Exploring the relationship between sensory and instrumental data with component-based methods

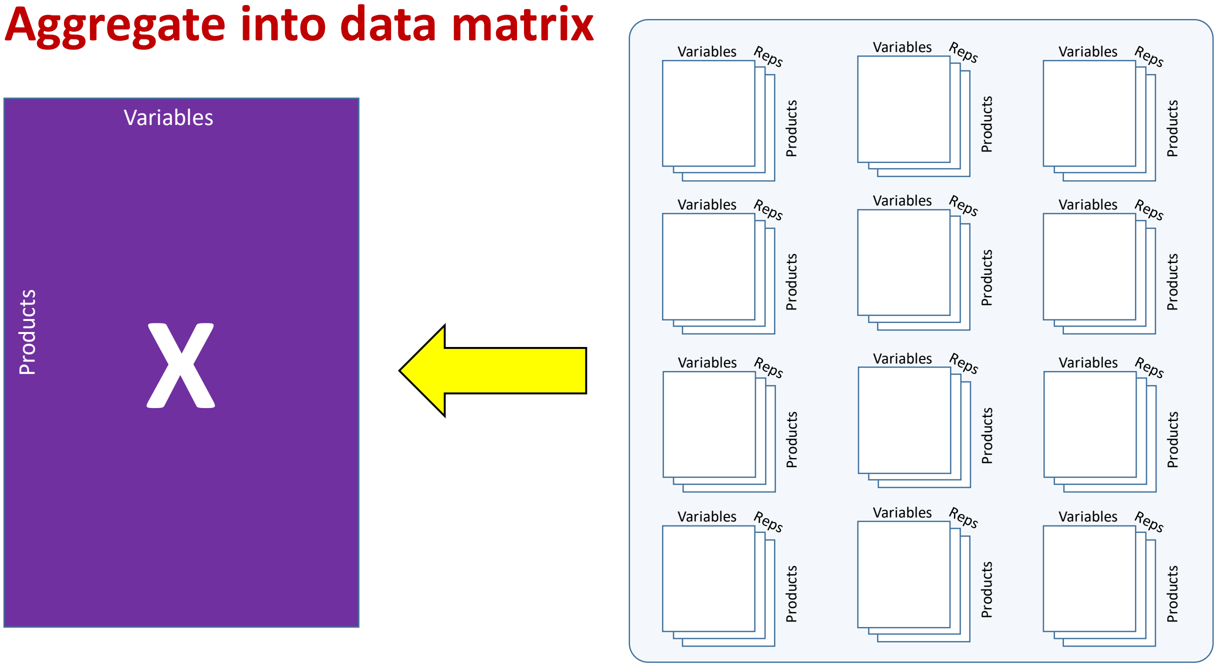
KoSFoST International Symposium and Annual Meeting 2025
Pioneering Future Connection in FoodTech
Gwangju, Korea · 2-4 July 2025

John Castura
Dr. Philos., M.Sc.
Research Fellow
Compusense

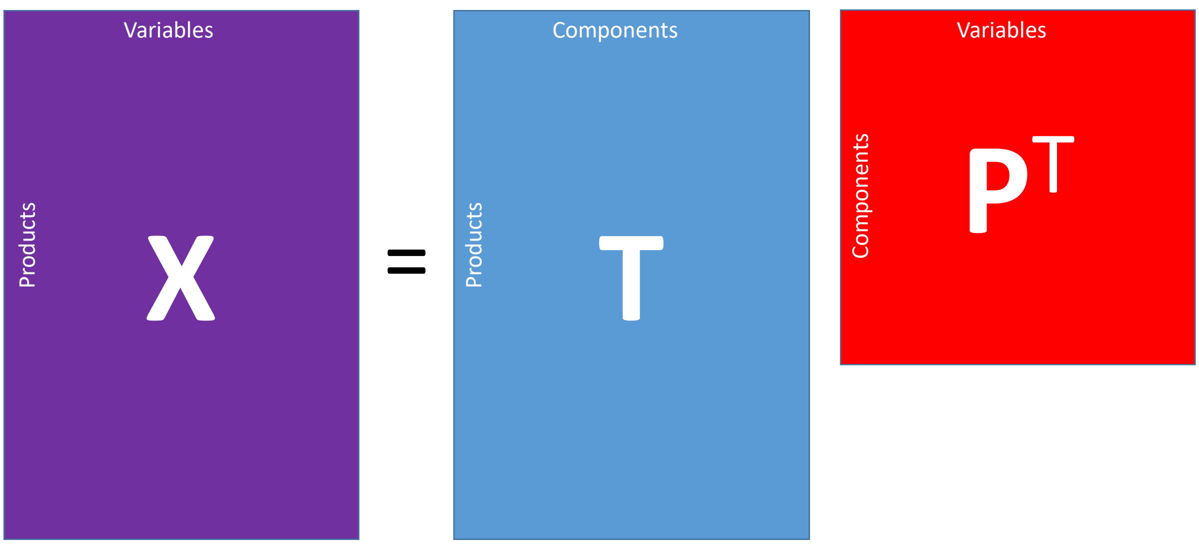
J.C. Castura, Compusense Inc.

KoSFoST 2025, 2-4 July, Gwangju, Korea

Aggregate into data matrix

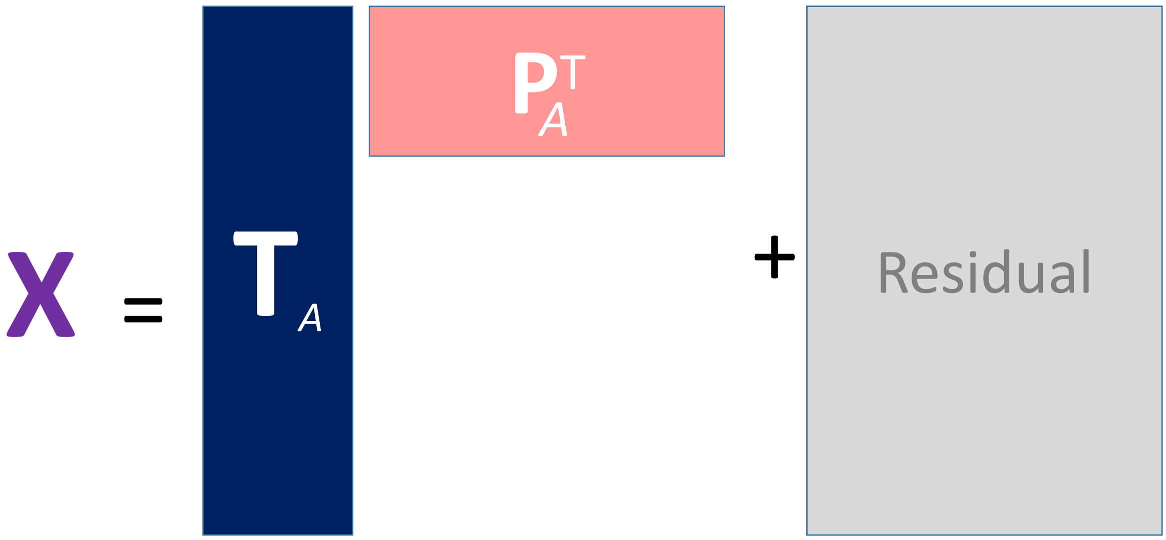


Principal component analysis

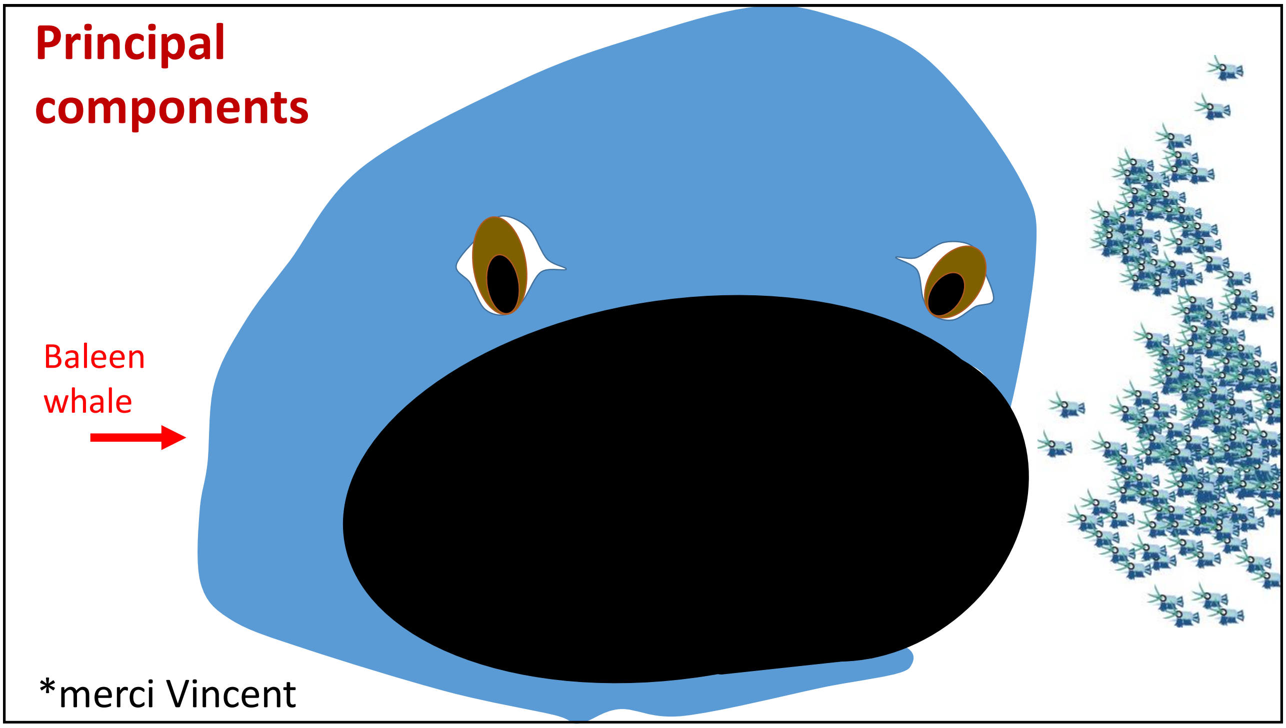


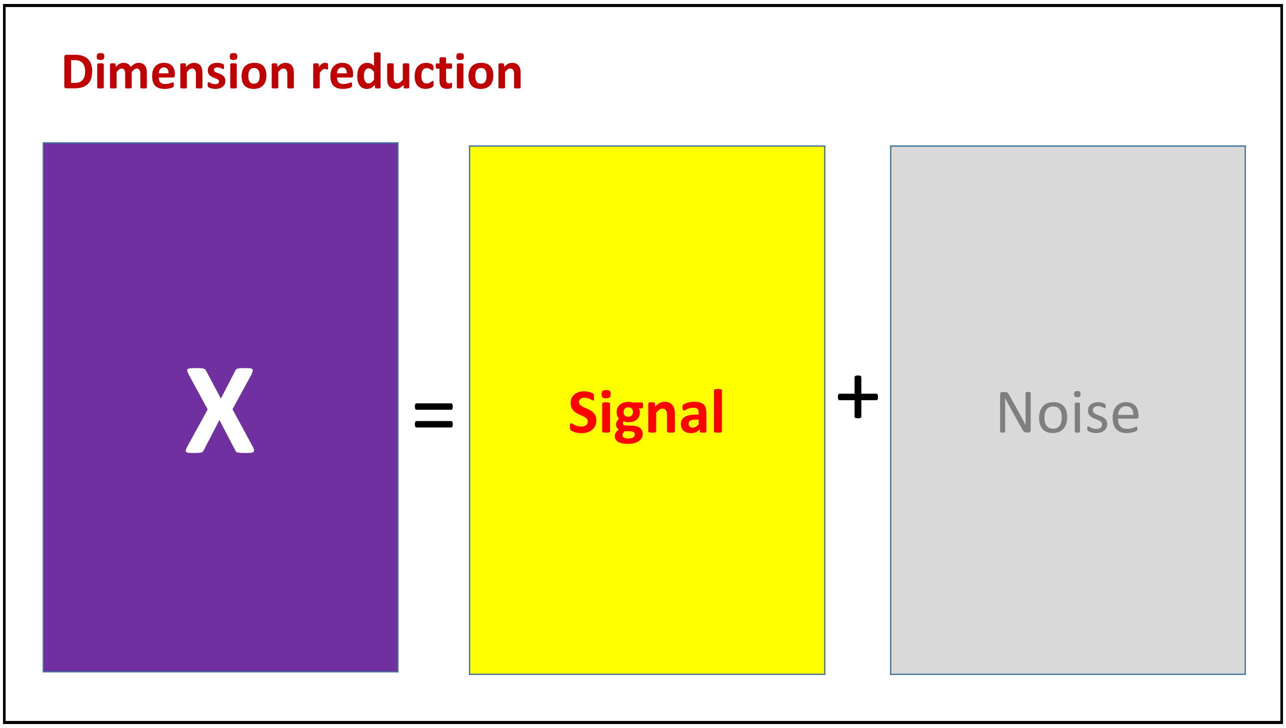
J.C. Castura, Compusense Inc.

Dimension reduction to *A* components

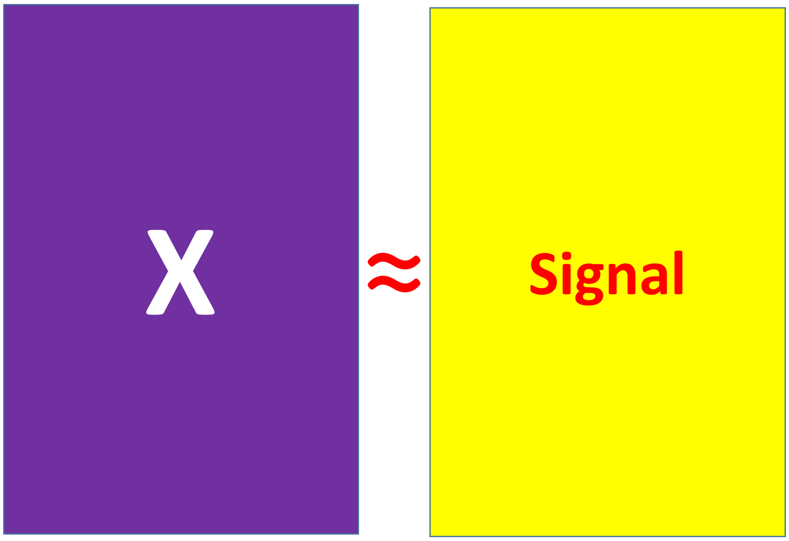


J.C. Castura, Compusense Inc.

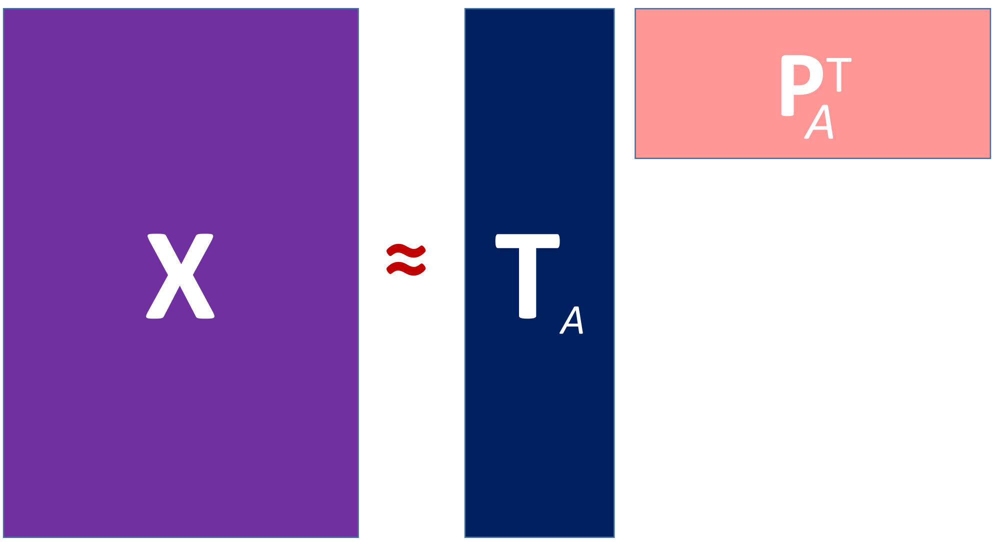

Principal
components
Baleen
whale
*merci Vincent


Dimension reduction
X
=
Signal
+
Noise

Dimension reduction

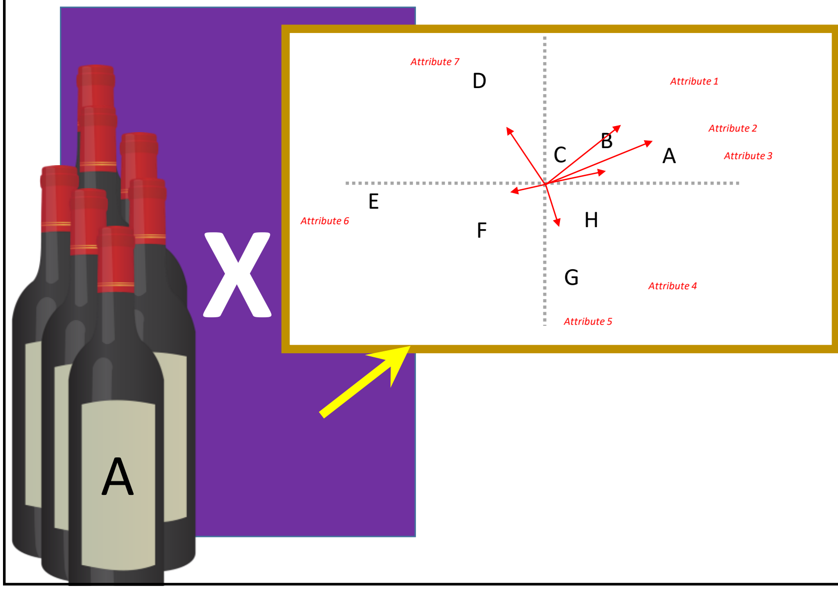


Dimension reduction to *A* components

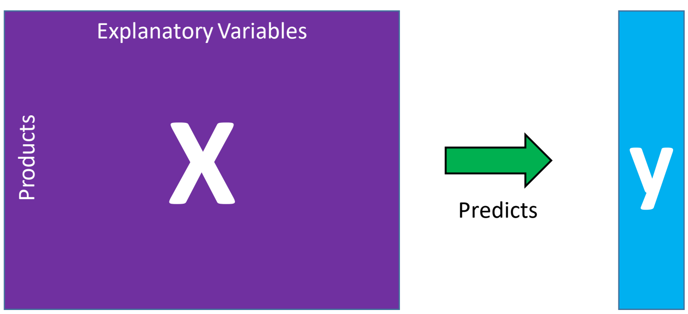


J.C. Castura, Compusense Inc.

