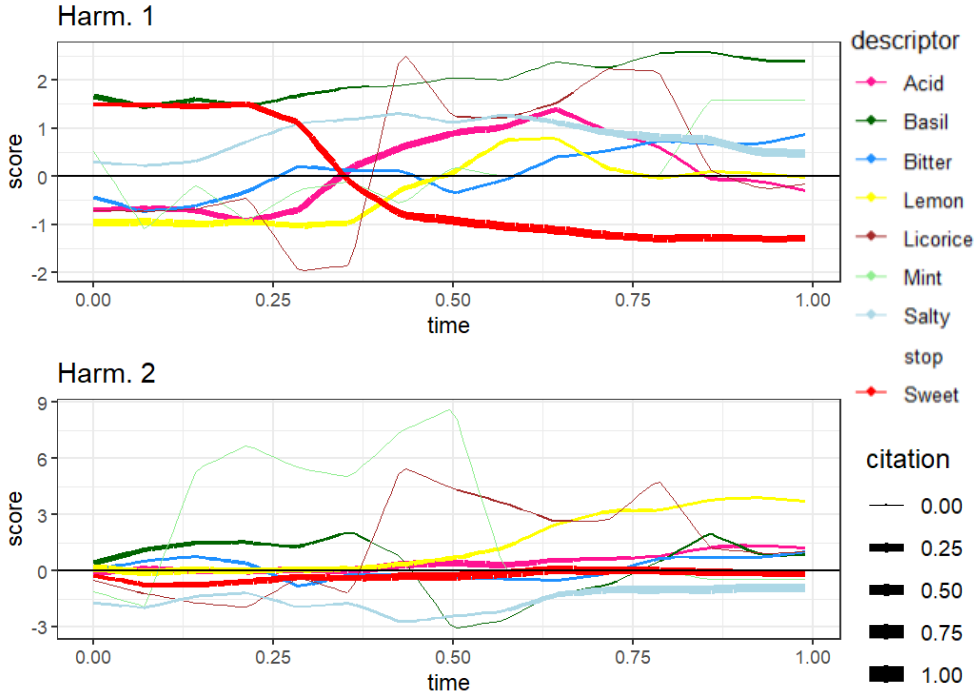
[2] Visalli, M., Beno, N., Nicolle, L., Schlich, P. (2023) Assessment of the validity and reliability of temporal sensory evaluation methods used with consumers on controlled stimuli delivered by a gustometer. Food Quality and Preference 110