Visualizing PCA results



Investigating data relationships



Chemistry

Sensory

Why component-based methods?

- Summarize and visualize complicated data
- Often relatively few underlying sources of variation
- Useful approximation of data
- Separation of signal/noise
- Outlier detection
- Confirm hypotheses
- Generate hypotheses

Næs, Varela, Castura, Bro & Tomic (2023). Why use component-based methods in sensory science? *Food Quality and Preference*, 112, 105028. https://doi.org/10.1016/j.foodqual.2023.105028

Data for
French Pinot Noir Wines

Volatile organic compounds (VOCs)

Headspace measurements of VOCs obtained from
headspace—solid phase micro-extraction—gas chromatography—mass spectrometry (HS-SPME-GC-MS)

X

1-hexanol	acetaldehyde	ethyl acetate	isoamyl acetate
1-octanol	acetic acid	ethyl butyrate	isoamyl propionate
1-phenoxy-2-propanol	alpha-ionone	ethyl caproate	isovaleric acid
2,3-butanedione	beta-ionone	ethyl isobutyrate	methional
2-ethylhexan-1-ol	butyl acetate	ethyl isovalerate	methionol
2-isobutyl-3-methoxypyrazine	butyric acid	ethyl lactate	pentyl propionate
2-methyl-1-butanol	damascenone	ethyl octanoate	phenol
2-methylbutyl acetate	dimethyl sulfide	ethyl propionate	phenylacetaldehyde
2-phenylethanol	ethyl 2-methylbutyrate	furaneol	phenylacetic acid
3-methyl-1-butanol	ethyl 3-hydroxybutyrate	hexyl acetate	propionic acid
4-ethyl-2-methoxyphenol	ethyl 6-hydroxyhexanoate	homofuraneol	trans-3-hexen-1-ol
4-ethylphenol			

Villière et al. (2019)

Pinot noir wines

T1 2010 Bourgogne PDO

T2 2009 Bourgogne PDO

T3 2009 Bourgogne PDO

T4 2009 Bourgogne Hautes Côtes de Beaune PDO

T5 2009 Savigny-lès-Beaune PDO

T6 2010 Maranges PDO

C **2009 Côte de Nuits-villages PDO**

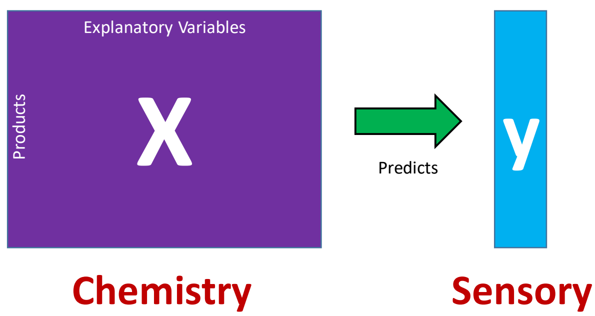
T7 2009 Ladoix PDO

Villière et al., 2019, https://doi.org/10.1016/j.dib.2019.103725

Sensory Variables

y

Artichoke	Cherry fresh	Hay	Smoky
Bell pepper	Cherry stone	Leather	Strawberry cooked
Blackberry fresh	Clove	Musk	Strawberry fresh
Blackcurrant bud	Cut grass	Pepper	Toasty
Blackcurrant fresh	Elderflower	Plum cooked	Undergrowth
Blueberry fresh	Ethanol	Plum fresh	Vanilla
Brioche	Firestone	Prune	Violet
Butter	Geranium	Raspberry fresh	Woody
Cherry cooked			



Principal component regression

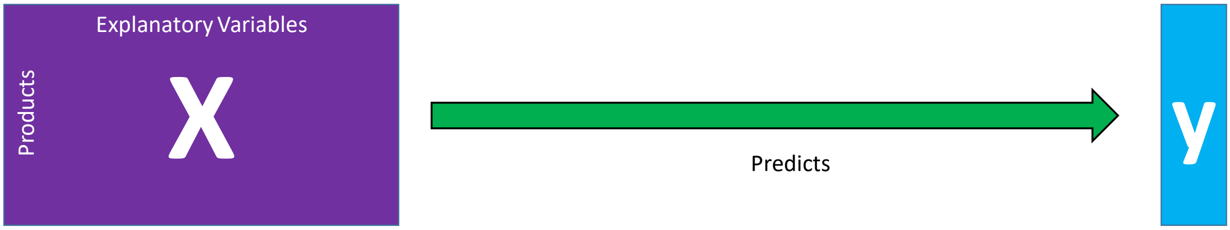
Unsupervised & Supervised

We want to ***predict*** the response **y** from multivariate **X**.

Response **y** is regressed on principal components of **X**.

Investigating data relationships

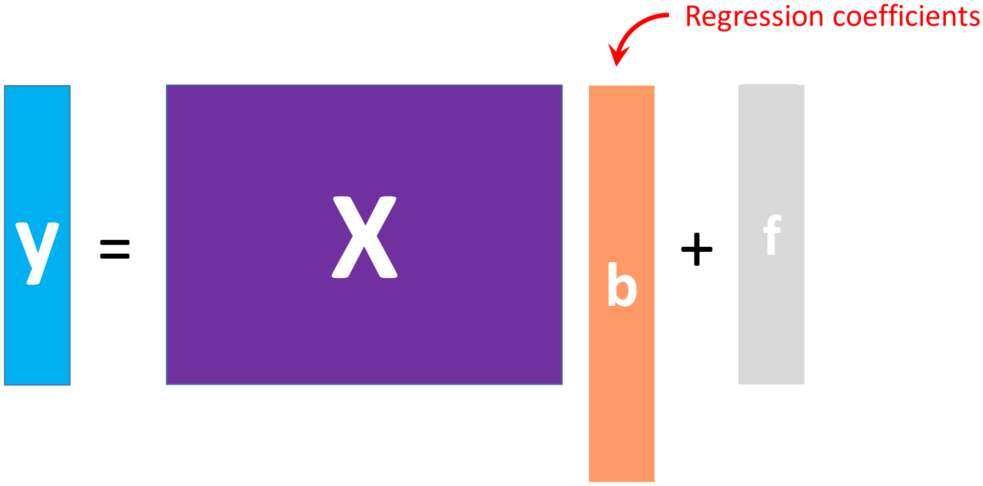
with **multiple linear regression**



J.C. Castura, Compusense Inc.

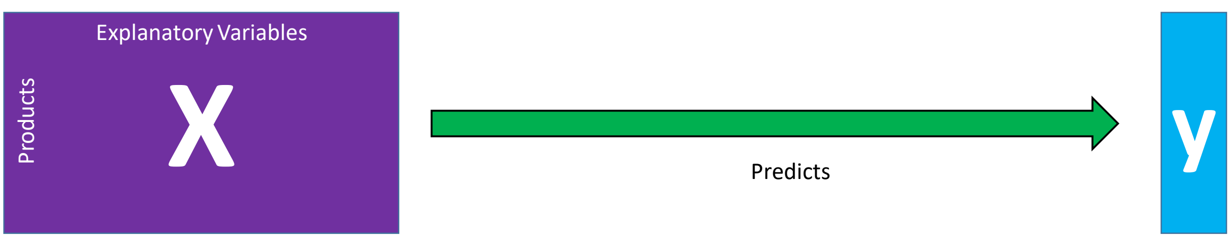
Investigating data relationships

with **multiple linear regression**



Investigating data relationships

with **multiple linear regression**



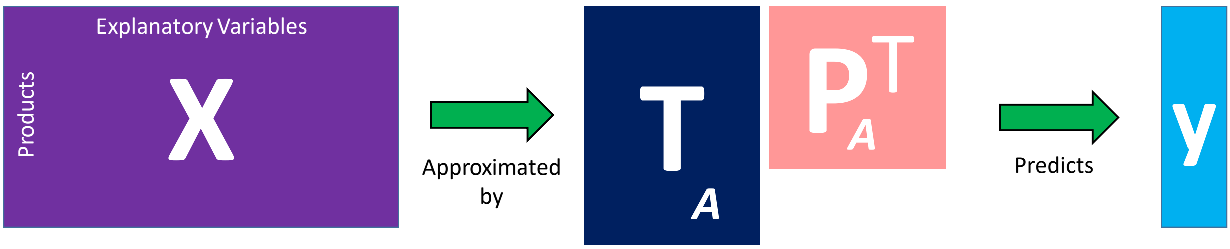
Multicollinearity?
More variables than objects?

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{f}$$

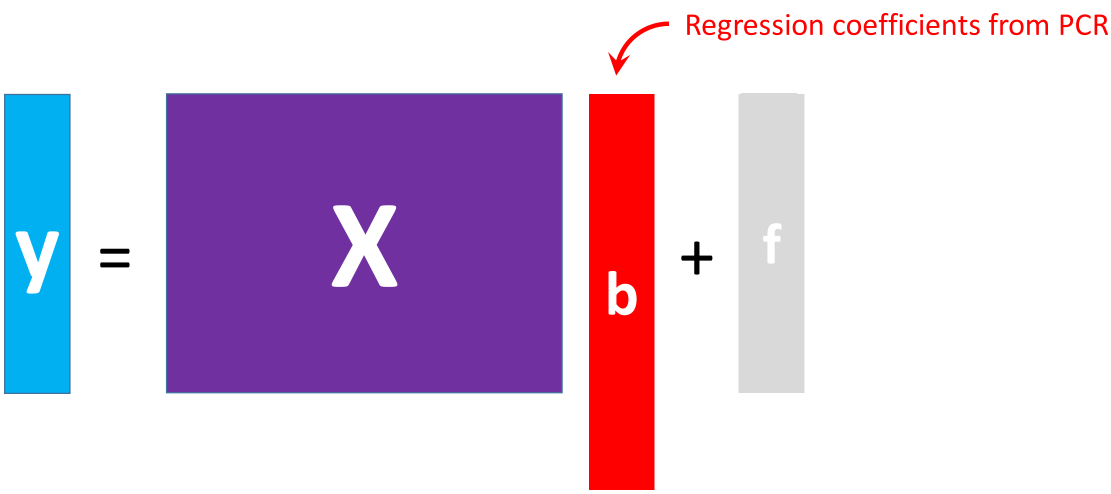
$$\mathbf{b} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

J.C. Castura, Compusense Inc.

Investigating data relationships
with principal component regression (PCR)



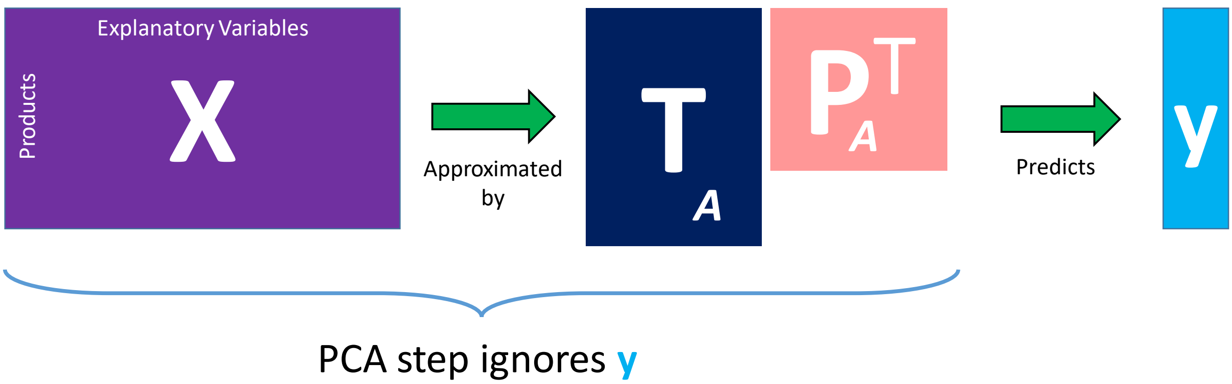
Investigating data relationships
with principal component regression (PCR)



J.C. Castura, Compusense Inc.

Investigating data relationships

with **principal component regression (PCR)**

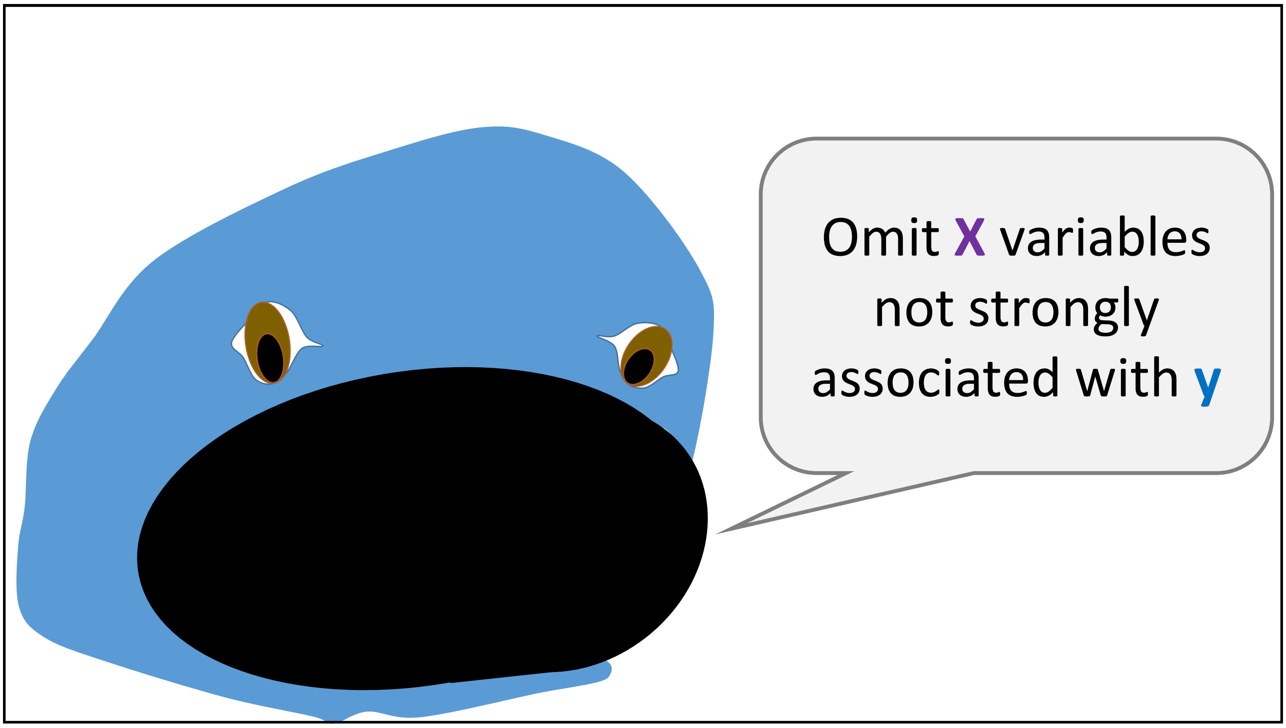


Supervised principal component regression (SPCR)

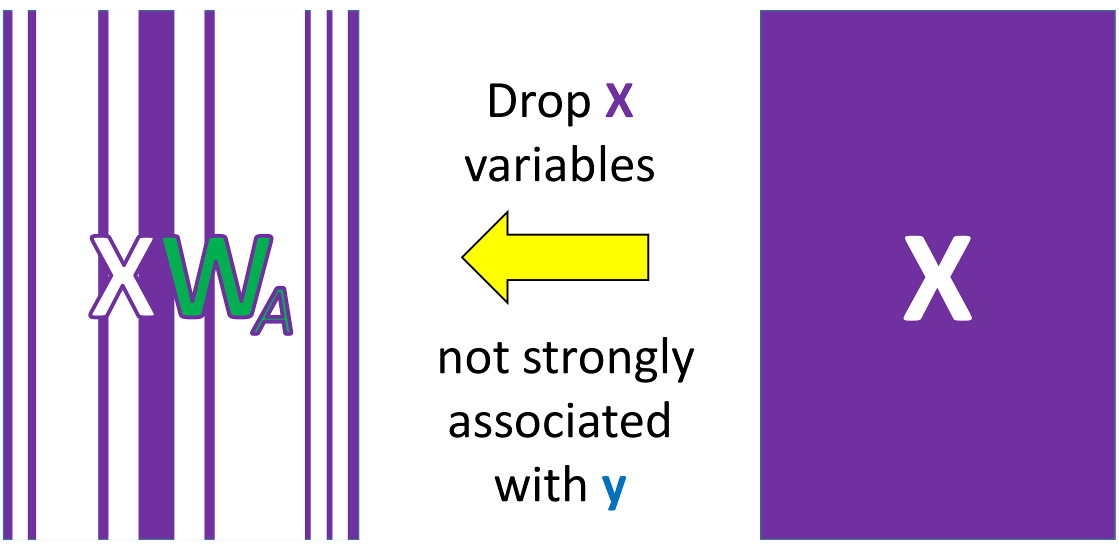
We want to ***predict*** the response y from the multivariate X.

Response y is regressed on principal components of X.

J.C. Castura, Compusense Inc.

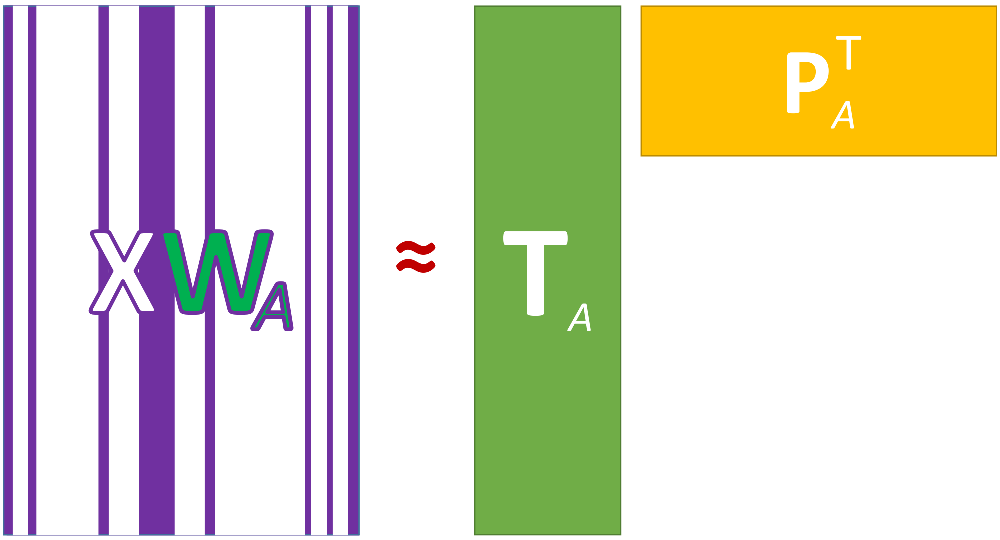


Supervised principal component regression (SPCR)



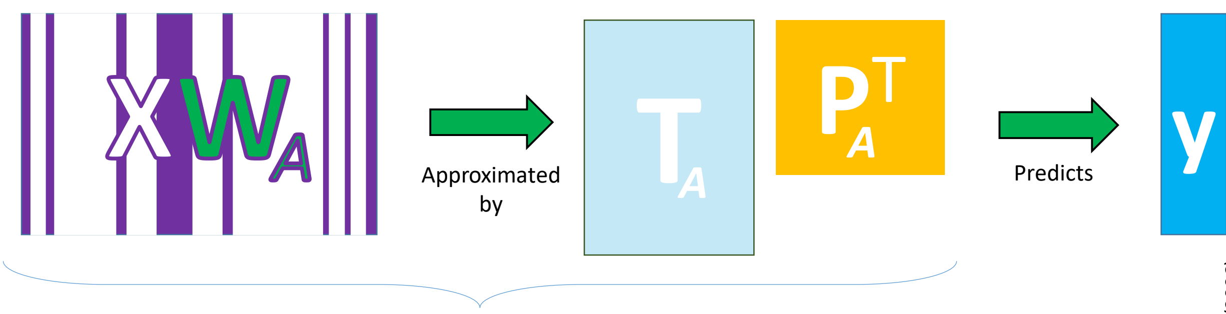
Bair, Hastie & Tibshirani (2006). Prediction by supervised principal components. JASA. oi:10.1198/016214505000000628

Dimension reduction in SPCR



Bair, Hastie & Tibshirani (2006)

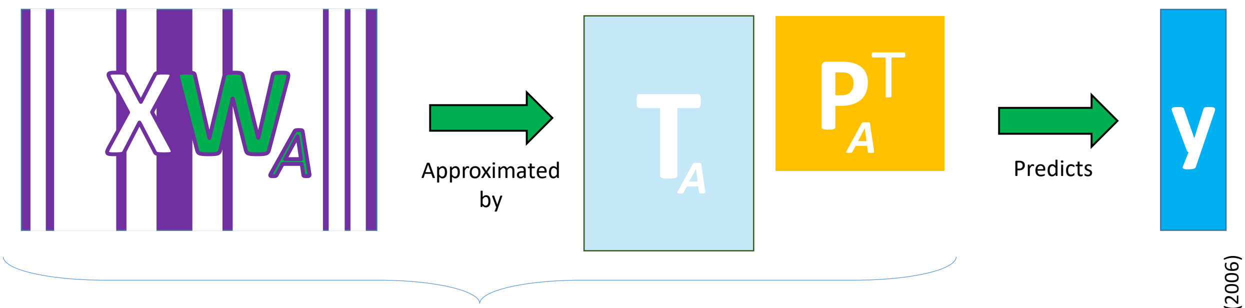
Supervised principal component regression (SPCR)



PCA of filtered X variables yields
Supervised Principal Components

Bair, Hastie & Tibshirani (2006)

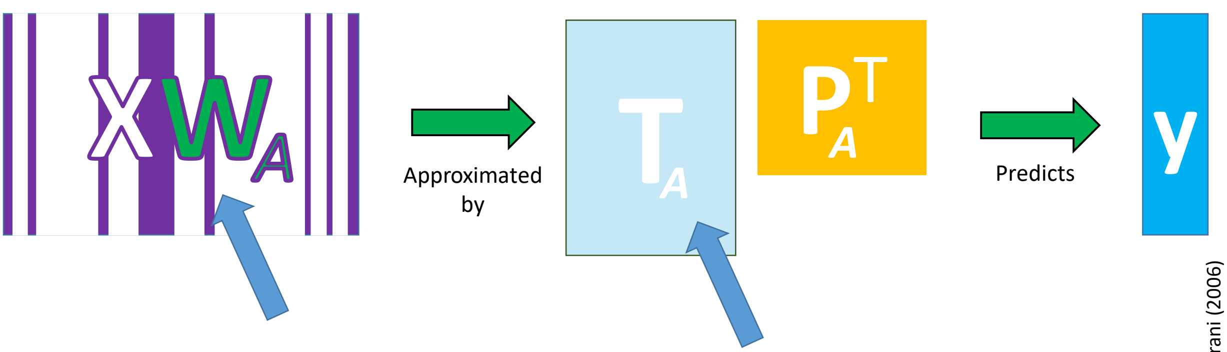
Supervised principal component regression (SPCR)



PCA of filtered X variables yields
Supervised Principal Components

Bair, Hastie & Tibshirani (2006)

Supervised principal component regression (SPCR)



Slope parameter: θ
Weight is 1 for **X** variables associated strongly enough with **y** and 0 otherwise.

Components parameter: ***A***
Determines number of components.

Bair, Hastie & Tibshirani (2006)

Many candidate values for parameters θ and ***A***

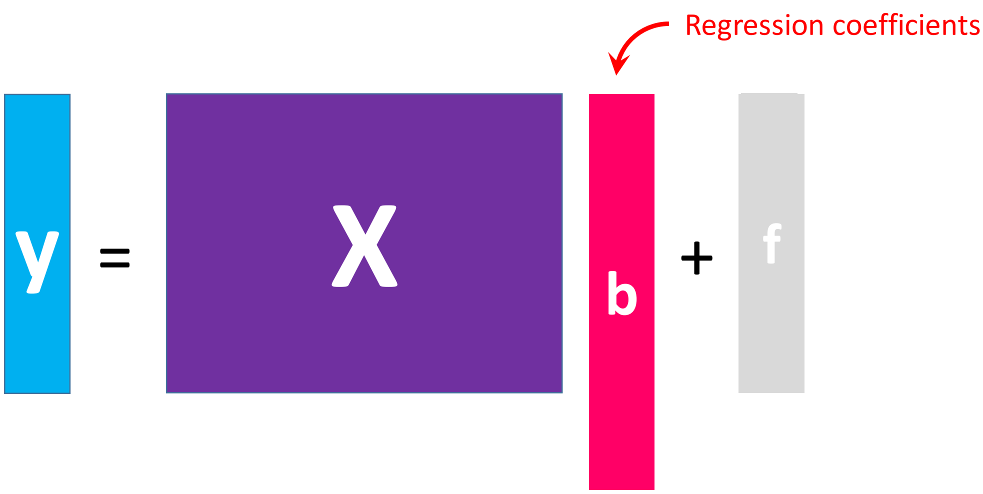
How to choose?

Strategy: Choose model that gives the most accuracy predictions for data withheld from model training

Bair et al.: 2-fold cross-validation (2-fold cv)

Us: leave-one-out cross-validation (loocv)

Supervised principal component regression (SPCR)



Bair, Hastie & Tibshirani (2006)

J.C. Castura, Compusense Inc.

Usually, when interpret results, we draw conclusions about…

relationships between variables,

relationships between objects, and

differences between objects.

Usually, when interpret results, we draw conclusions about…

relationships between variables,

relationships between objects, and

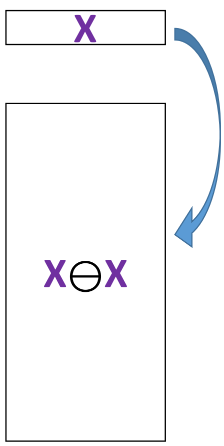
differences between objects.

Investigating paired comparisons after principal component analysis

Castura, J.C., Varela, P, & Næs, T. (2023).
Investigating paired comparisons after principal component analysis. *Food Quality and Preference*, 106, 104814.
https://doi.org/10.1016/j.foodqual.2023.104814

Investigating paired comparisons after PCA

"Crossdiff-unfolding"



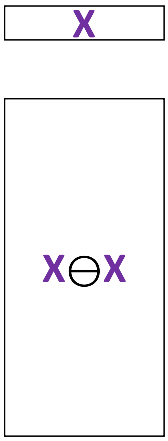
X is a column-centered ($J \times M$) matrix

Every row is subtracted from every row

X⊖X is a column-centered ($J^2 \times M$) matrix

Investigating paired comparisons after PCA

“Crossdiff-unfolding”

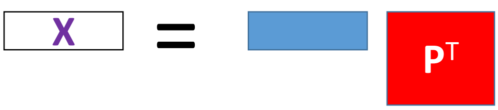


The covariance matrix of **X** and the covariance matrix of **X⊖X** are identical except for a multiplier.

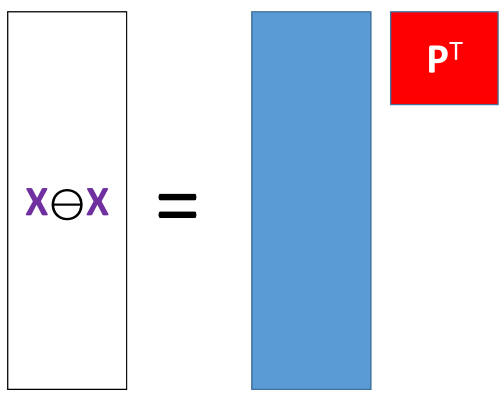
Next, consider PCA of **X** and PCA of **X⊖X**.

Investigating paired comparisons after PCA

PCA of X



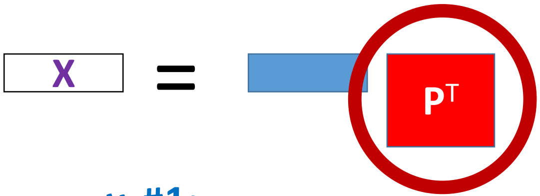
PCA of X⊖X



J.C. Castura, Compusense Inc.

Investigating paired comparisons after PCA

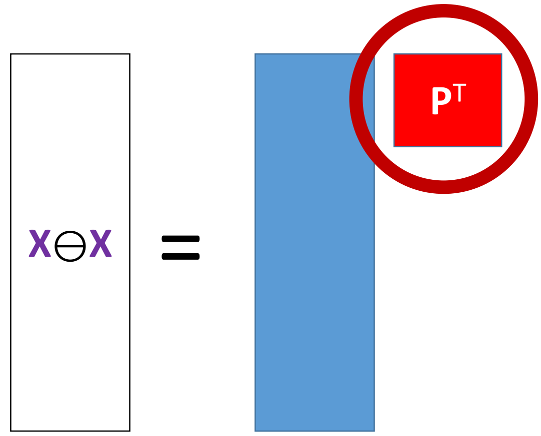
PCA of X



Key result #1:

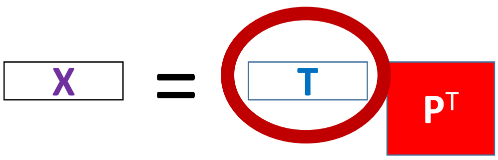
Loading matrices obtained from these two PCA solutions are identical.

PCA of $X \ominus X$



Investigating paired comparisons after PCA

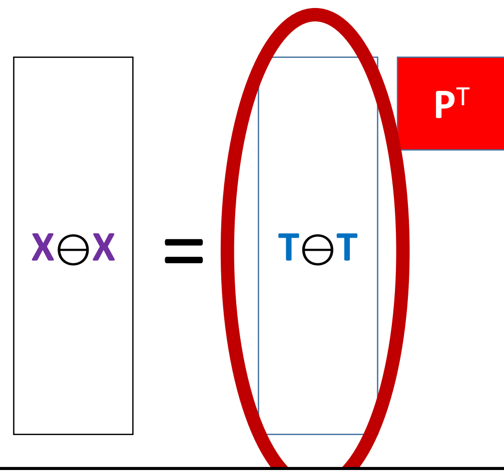
PCA of X



Key result #2:

If we crossdiff-unfold scores from PCA of X, we get scores from PCA of $X \ominus X$.

PCA of $X \ominus X$



J.C. Castura, Compusense Inc.

Paired comparisons

This demonstrates that objects and all their paired comparisons are optimally investigated **in the same principal components**.

Paired comparisons

Therefore, we can just do PCA of **X** and know the PCA of **X**⊖**X** without actually doing this PCA.

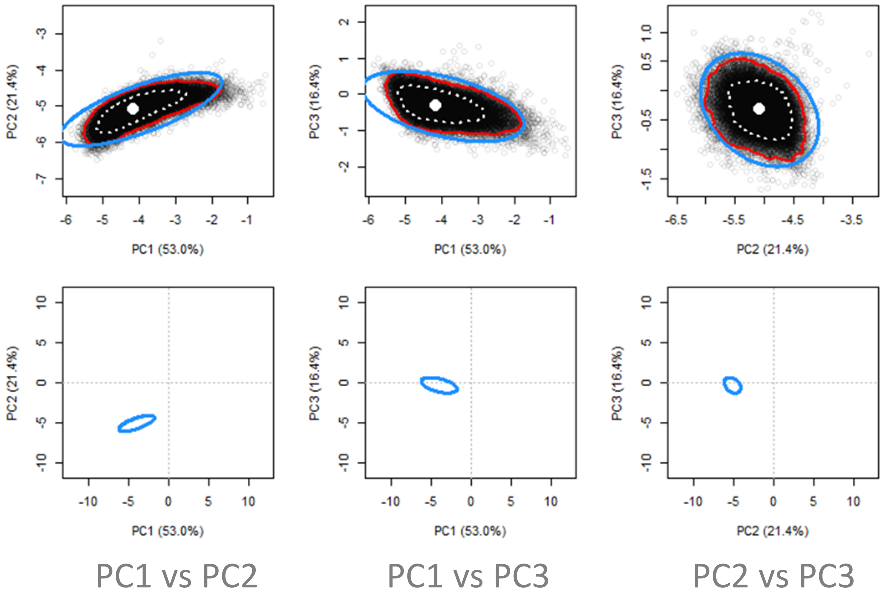
This lays necessary theoretical groundwork to justify a strategy for doing paired comparisons after PCA.

We need to **account for mutual dependencies in the data** when investigating paired comparisons.

Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis

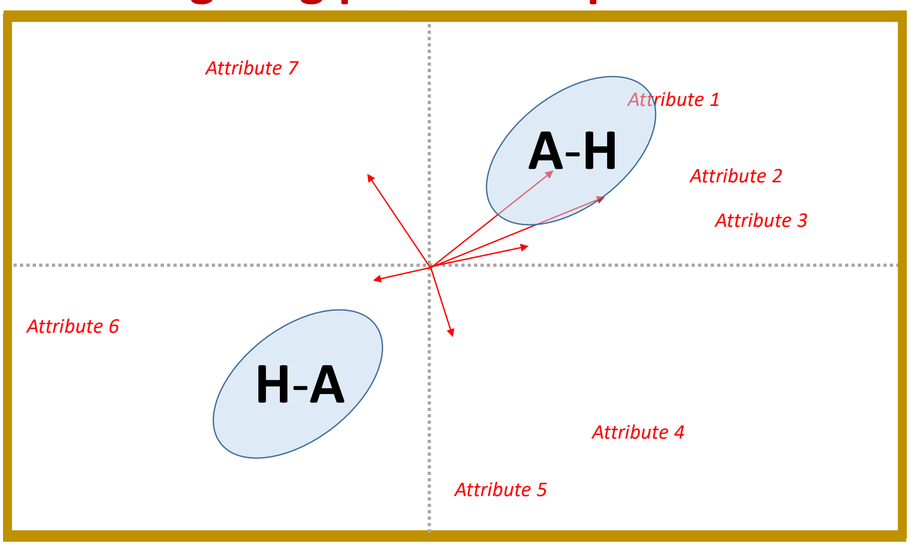
Castura, J.C., Varela, P, & Næs, T. (2023).
Evaluation of complementary numerical and visual approaches for investigating pairwise comparisons after principal component analysis. *Food Quality and Preference*, 107, 104843. https://doi.org/10.1016/j.foodqual.2023.104843

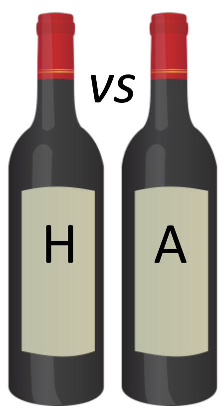
…complementary numerical and visual approaches…



Suppl. Fig. 1a from Castura, Varela & Næs (2023a)

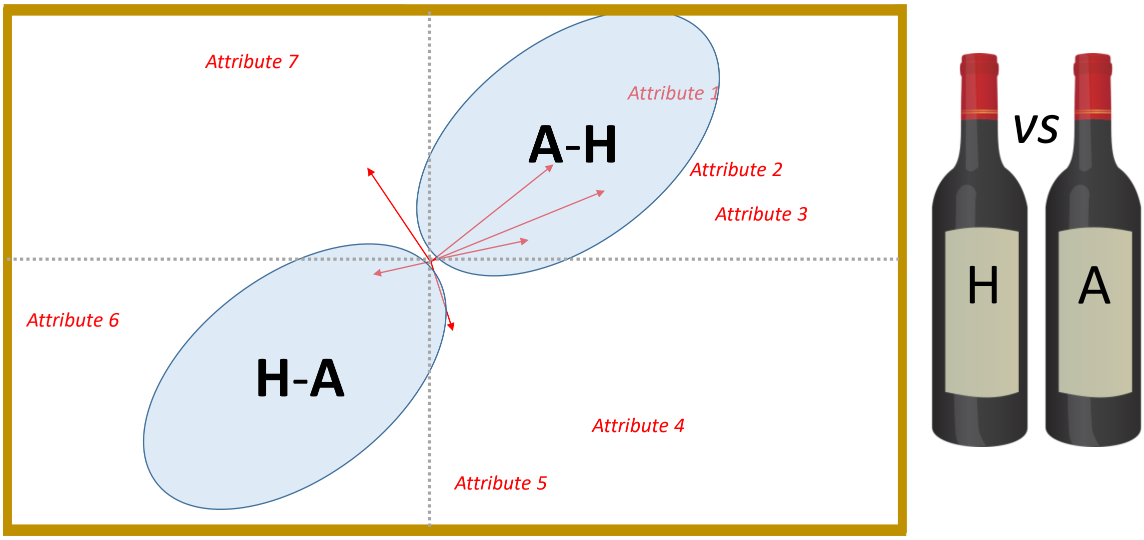
Investigating paired comparisons



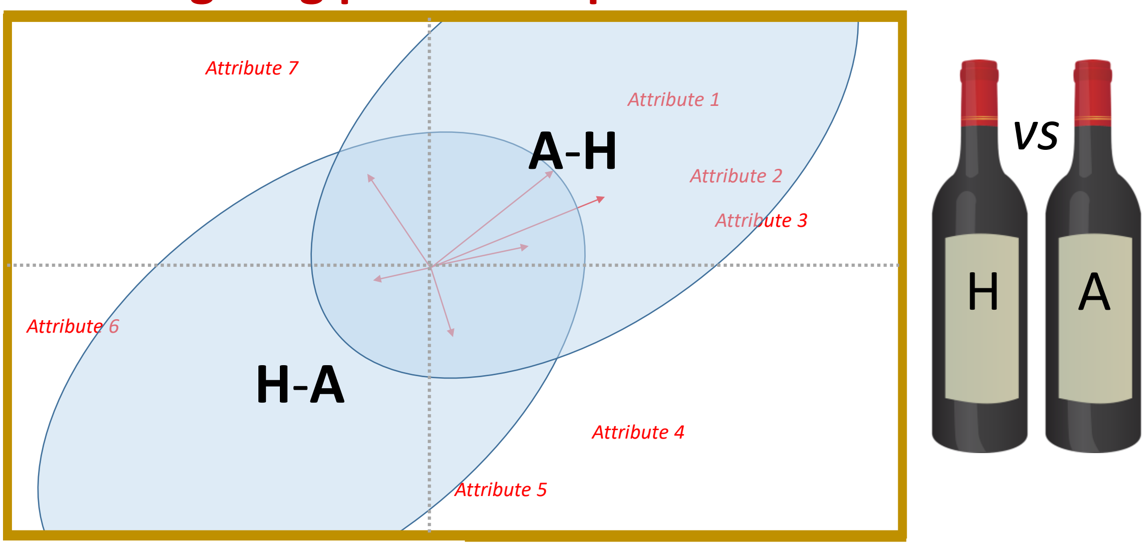


J.C. Castura, Compusense Inc.