CFDA results

- Linewidth of the harmonic proportional to the number of citations

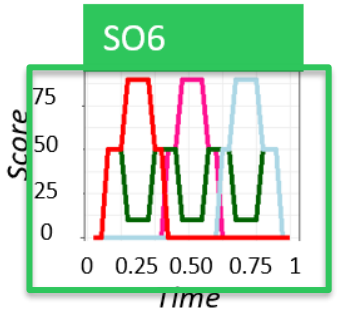


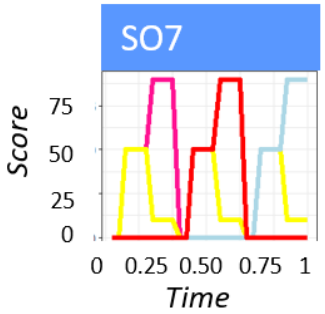


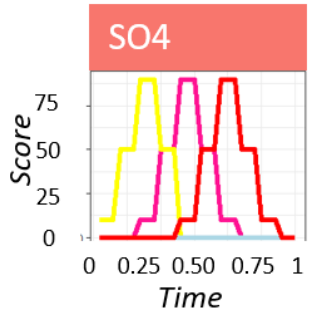


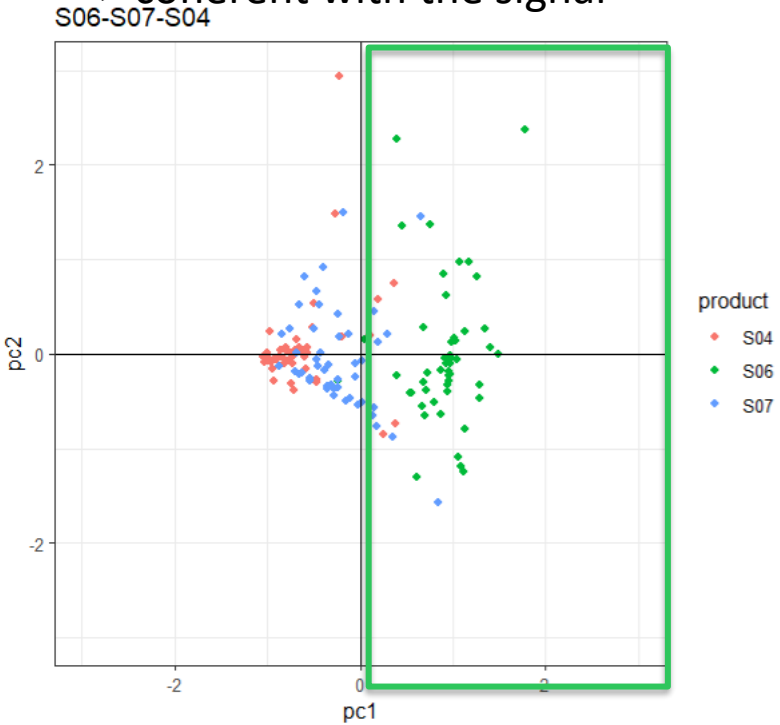
CFDA results

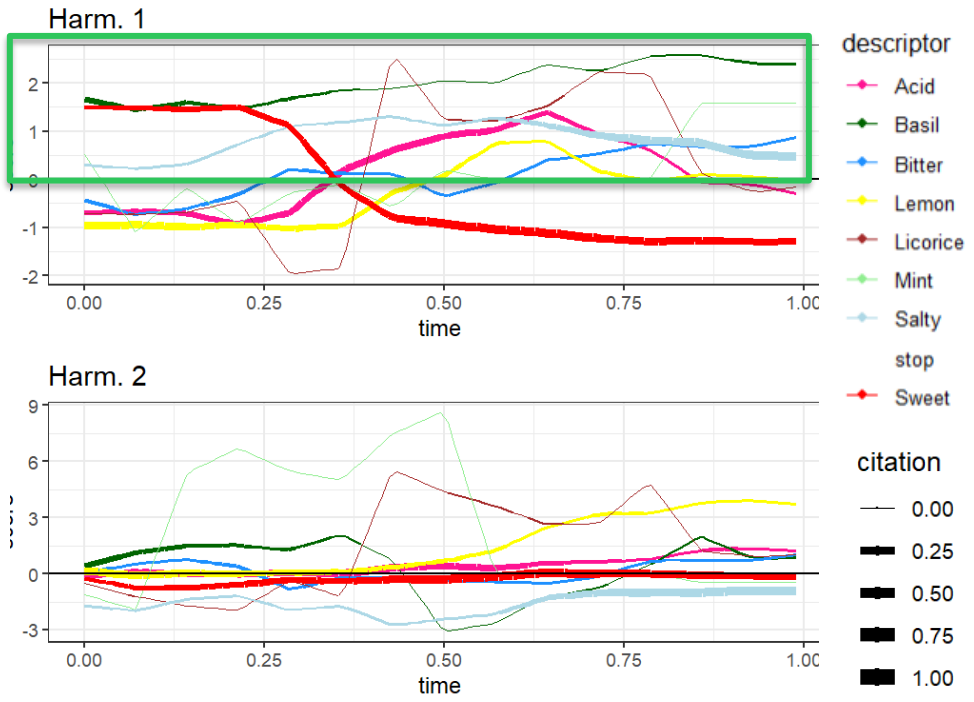
- Positive scores on the first component: related to signal S06
- **Sweet** at the beginning, then **Acid**. Basilic all the time
- => coherent with the signal