Investigating paired comparisons



Investigating paired comparisons



In some studies, only a subset of paired comparisons is of primary interest…

Investigating only a subset of paired comparisons after principal component analysis

Castura, J.C., Varela, P., & Næs, T. (2023).
Investigating only a subset of paired comparisons after principal component analysis. *Food Quality and Preference*, 110, 104941.
https://doi.org/10.1016/j.foodqual.2023.104941

Investigating a subset of paired comparisons after PCA

Example where only a subset of paired comparisons are of primary interest:

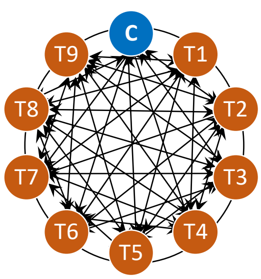
Many Test Products *vs* One Control

Focus on Test-Control pairs *...not Test-Test pairs*

1 Control vs 9 Test Products

X⊖X

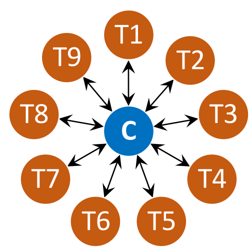
All Pairs



45 paired comparisons
90 paired differences

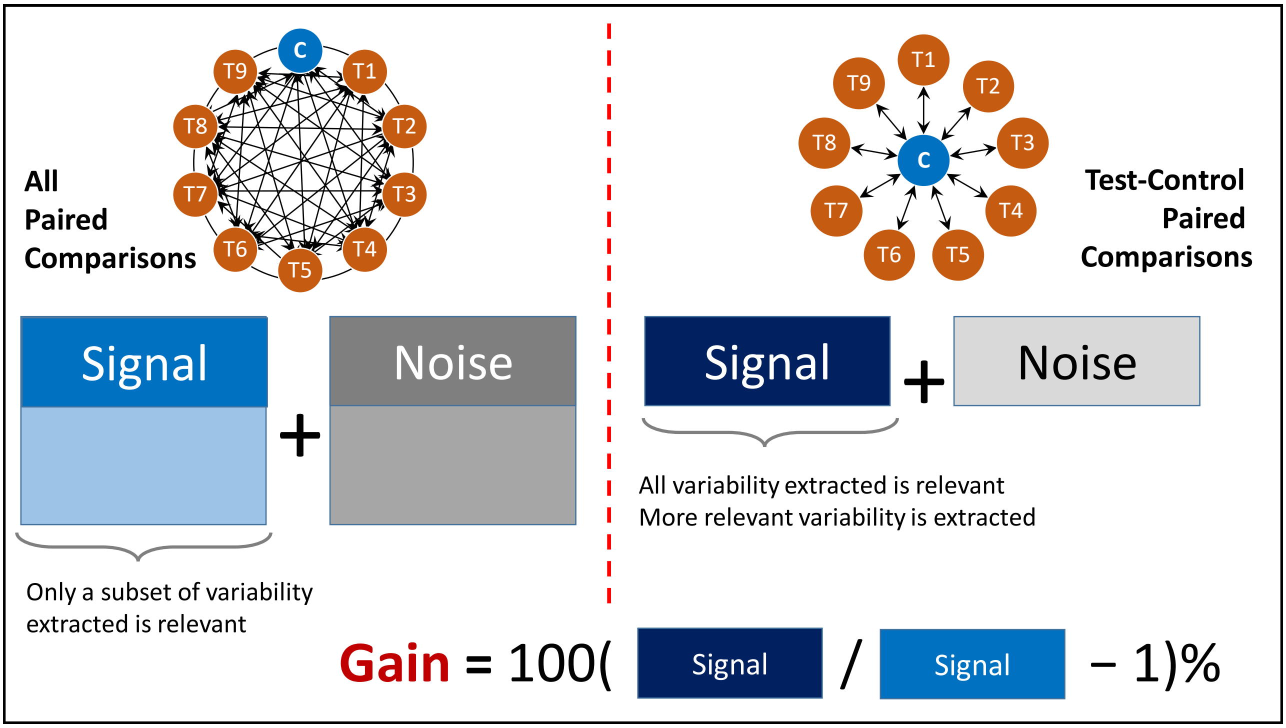
Δ*

Test-Control Pairs



9 paired comparisons
18 paired differences

Castura, Cariou & Næs (2024)

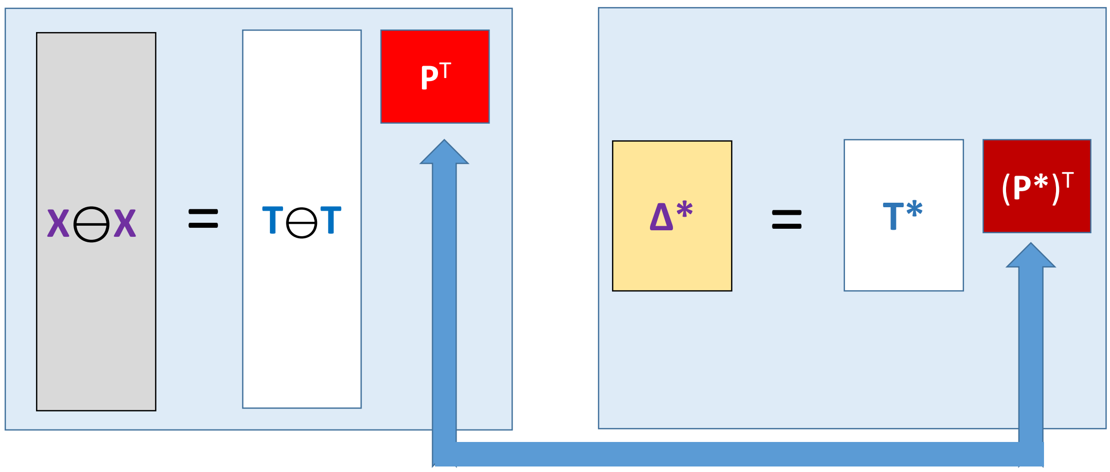


Investigating only a subset of paired comparisons

"…the interrelationships between the variables might be different for the subset of paired comparisons than it is for all paired comparisons. So the covariance matrix for a matrix of all paired comparisons and the covariance matrix of selected paired comparisons will differ depending on the data. "

Castura, J.C., Varela, P., & Næs, T. (2023). Investigating only a subset of paired comparisons after principal component analysis. *Food Quality and Preference*, 110, 104941.

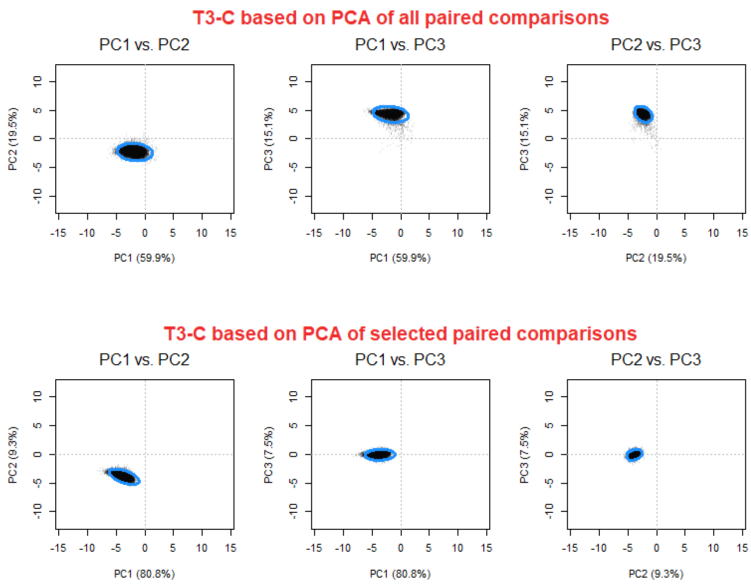
Investigating a subset of paired comparisons after PCA



Δ* contains a subset of the rows in **X⊖X**

Loading matrices differ

Investigating a subset of paired comparisons after PCA



Castura, Varela, & Næs (2023) [eComponent]
doi:10.1016/j.foodqual.2023.104941

Gain:

1 PC: **15%**

2 PCs: **14%**

3 PCs: **1%**

Investigating a subset of paired comparisons after PCA

Another example where only a subset of paired comparisons are of primary interest:

Temporal sensory data

Focus:

Paired comparisons *within* timepoints, **not** *across* timepoints

Investigating a subset of paired comparisons after PCA

All Pairs

- 8 yogurts × 56 timepoints
- 448 combinations
- All pairs = 100,028

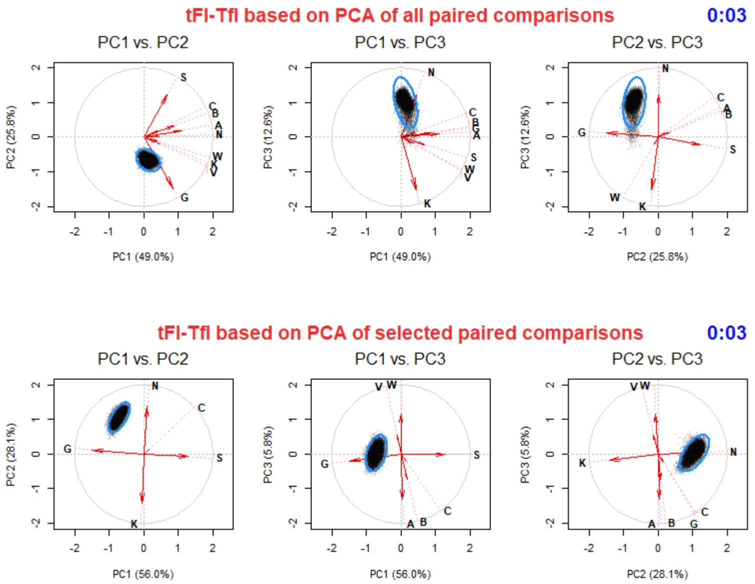
X⊖X has dimension 100028 × 10

Within-timepoint Pairs

- 28 within-timepoint pairs
- 56 timepoints
- C = 28×56 = 1568

Δ* matrix has dimension 3136 × 10

Investigating a subset of paired comparisons after PCA



Castura, Varela, & Næs (2023) [eComponent]
doi:10.1016/j.foodqual.2023.104941

Gain:

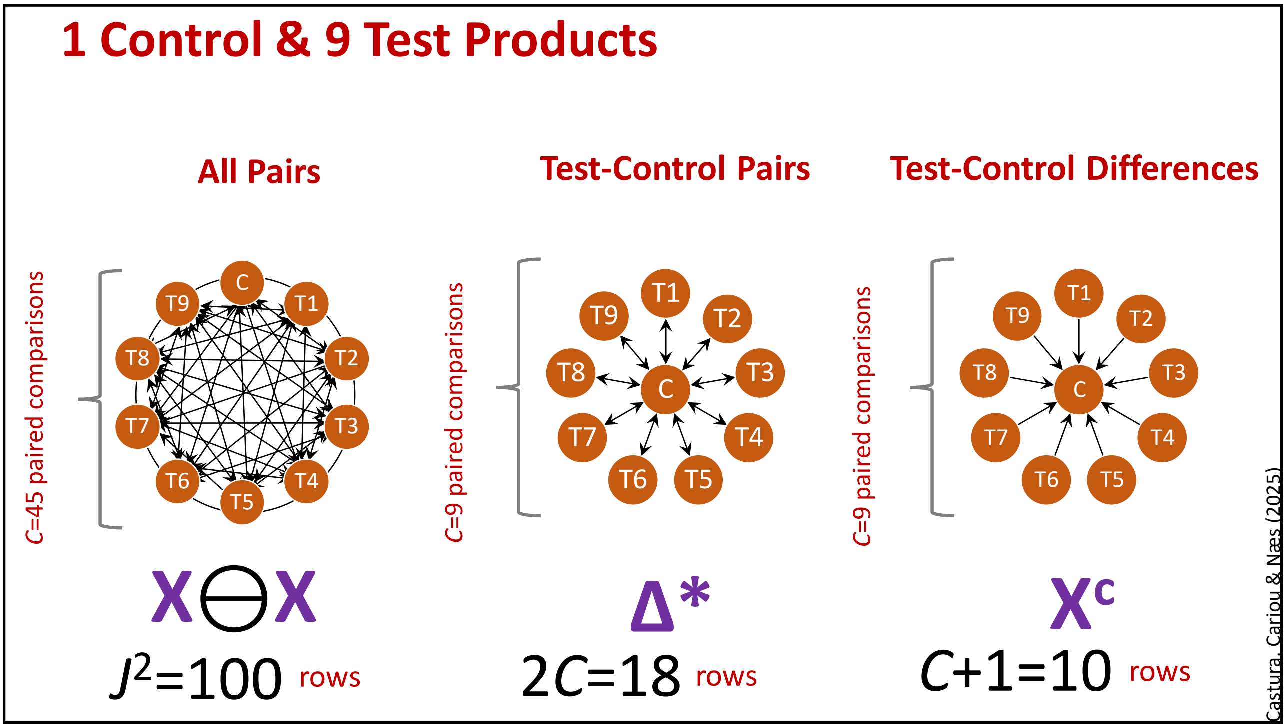
1 PC: **>3500%**

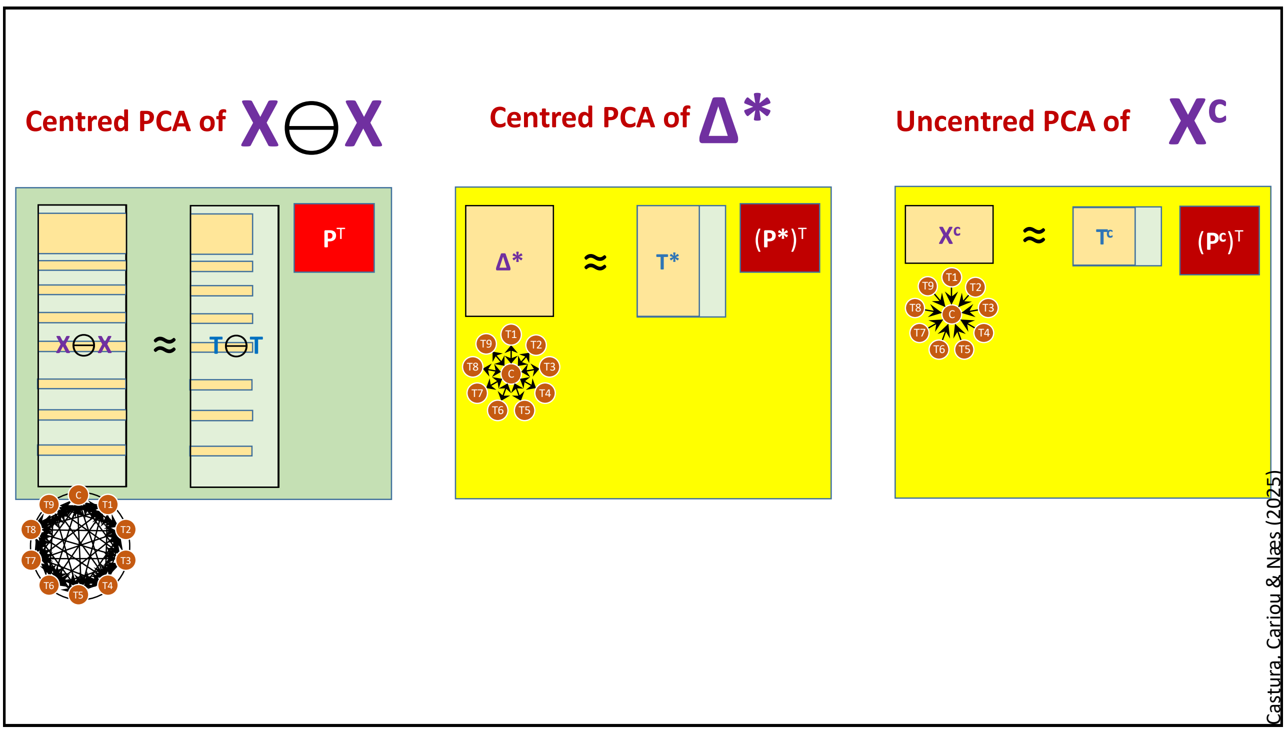
2 PCs: **52%**

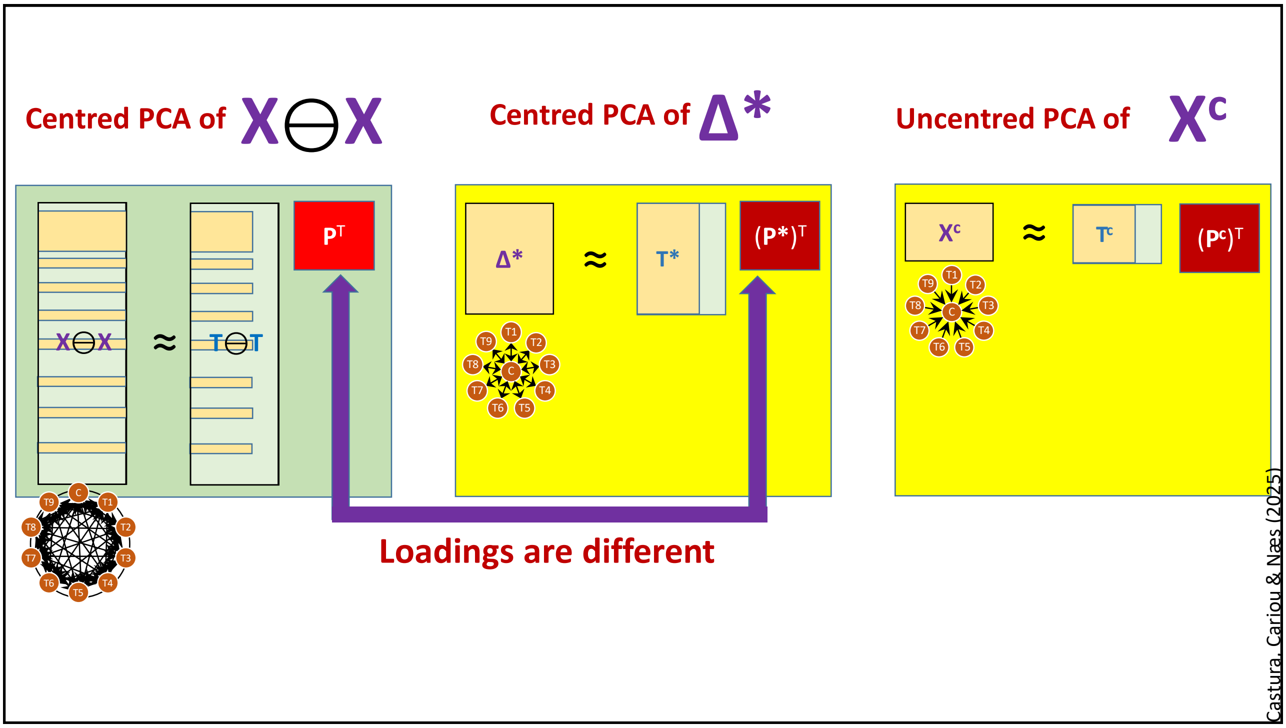
3 PCs: **1%**

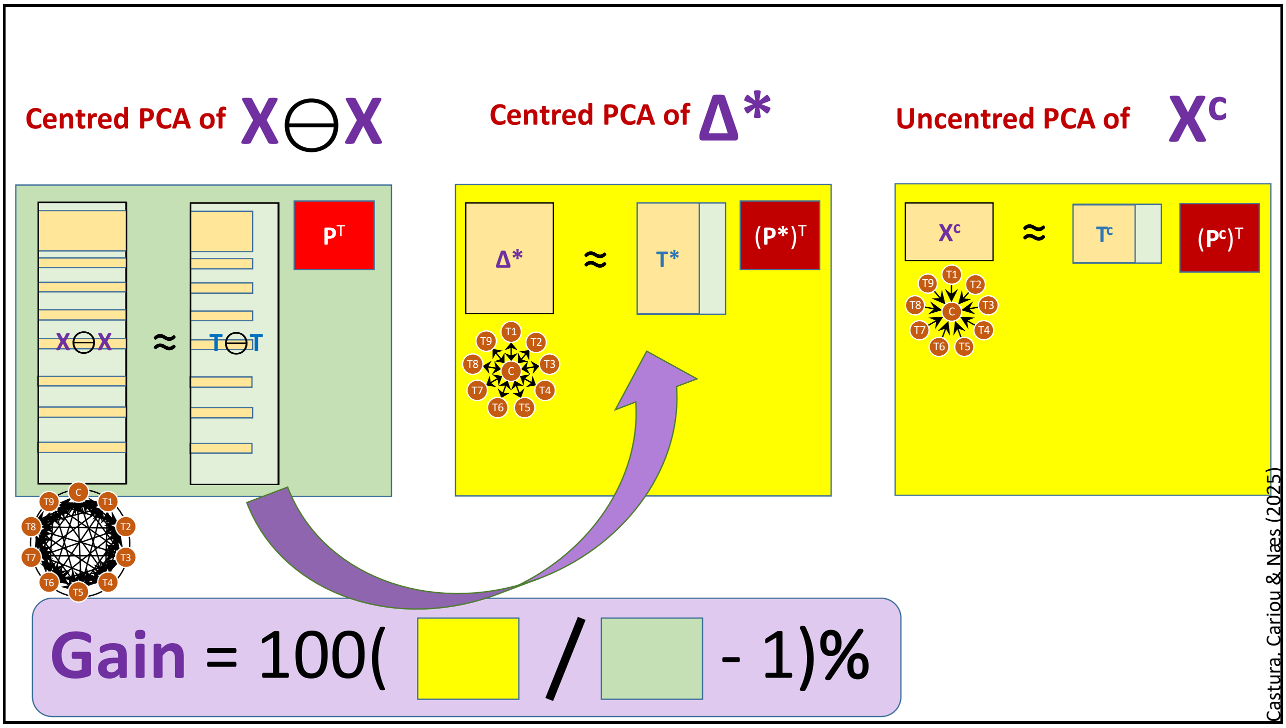
Investigating control-centred results after uncentred principal component analysis

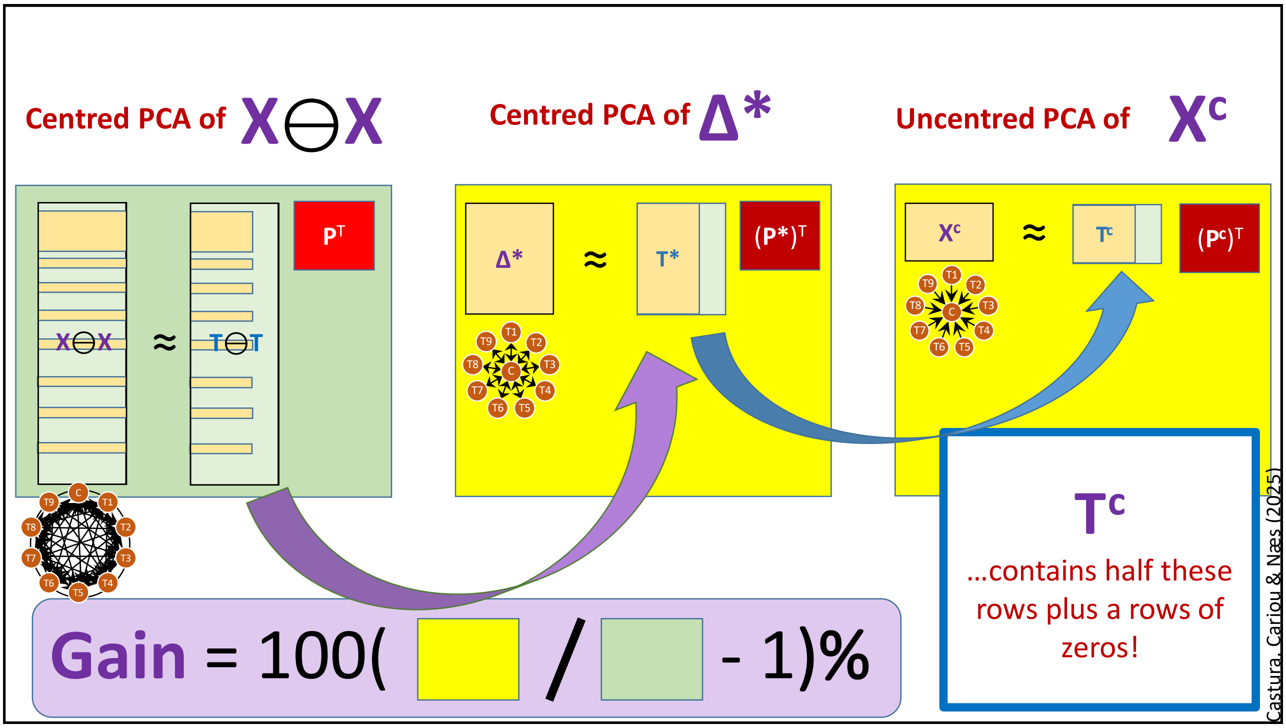
Castura, J.C., Cariou, V., & Næs, T. (2025). Investigating control-centred results after uncentred principal component analysis. *Zenodo.* Preprint.
[Manuscript under review. Preprint not peer reviewed.
This preprint to be updated very soon!].
https://doi.org/10.5281/zenodo.15073361

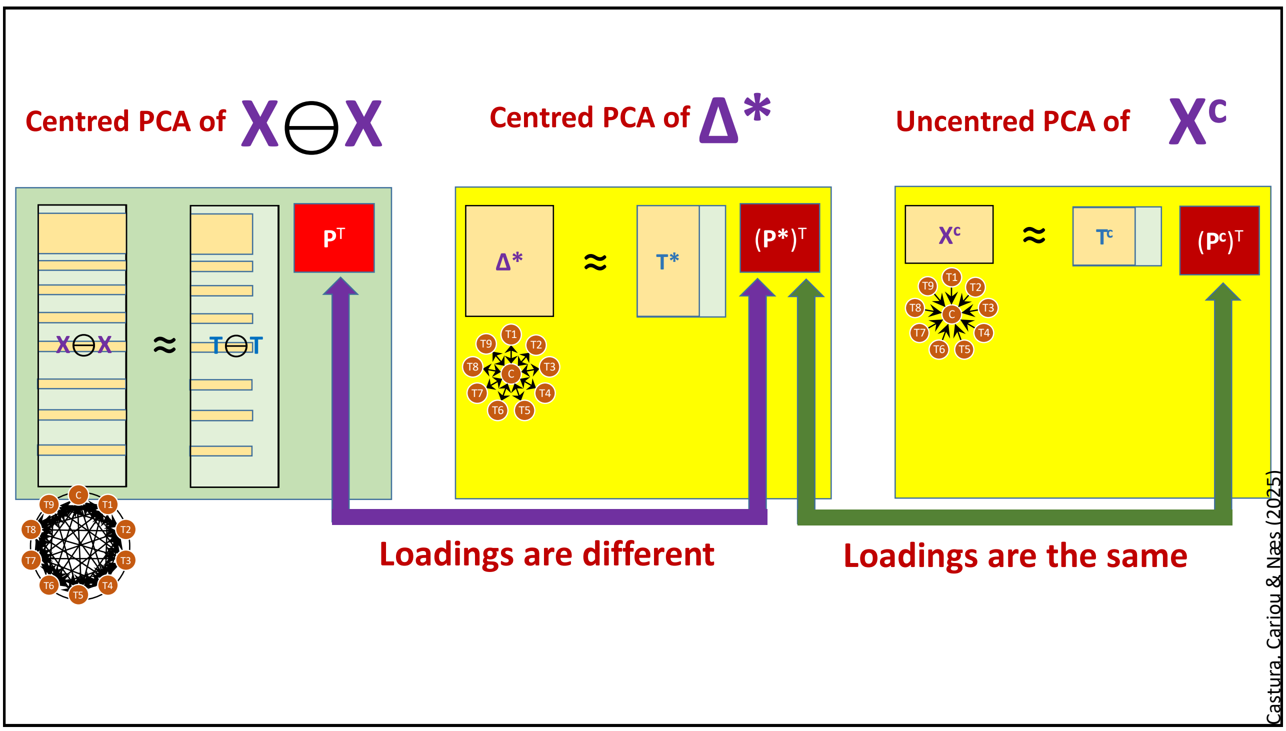

1 Control & 9 Test Products
All Pairs
Test-Control Pairs
Test-Control Differences
C=45 paired comparisons
C=9 paired comparisons
C=9 paired comparisons
C
T1
T2
T3
T4
T5
T6
T7
T8
T9
X⊖X
Δ*
X^c
J²=100 rows
2C=18 rows
C+1=10 rows
Castura, Cariou & Næs (2025)

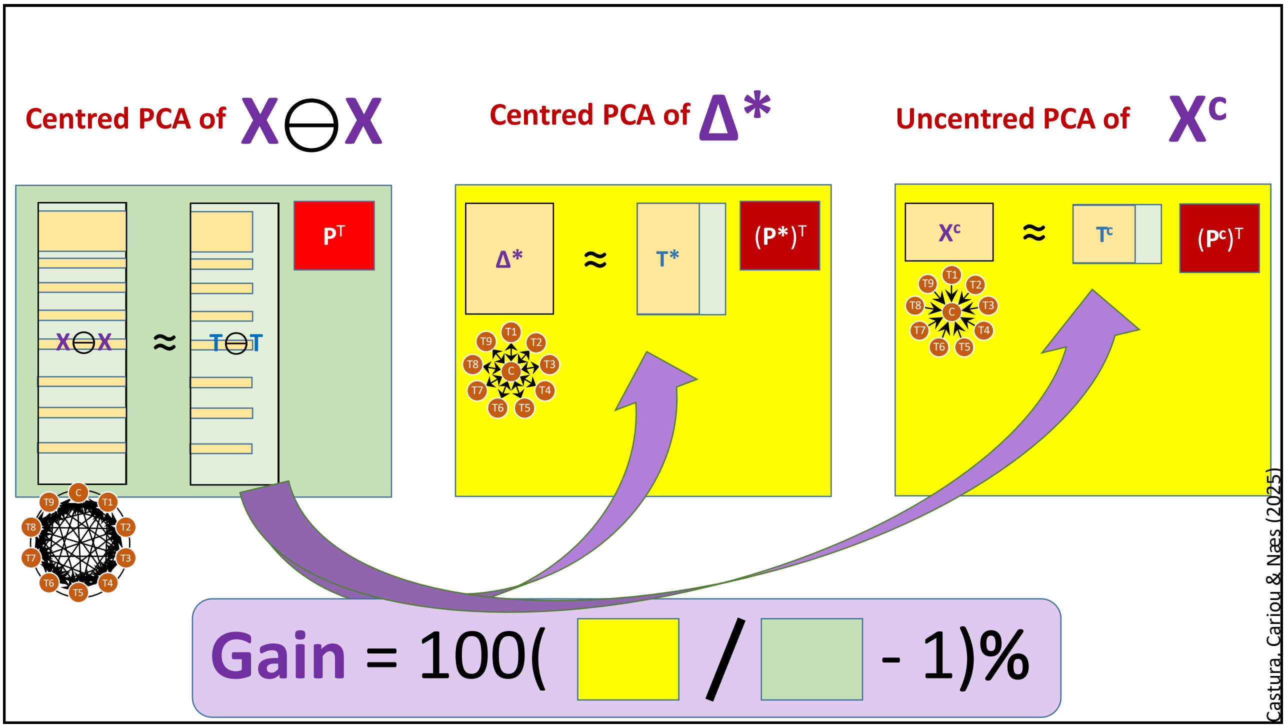

Centred PCA of X⊖X
Centred PCA of Δ*
Uncentred PCA of X^c
X⊖X
≈
T⊖T
P^T
Δ*
≈
T*
(P*)^T
X^c
≈
T^c
(P^c)^T
Castura, Cariou & Næs (2025)

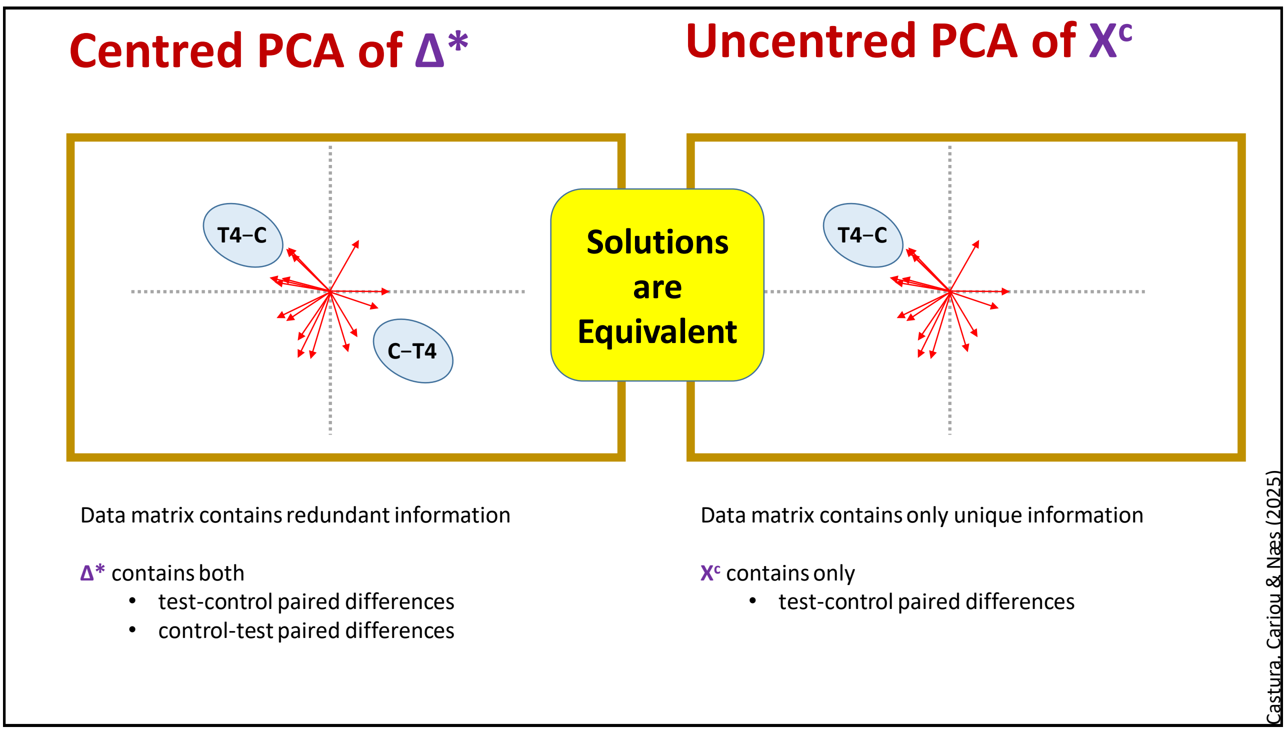

Centred PCA of X⊖X
Centred PCA of Δ*
Uncentred PCA of X^C
X⊖X
≈
T⊖T
P^T
Δ*
≈
T*
(P*)^T
X^c
≈
T^c
(P^c)^T
Loadings are different
Castura, Cariou & Næs (2025)


Centred PCA of X⊖X
Centred PCA of Δ*
Uncentred PCA of X^C
X⊖X
≈
T⊖T
P^T
Δ*
≈
T*
(P*)^T
X^c
≈
T^c
(P^c)^T
Gain = 100(/ - 1)%
Castura, Cariou & Næs (2025)