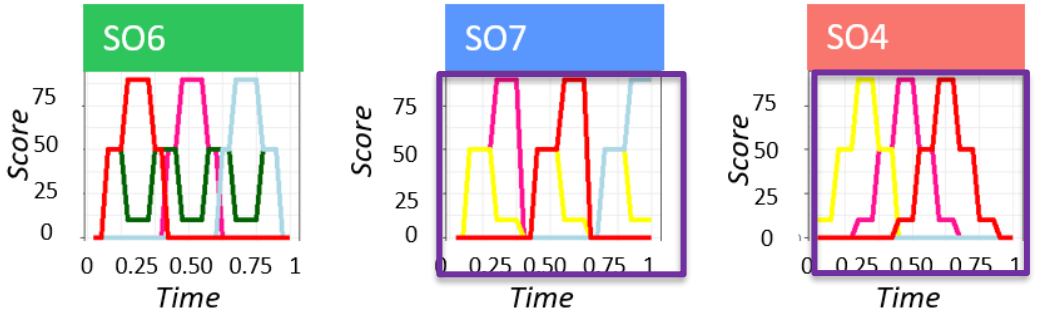


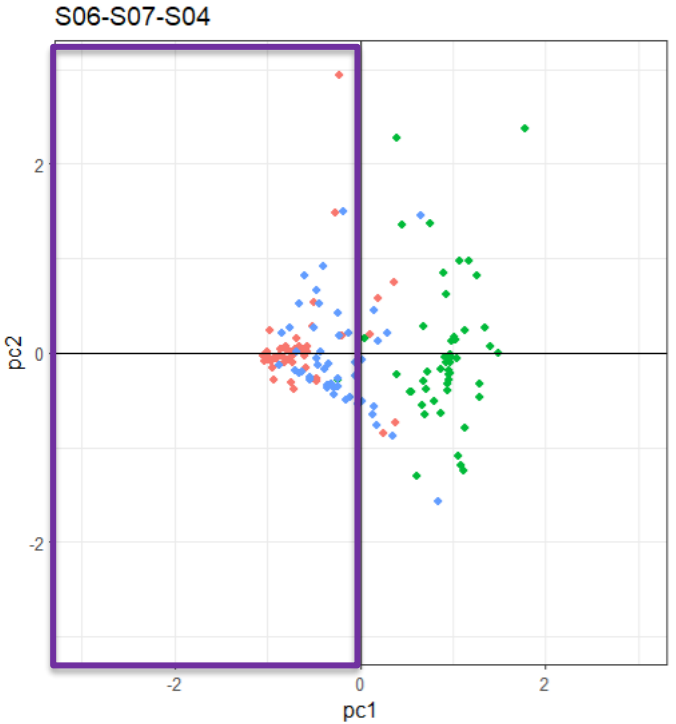


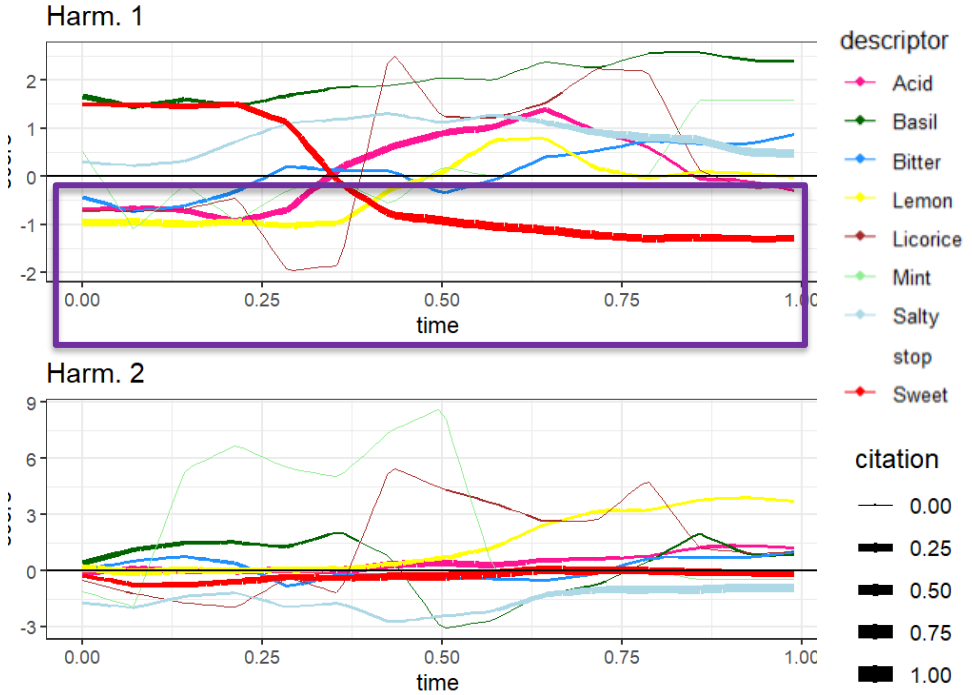


CFDA results

- Negative scores on the first component: related to signal S07 and S04
- **Acid** and Lemon at the begining then **Sweet** => coherent with the signal