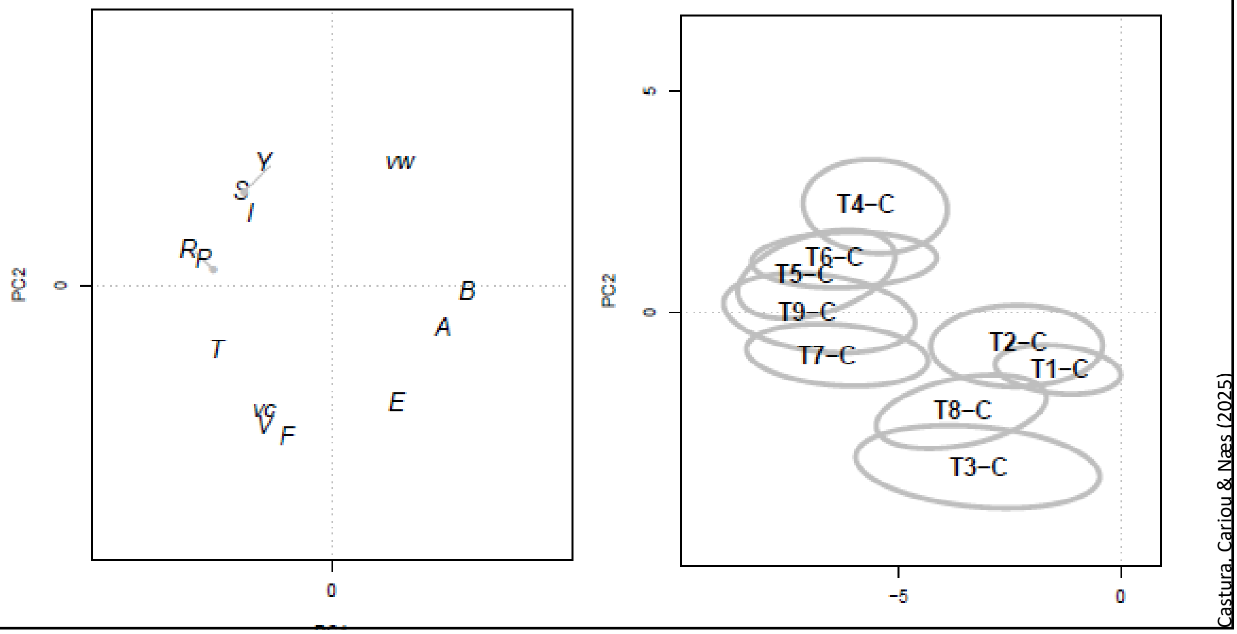

Centred PCA of X⊖X
Centred PCA of Δ*
Uncentred PCA of Xᶜ
Pᵀ
X⊖X
≈
T⊖T
Δ*
≈
T*
(P*)ᵀ
Xᶜ
≈
Tᶜ
(Pᶜ)ᵀ
Gain = 100(/ - 1)%
Tᶜ
…contains half these rows plus a rows of zeros!
Castura, Cariou & Næs (2025)


Centred PCA of X⊖X
Centred PCA of Δ*
Uncentred PCA of Xᶜ
Pᵀ
X⊖X
≈
T⊖T
Δ*
≈
T*
(P*)ᵀ
Xᶜ
≈
Tᶜ
(Pᶜ)ᵀ
Loadings are different
Loadings are the same
Castura, Cariou & Næs (2025)

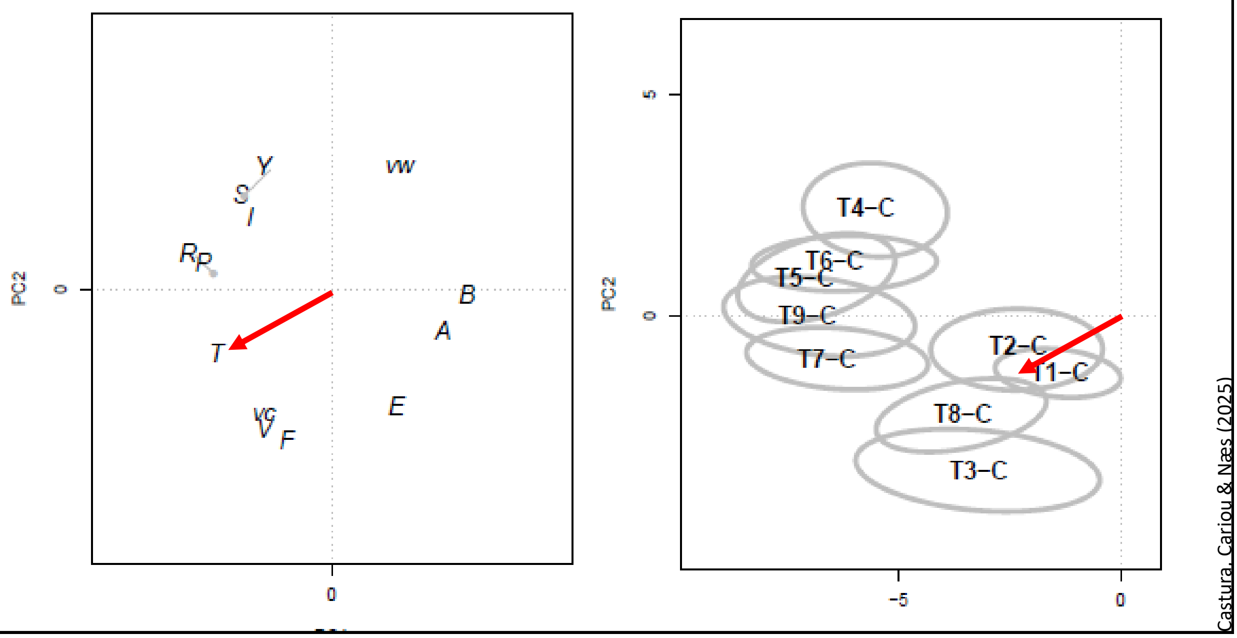

Centred PCA of X⊖X
Centred PCA of Δ*
Uncentred PCA of Xᶜ
Pᵀ
X⊖X
≈
T⊖T
Δ*
≈
T*
(P*)ᵀ
Xᶜ
≈
Tᶜ
(Pᶜ)ᵀ
Gain = 100(/ - 1)%
Castura, Cariou & Næs (2025)


Centred PCA of Δ*
Uncentred PCA of Xᶜ
T4–C
C–T4
Solutions are Equivalent
T4–C
Data matrix contains redundant information
Δ* contains both
• test-control paired differences
• control-test paired differences
Data matrix contains only unique information
Xᶜ contains only
• test-control paired differences
Castura, Cariou & Næs (2025)

Uncentred PCA of X^c



Uncentred PCA of X^c



For investigating Test-Control paired comparisons…

Advantages of uncentred PCA of X^c

- "best" subspace for investigating the relevant test-control pairs
- components ordered by importance
- these particular uncentred PCA results have a conventional "variance interpretation"
- origin interpretation: "no difference from control"

Castura, Cariou & Næs (2025)

Supervised principal component regression of selected paired comparisons

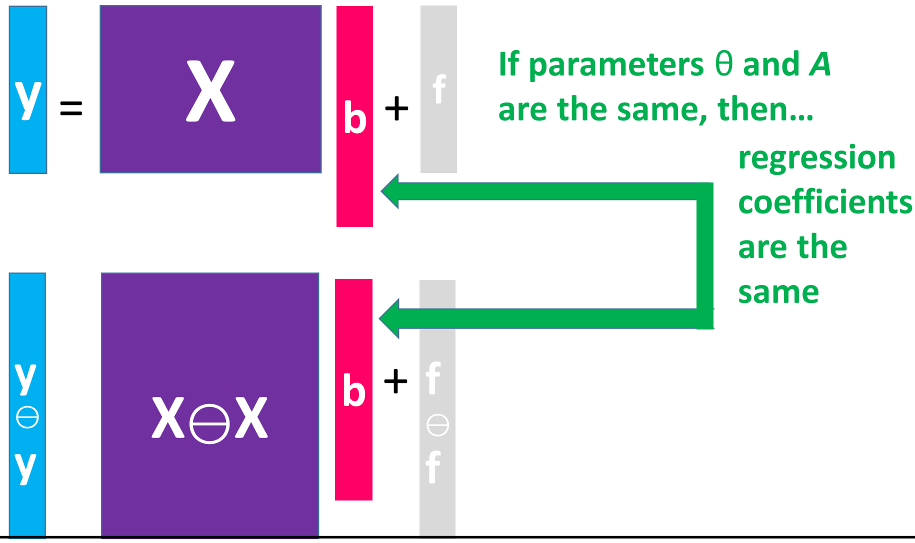
Castura, J.C., & Tomic, O. (2024). Supervised principal component regression of select paired comparisons. *Zenodo*. [Manuscript under review. Preprint not peer reviewed]. https://doi.org/10.5281/zenodo.11663995

This preprint to be updated very soon!

Castura, J.C., & Tomic, O. (2025). Supervised principal component regression of select paired comparisons. [To be uploaded soon to the *Zenodo* preprint server!] [Manuscript under review. Preprint not peer reviewed].

Investigating data relationships

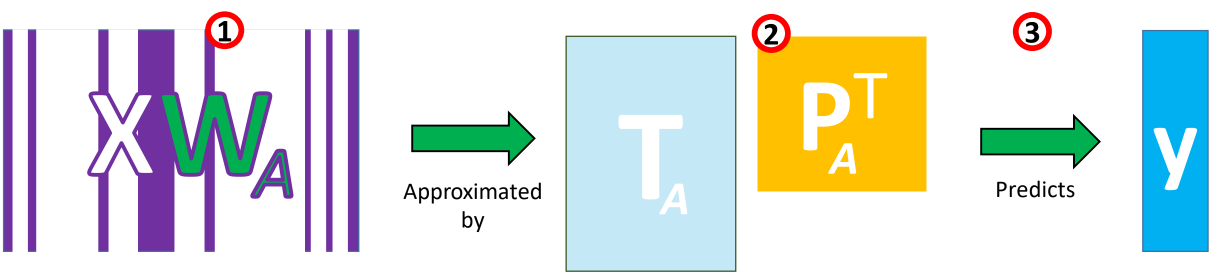
with **supervised principal component regression (SPCR)**



Castura & Tomic (2024)

Investigating data relationships

with **supervised principal component regression (SPCR)**



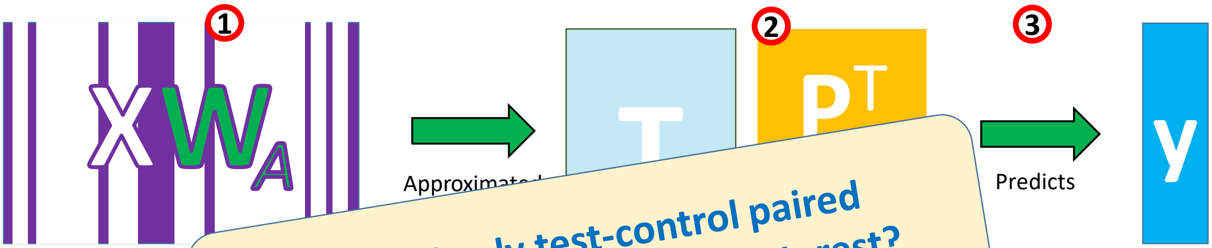
SPCR of **X** and **y** is equivalent to SPCR of **X**⊖**X** and **y**⊖**y** because regression coefficients are the same, which occurs the following are the same:

1. **X** weights
2. **X** loadings
3. **Y** loadings

Castura & Tomic (2024)

Investigating data relationships

with **supervised principal component regression (SPCR)**



What if only test-control paired differences are of primary interest?

SPCR of X a... and y⊖y because regression coefficients ... which occurs the following are the same:

1. X weights
2. X loadings
3. Y loadings

Castura & Tomic (2024)

Investigating data relationships

with **supervised principal component regression (SPCR)**

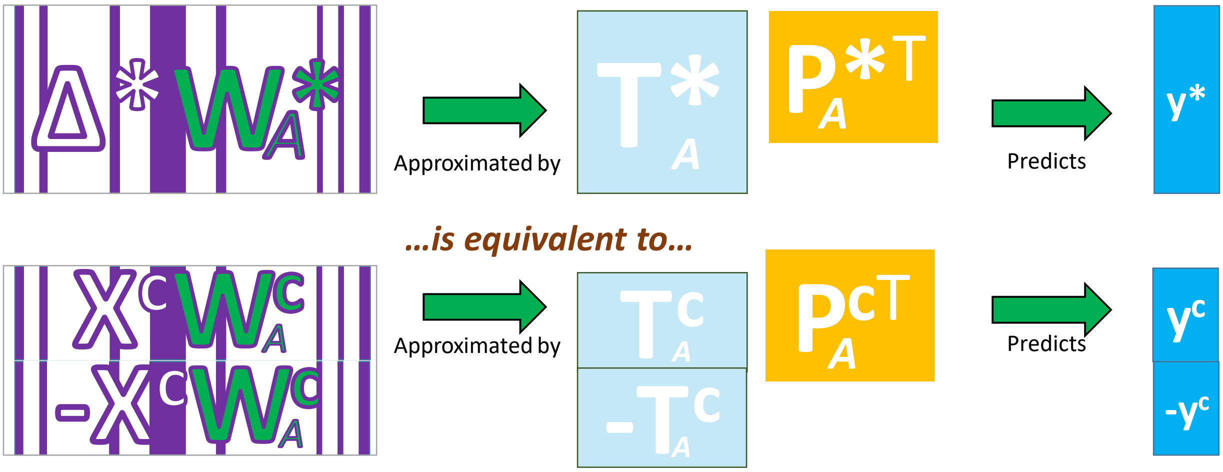
We know we can focus on paired comparisons or paired differences since…

…however, using Δ* introduces a subtle problem…

Castura & Tomic (2025)

Investigating data relationships

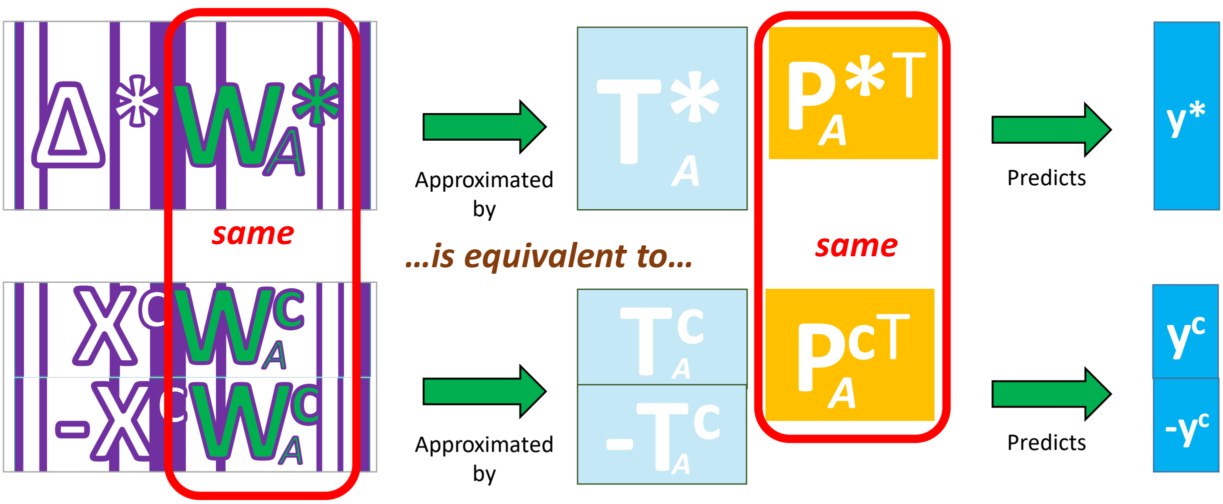
with **supervised principal component regression (SPCR)**



Castura & Tomic (2025)

Investigating data relationships

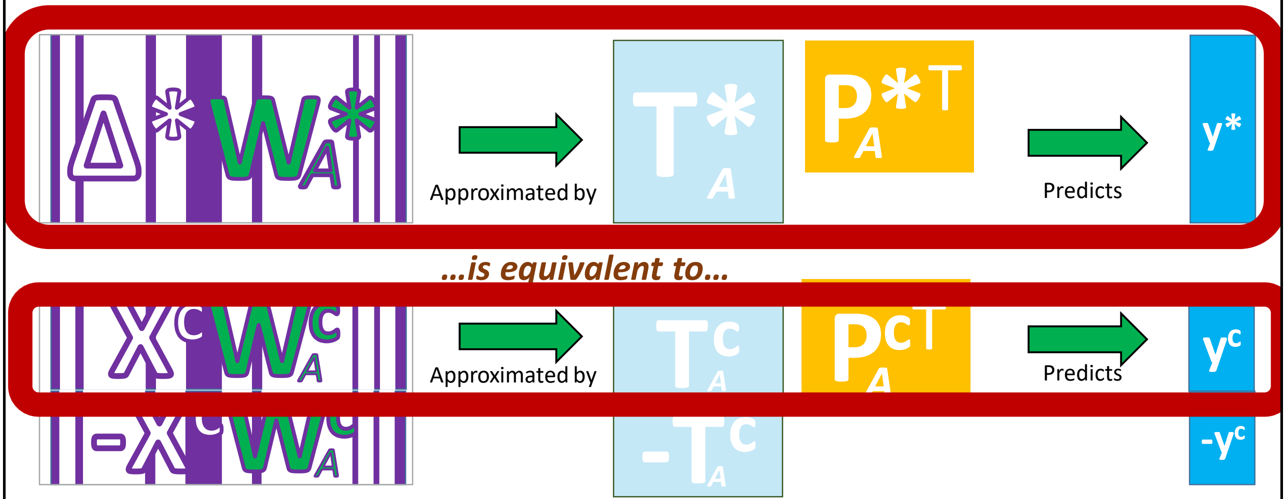
with **supervised principal component regression (SPCR)**



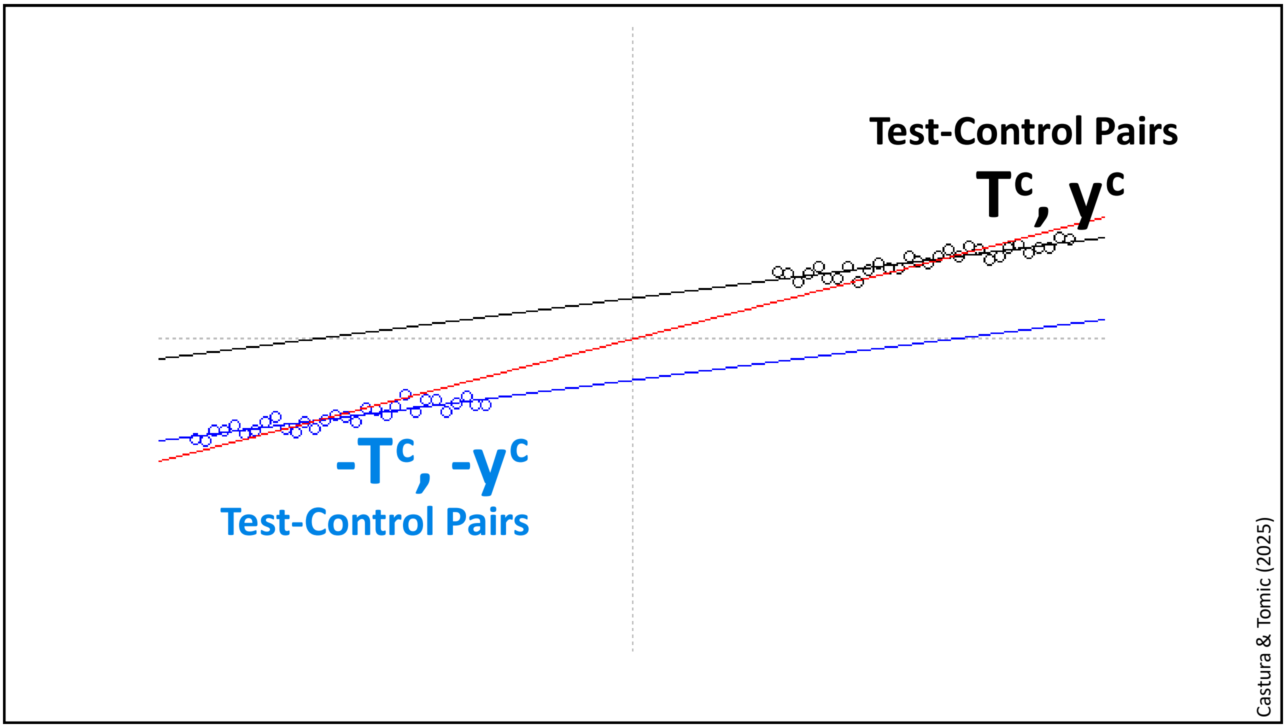
Castura & Tomic (2025)

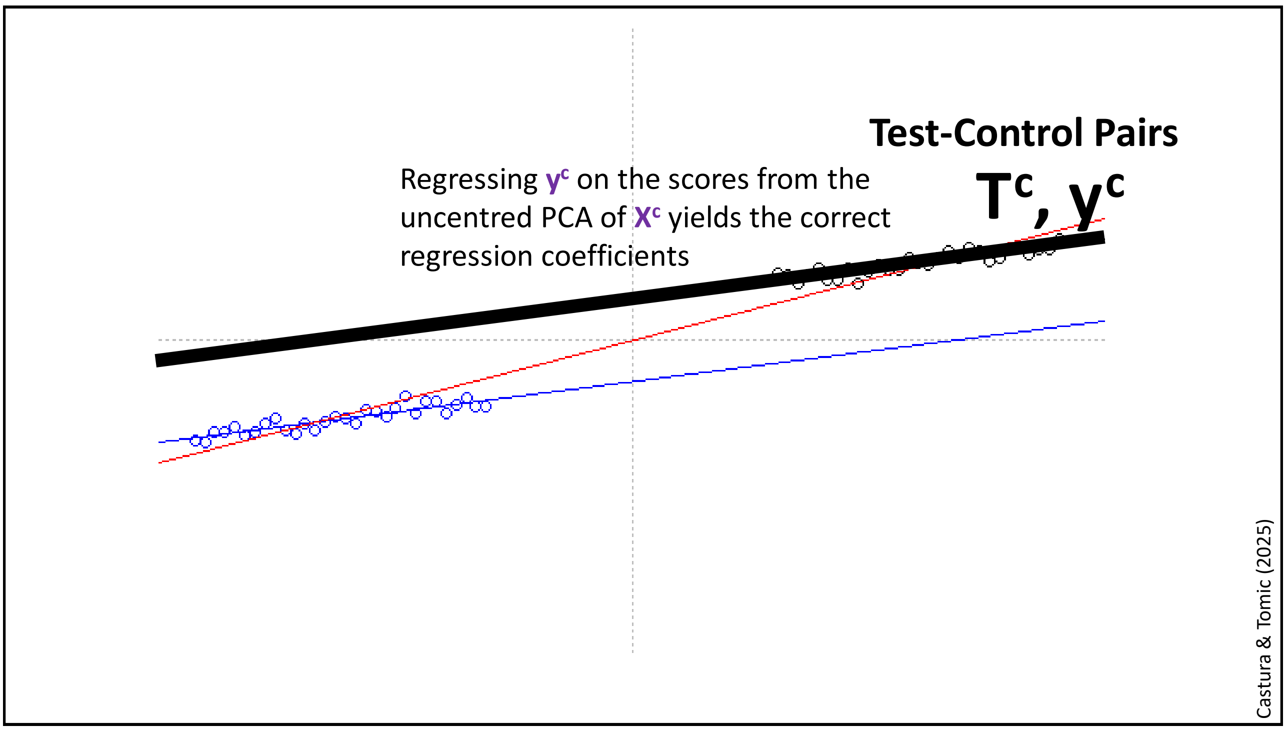
Investigating data relationships

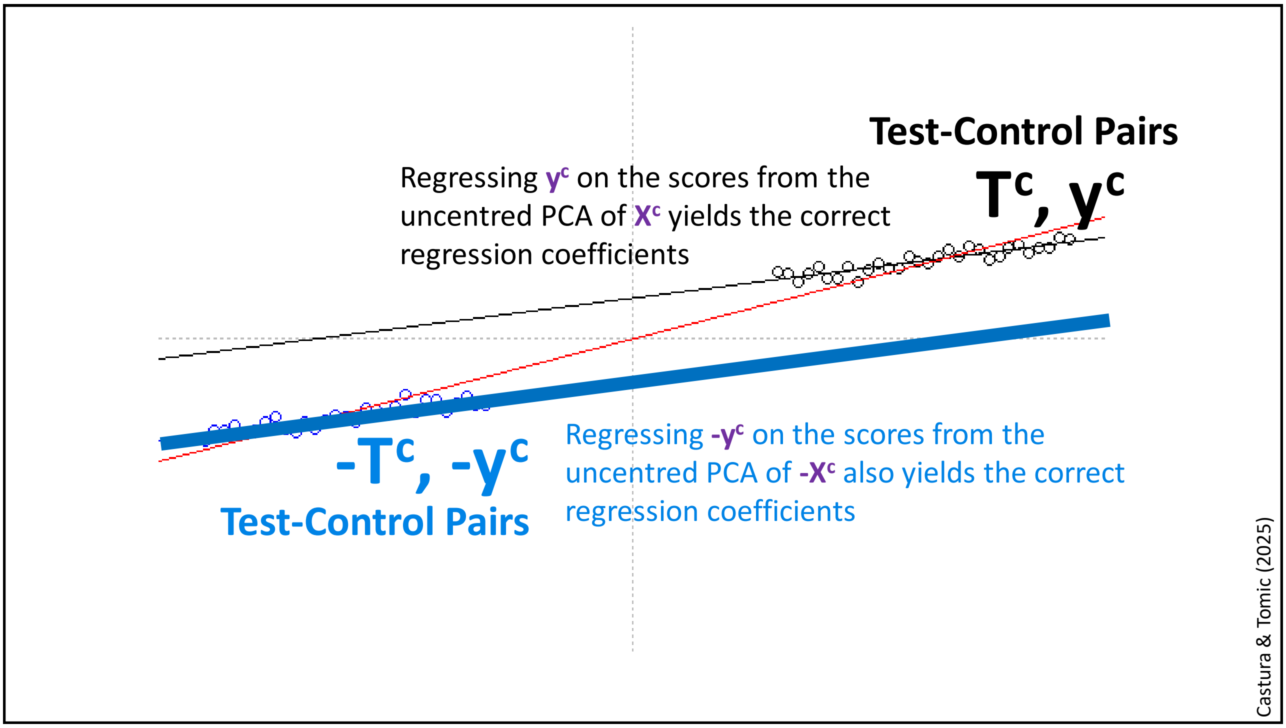
with **supervised principal component regression (SPCR)**

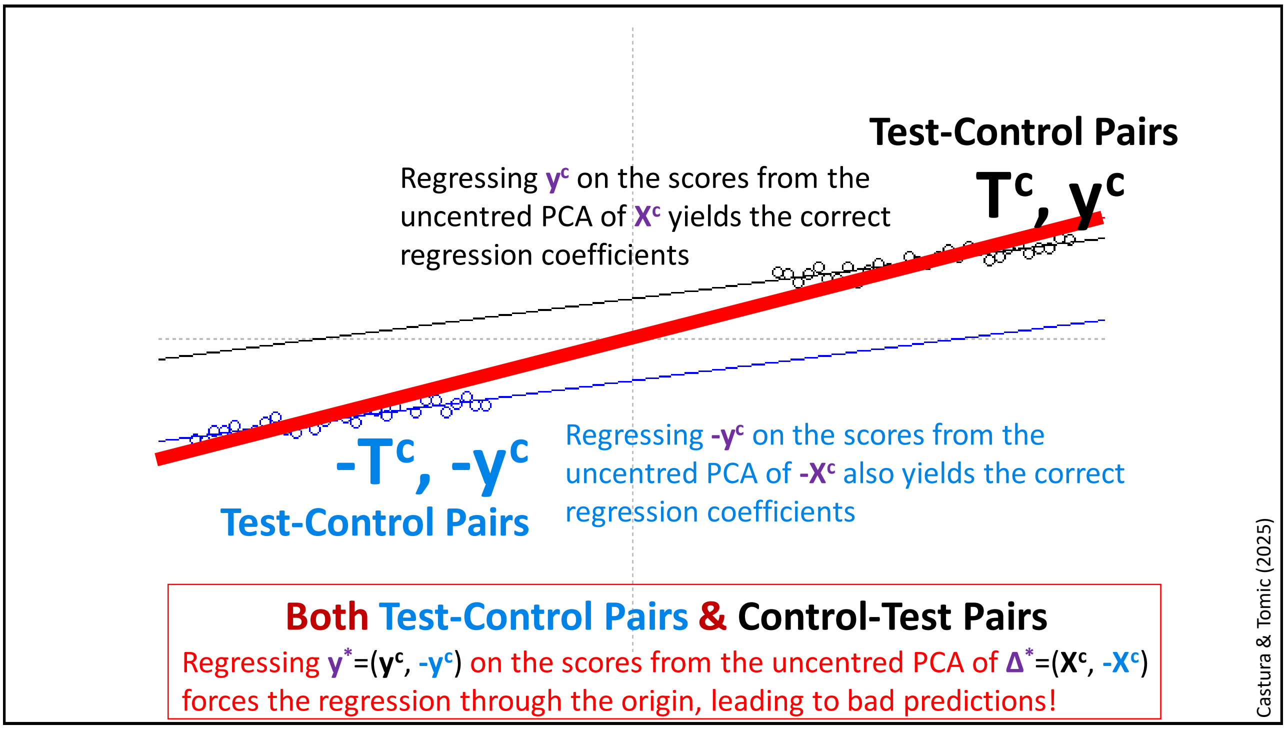


...however, using **Δ*** introduces a subtle problem...


Test-Control Pairs
T^c, y^c
$-T^c, -y^c$
Test-Control Pairs
Castura & Tomic (2025)


Test-Control Pairs
T^c, y^c
Regressing y^c on the scores from the uncentred PCA of X^c yields the correct regression coefficients
Castura & Tomic (2025)


Test-Control Pairs
Tc, yc
Regressing yc on the scores from the uncentred PCA of Xc yields the correct regression coefficients
-Tc, -yc
Test-Control Pairs
Regressing -yc on the scores from the uncentred PCA of -Xc also yields the correct regression coefficients
Castura & Tomic (2025)


Test-Control Pairs
Tc, yc
Regressing yc on the scores from the uncentred PCA of Xc yields the correct regression coefficients
-Tc, -yc
Test-Control Pairs
Regressing -yc on the scores from the uncentred PCA of -Xc also yields the correct regression coefficients
Both Test-Control Pairs & Control-Test Pairs
Regressing y*=(yc, -yc) on the scores from the uncentred PCA of Δ*=(Xc, -Xc) forces the regression through the origin, leading to bad predictions!
Castura & Tomic (2025)

Many candidate values for
parameters θ and ***A***

Since we are focused on paired comparisons, we **adapted** **leave-one-out cross-validation** to handle ***dependent data***.

Castura & Tomic (2025)

Conventional loocv	**Adapted loocv**
One row left out per fold.	One object left out per fold.
Each row is an object.	Multiple rows are left out.
No information from the left-out object leaks into the model during training.	No information from the left-out object can leak into the model during training.
Software for conventional loocv widely available.	Functions for adapted loocv must be coded.

Castura & Tomic (2025)

Predicted error sum of squares (PRESS)

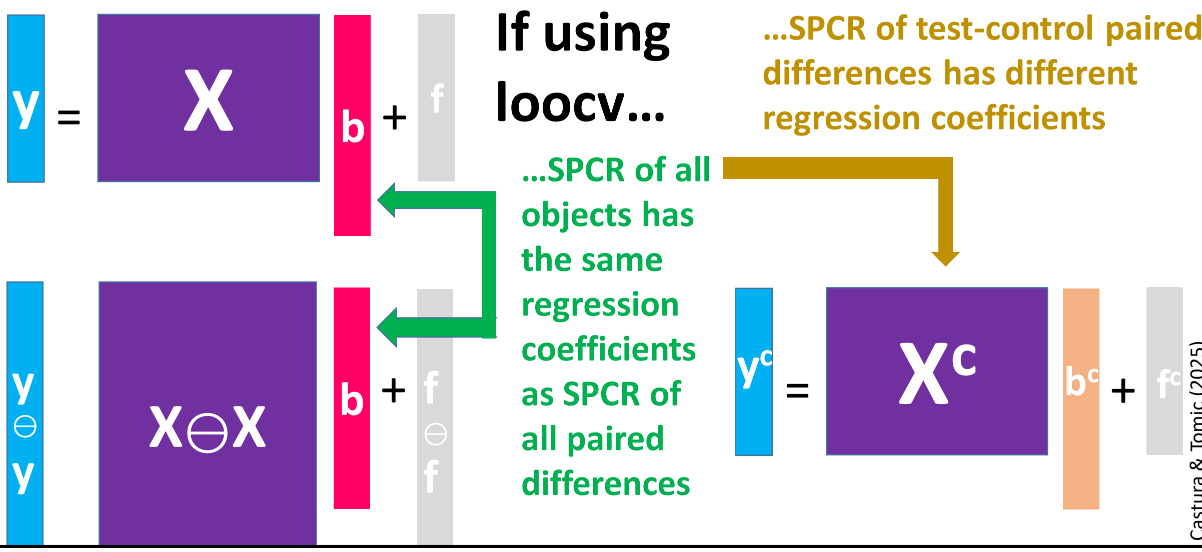
Models evaluated by calculating PRESS statistic calculated for left-out test-control paired differences.

$$\sum (predicted - observed)^2$$

Castura & Tomic (2025)

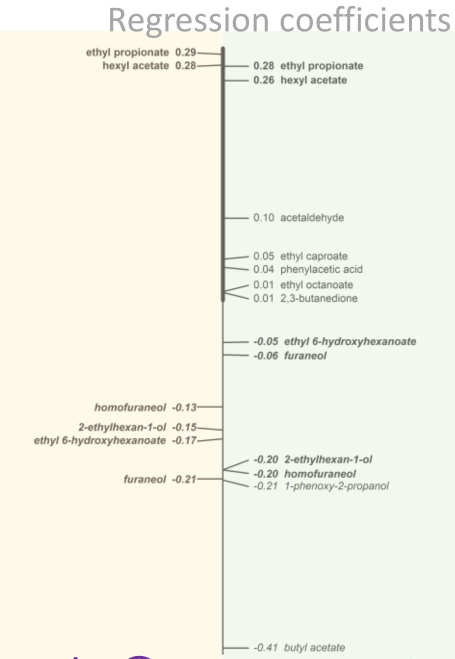
Investigating data relationships

with **supervised principal component regression (SPCR)**



Castura & Tomic (2025)

Cooked cherry aroma



6 Predictors
2 PCs
PRESS 1.75

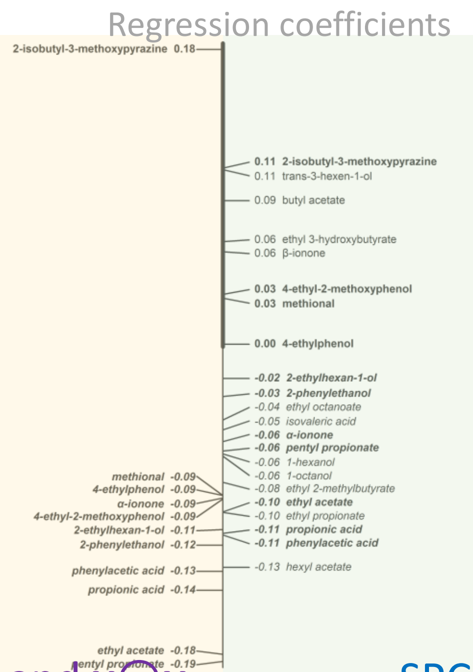
13 Predictors
2 PCs
***PRESS* 1.48**

SPCR of **X**⊖**X** and **y**⊖**y**

SPCR of $\mathbf{X}^c$ and $\mathbf{y}^c$

Castura & Tomic (2025)