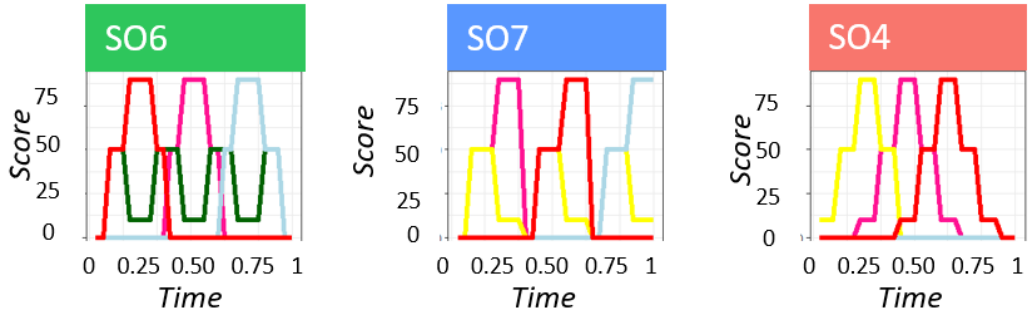


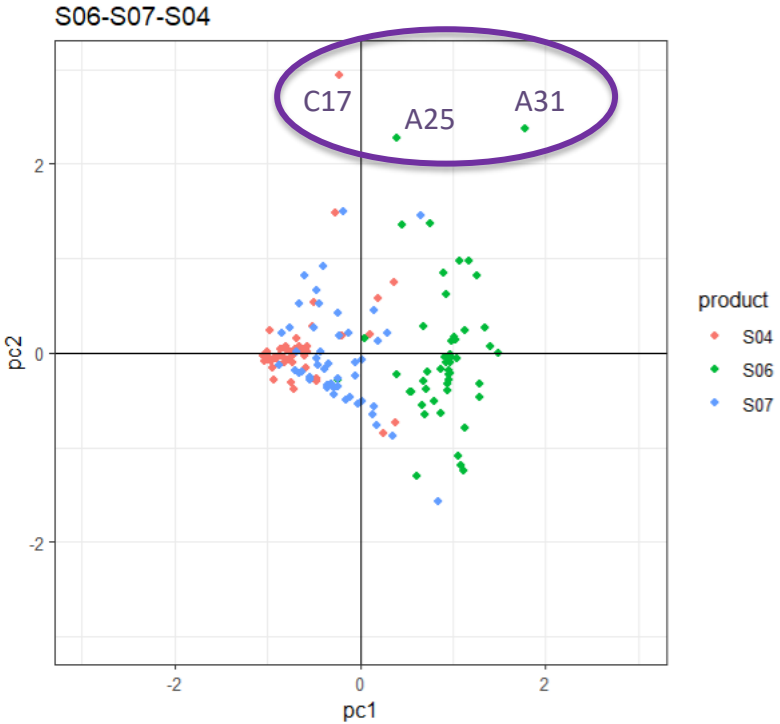


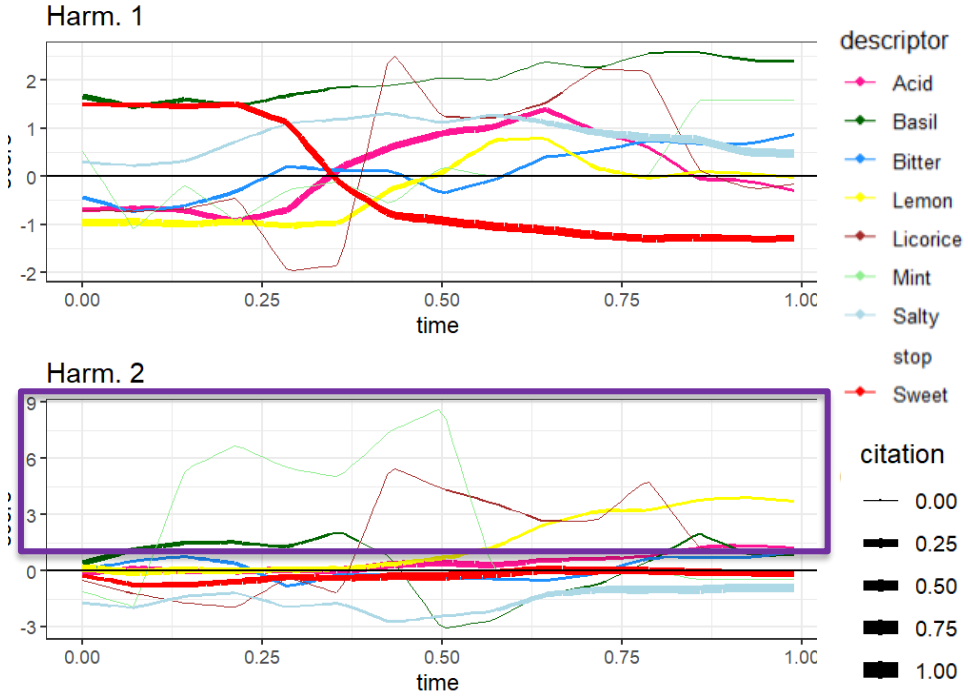


CFDA results

- Positive scores on the second component: Mint then Licorice and finally Lemon
  = DISTRACTORS: outlier detection





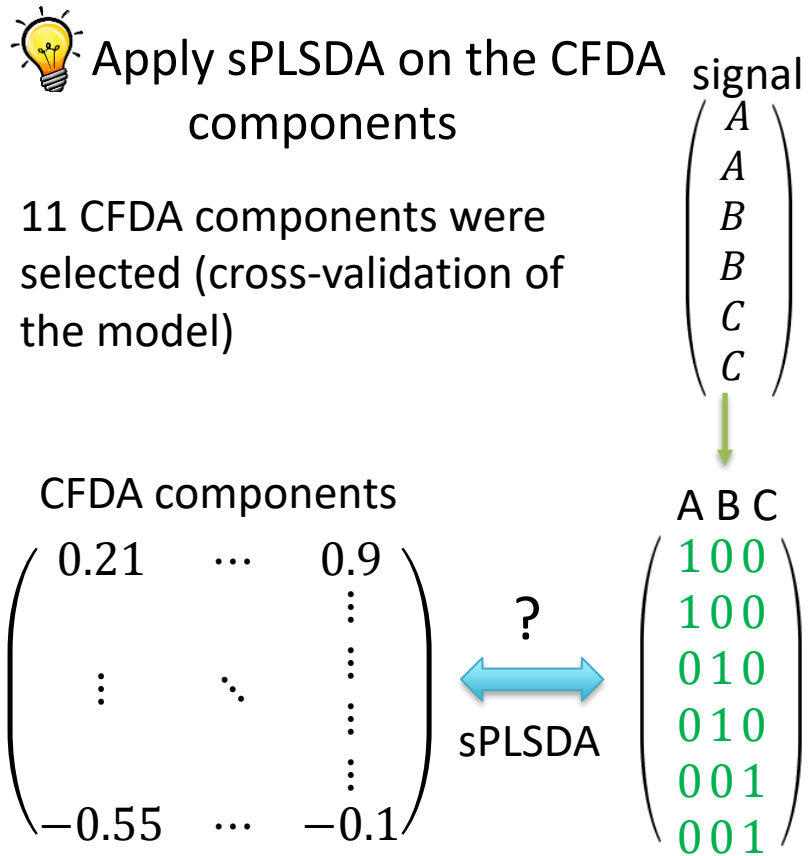


How to characterize the three signals (if they were unknown)

Objective: attempt to explain the differences between the signals

Apply sPLSDA on the CFDA components

11 CFDA components were selected (cross-validation of the model)



Partial Least Square regression (PLS):
a multivariate statistical method applied to two matrices X_1(n, p_1), and X_2 (n, p_2), that finds the weights $\boldsymbol{w_1, w_2}$ **maximizing**

$$\boldsymbol{cov(X_1 w_1, X_2 w_2)}$$

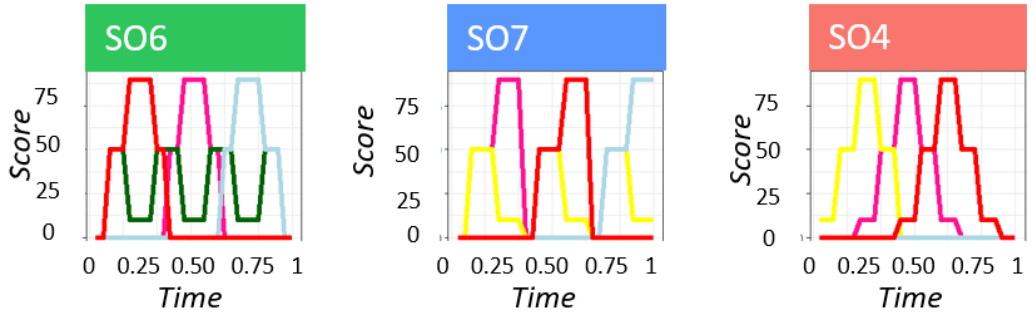
under constraints of unit variance of the weights $||w_1||_2 = ||w_1||_2 = 1$

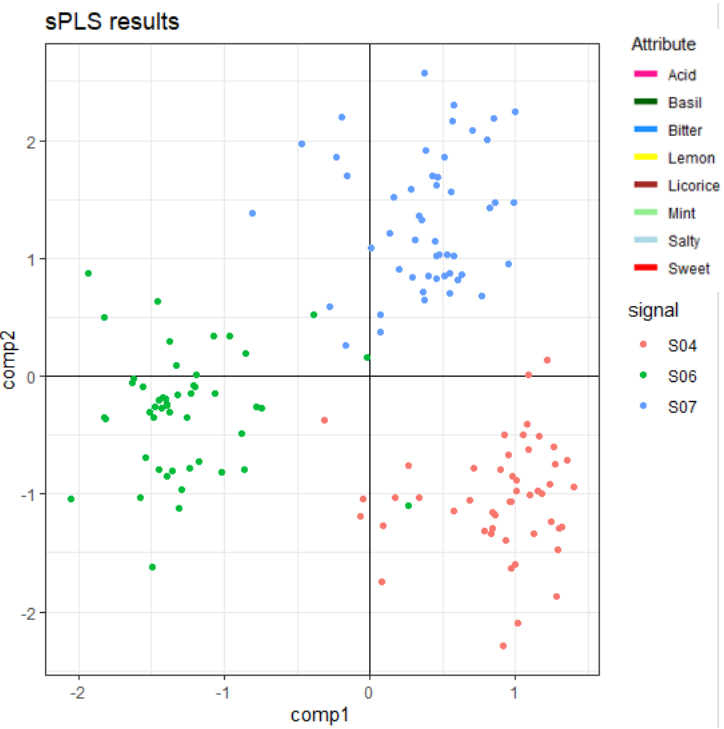
Sparse Partial Least Square (sPLS):
Adding constraints of sparsity: $||w_i||_1 < \sigma_i$ with $0 < \sigma_i < 1$ for selecting variables (more variables than individuals)

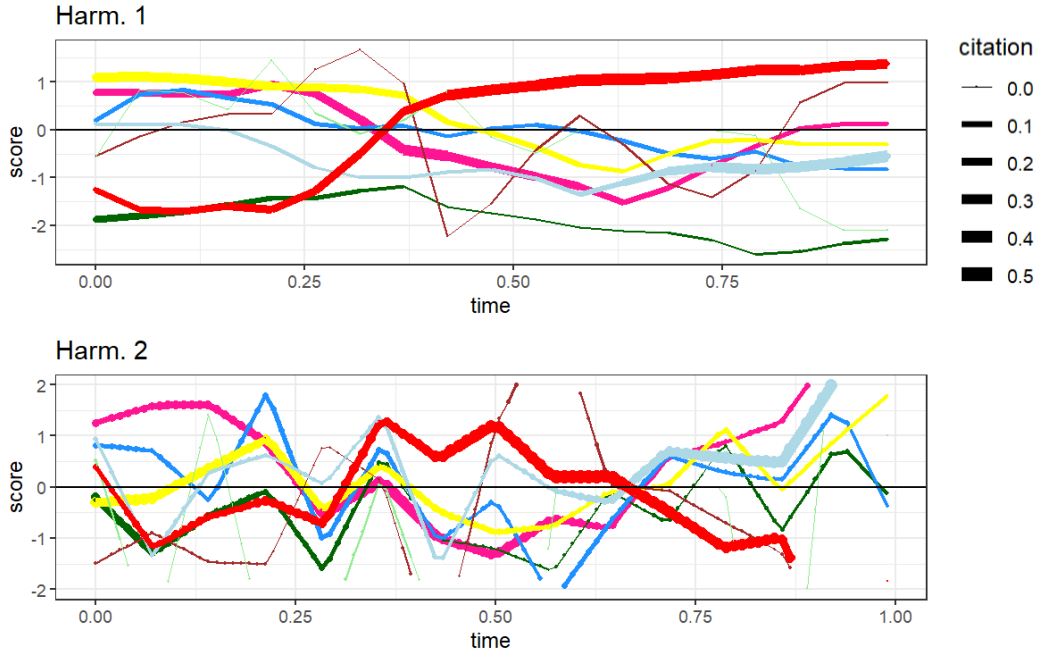
sPLS - Discriminant Analysis (sPLS-DA): consists in choosing X_2 as a dummy matrix corresponding to a qualitative variable

sPLS-DA results

The three signals are well discriminated.
First axis similar to the CFDA's one.

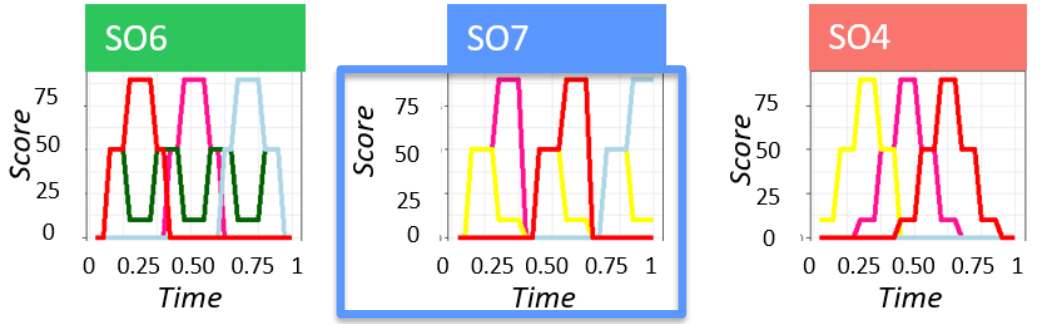


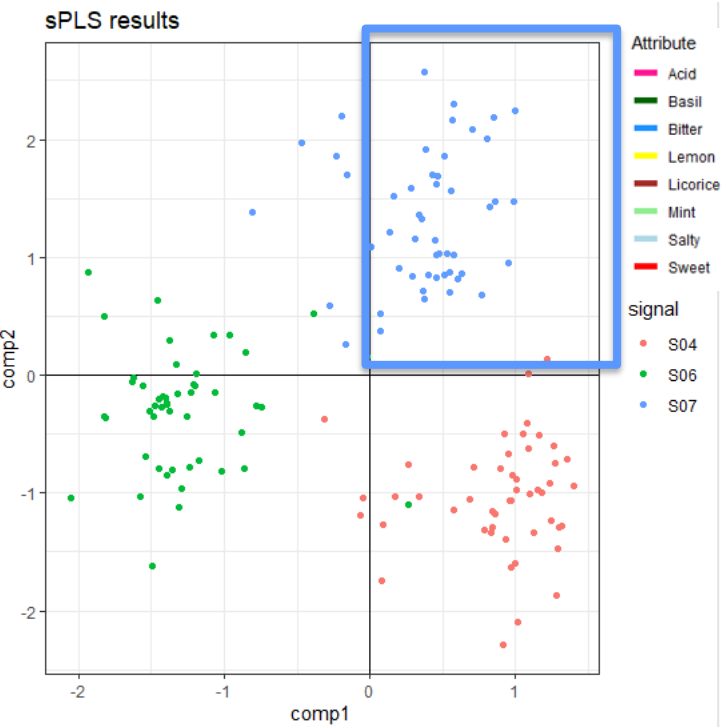


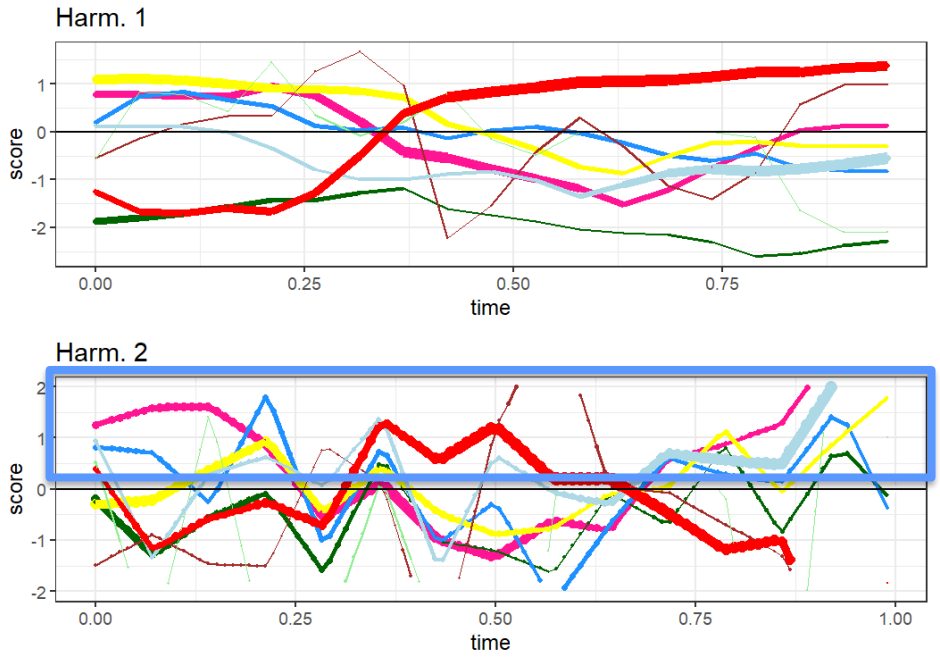


sPLS-DA results

The second axis discriminated S07 and S04. S07 is Acid then Sweet in the middle of the tasting and Salty at the end

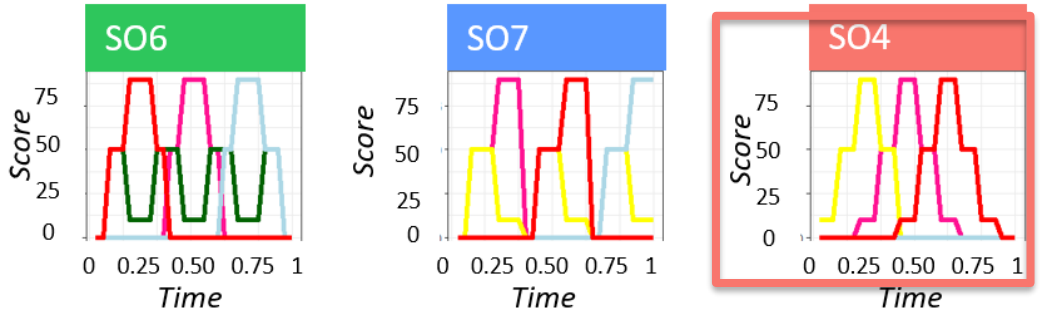


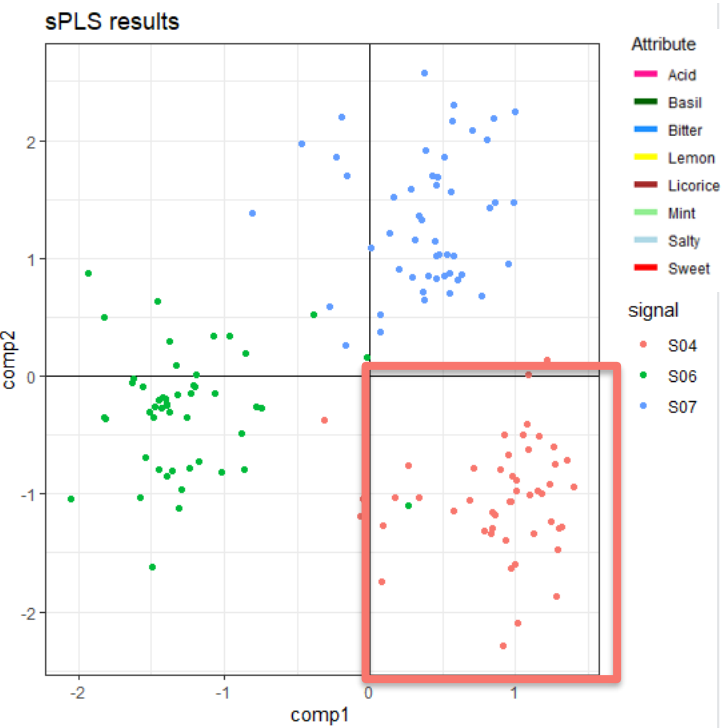


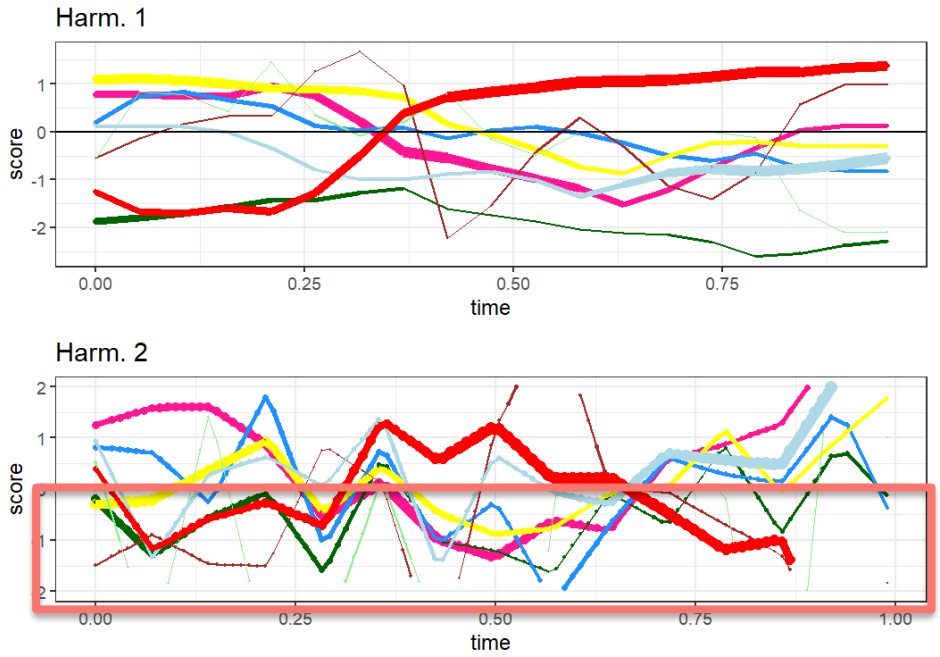


sPLS-DA results

The second axis discriminated S07 and S04. S04 is Sweet at the end of the tasting.







Synthesis

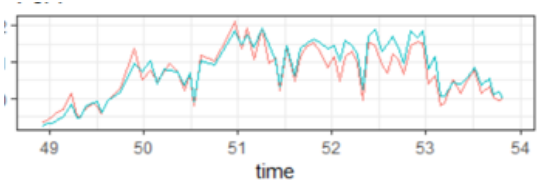
CFDA works well for these TDS data and returns results that are coherent with the original signals.

<u>Some important technical parameters:</u>
Selection of the basis of functions (number, order, … ?)
Parameters of sPLSDA to be discussed

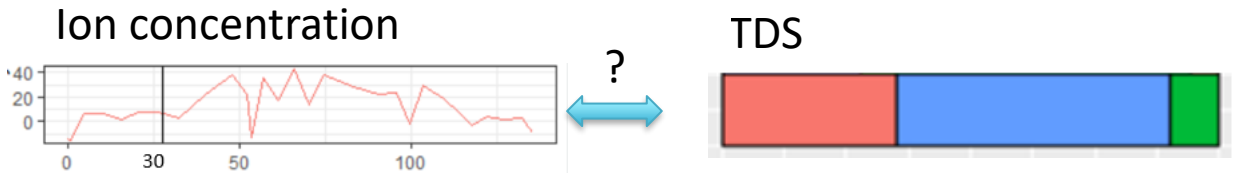
Use of FDA techniques to other temporal sensory analysis protocols

- Multivariate CFDA with Temporal Check All That Apply

- Multivariate FDA for temporal intensity values



- Conjoint analysis of chemico-physical data (quantitative, analyzed with FDA) and sensory TDS data (qualitative, analyzed with CFDA)



Thanks!

Thanks for your attention!

ELSEVIER

Food Quality and Preference

Volume 109, July 2023, 104893

Analyzing temporal dominance of sensations data with categorical functional data analysis

Caroline Peltier [a b], Michel Visalli [a b], Pascal Schlich [a b], Hervé Cardot [c]

caroline.peltier@inrae.fr

ChemoSens