Fresh cherry aroma



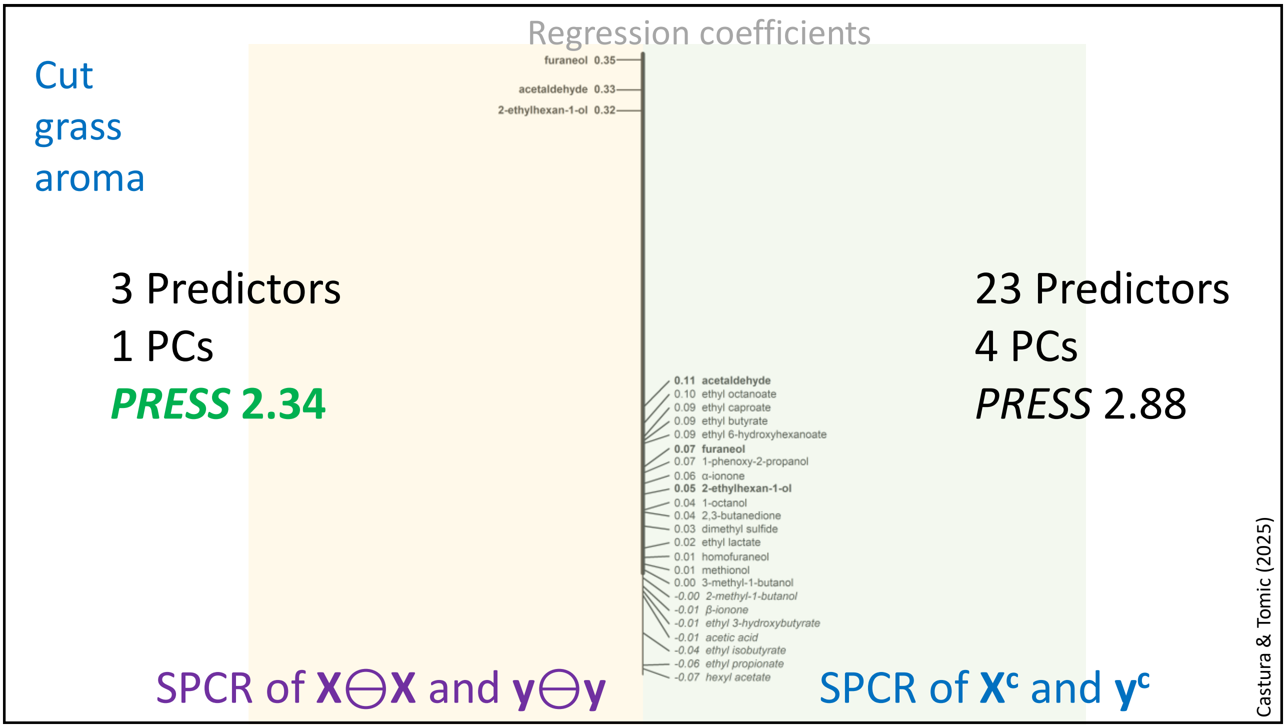
11 Predictors
1 PCs
PRESS 1.67

22 Predictors
3 PCs
***PRESS* 0.89**

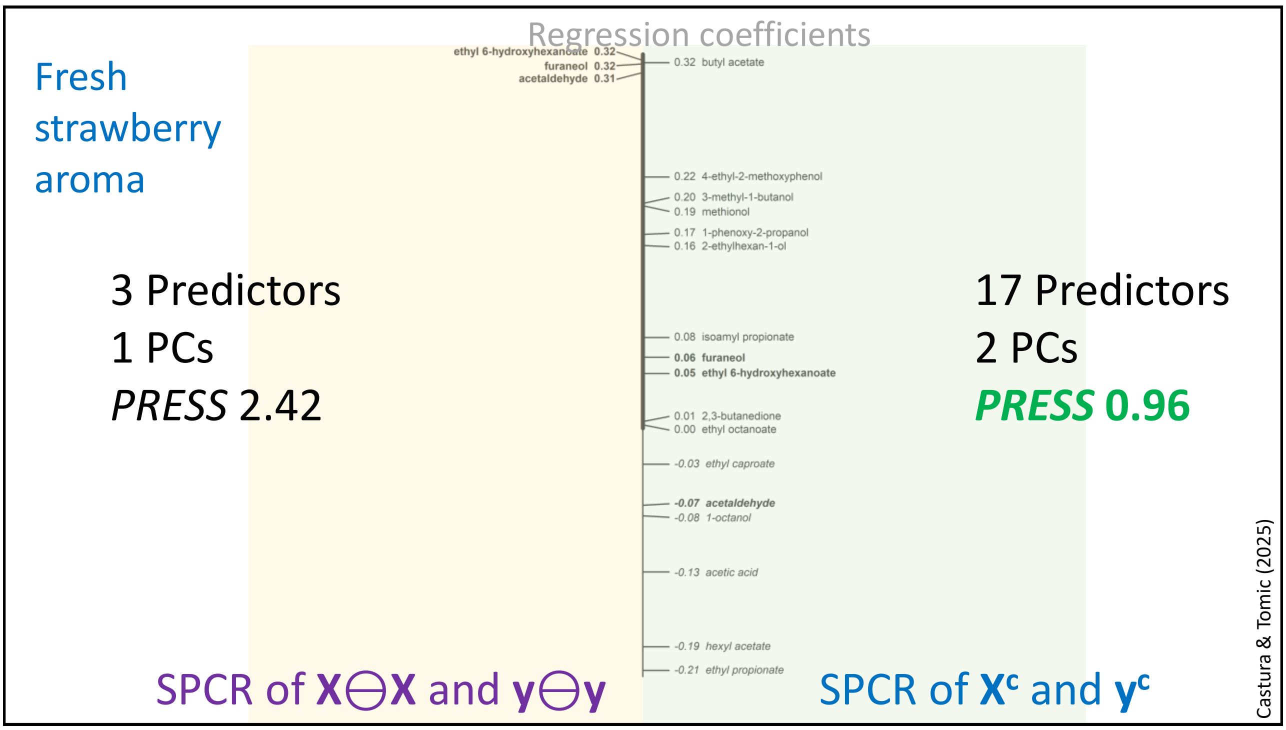
SPCR of **X**⊖**X** and **y**⊖**y**

SPCR of $\mathbf{X}^c$ and $\mathbf{y}^c$

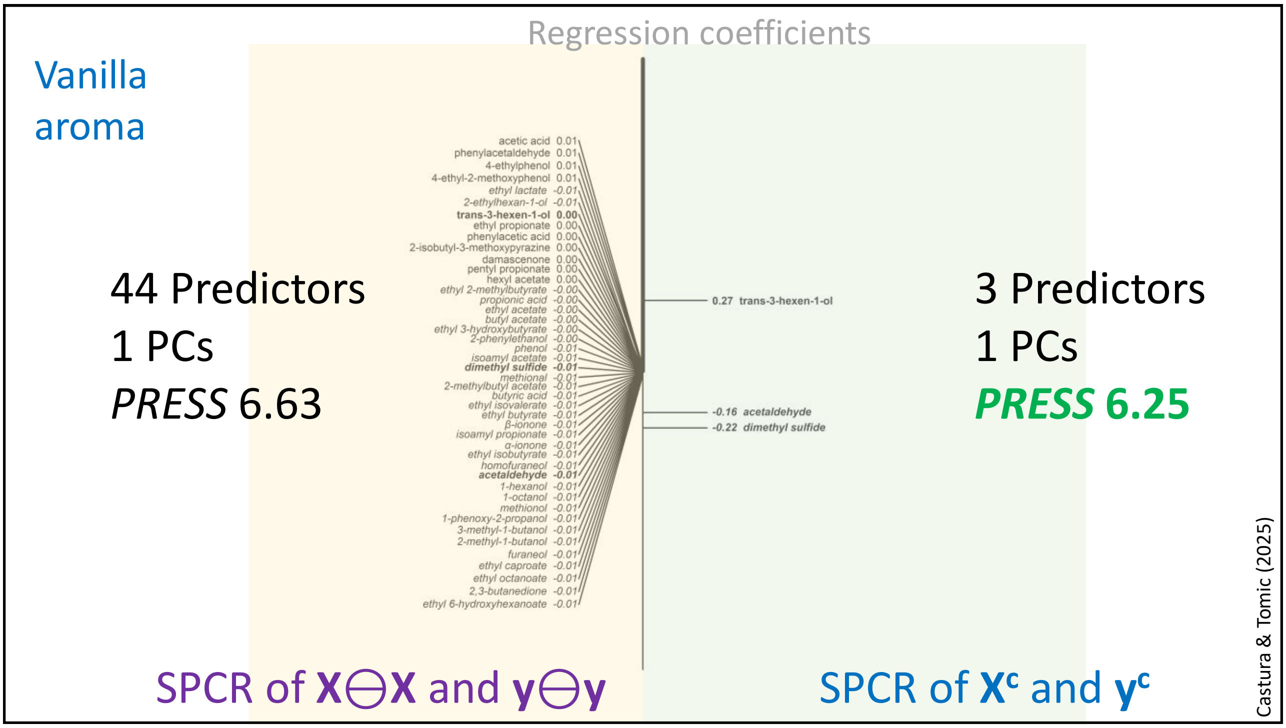
Castura & Tomic (2025)



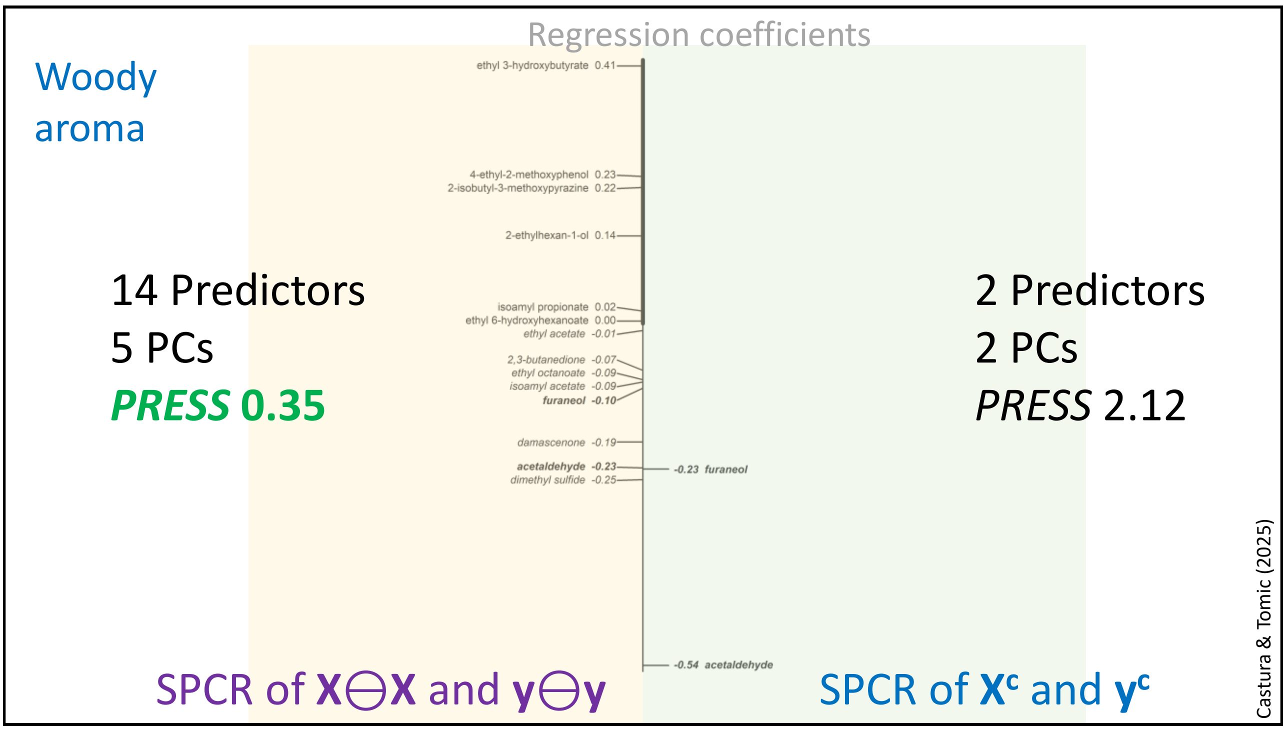
Regression coefficients
Cut grass aroma
furaneol 0.35
acetaldehyde 0.33
2-ethylhexan-1-ol 0.32
3 Predictors
1 PCs
PRESS 2.34
0.11 acetaldehyde
0.10 ethyl octanoate
0.09 ethyl caproate
0.09 ethyl butyrate
0.09 ethyl 6-hydroxyhexanoate
0.07 furaneol
0.07 1-phenoxy-2-propanol
0.06 α-ionone
0.05 2-ethylhexan-1-ol
0.04 1-octanol
0.04 2,3-butanedione
0.03 dimethyl sulfide
0.02 ethyl lactate
0.01 homofuraneol
0.01 methionol
0.00 3-methyl-1-butanol
-0.00 2-methyl-1-butanol
-0.01 β-ionone
-0.01 ethyl 3-hydroxybutyrate
-0.01 acetic acid
-0.04 ethyl isobutyrate
-0.06 ethyl propionate
-0.07 hexyl acetate
23 Predictors
4 PCs
PRESS 2.88
SPCR of X⊖X and y⊖y
SPCR of Xc and yc
Castura & Tomic (2025)



Regression coefficients
Fresh strawberry aroma
ethyl 6-hydroxyhexanoate 0.32
furaneol 0.32
acetaldehyde 0.31
3 Predictors
1 PCs
PRESS 2.42
0.32 butyl acetate
0.22 4-ethyl-2-methoxyphenol
0.20 3-methyl-1-butanol
0.19 methionol
0.17 1-phenoxy-2-propanol
0.16 2-ethylhexan-1-ol
0.08 isoamyl propionate
0.06 furaneol
0.05 ethyl 6-hydroxyhexanoate
0.01 2,3-butanedione
0.00 ethyl octanoate
-0.03 ethyl caproate
-0.07 acetaldehyde
-0.08 1-octanol
-0.13 acetic acid
-0.19 hexyl acetate
-0.21 ethyl propionate
17 Predictors
2 PCs
PRESS 0.96
SPCR of X⊖X and y⊖y
SPCR of Xc and yc
Castura & Tomic (2025)



Regression coefficients
Vanilla aroma
44 Predictors
1 PCs
PRESS 6.63
0.27 trans-3-hexen-1-ol
-0.16 acetaldehyde
-0.22 dimethyl sulfide
3 Predictors
1 PCs
PRESS 6.25
SPCR of X⊖X and y⊖y
SPCR of Xᶜ and yᶜ
Castura & Tomic (2025)

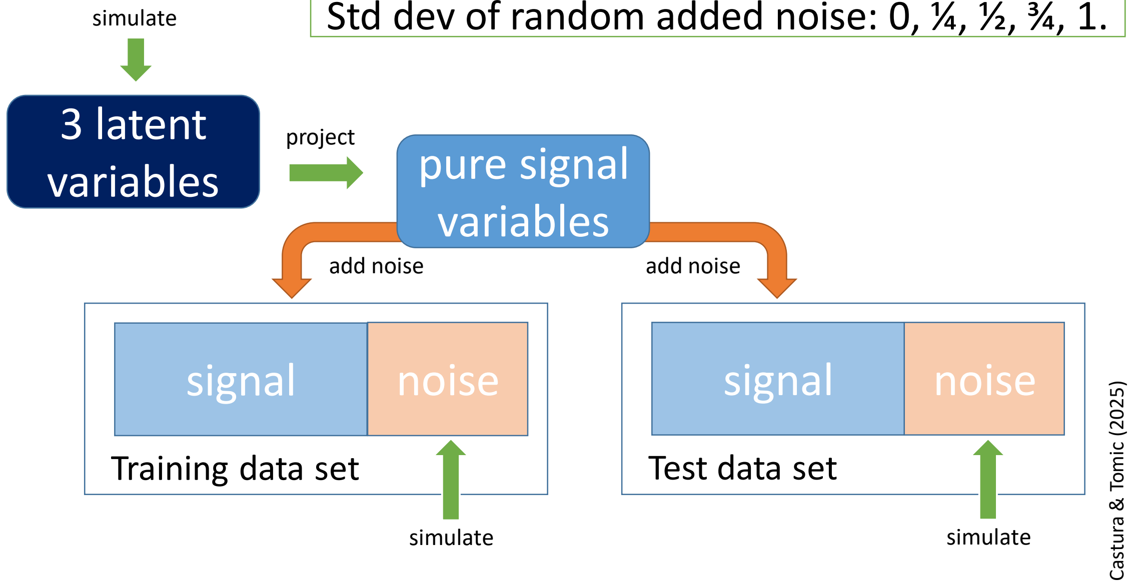


Regression coefficients
Woody aroma
ethyl 3-hydroxybutyrate 0.41
4-ethyl-2-methoxyphenol 0.23
2-isobutyl-3-methoxypyrazine 0.22
2-ethylhexan-1-ol 0.14
isoamyl propionate 0.02
ethyl 6-hydroxyhexanoate 0.00
ethyl acetate -0.01
2,3-butanedione -0.07
ethyl octanoate -0.09
isoamyl acetate -0.09
furaneol -0.10
damascenone -0.19
acetaldehyde -0.23
dimethyl sulfide -0.25
-0.23 furaneol
-0.54 acetaldehyde
14 Predictors
5 PCs
PRESS 0.35
2 Predictors
2 PCs
PRESS 2.12
SPCR of X⊖X and y⊖y
SPCR of Xᶜ and yᶜ
Castura & Tomic (2025)

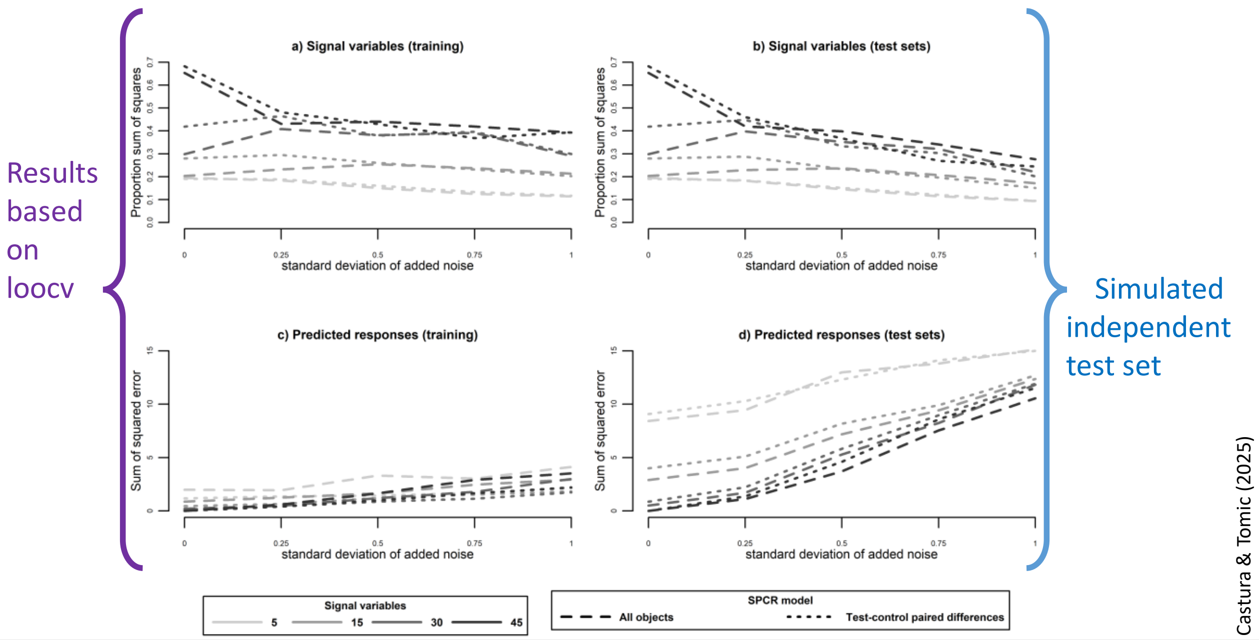
Simulation study

Each condition was a combination of…

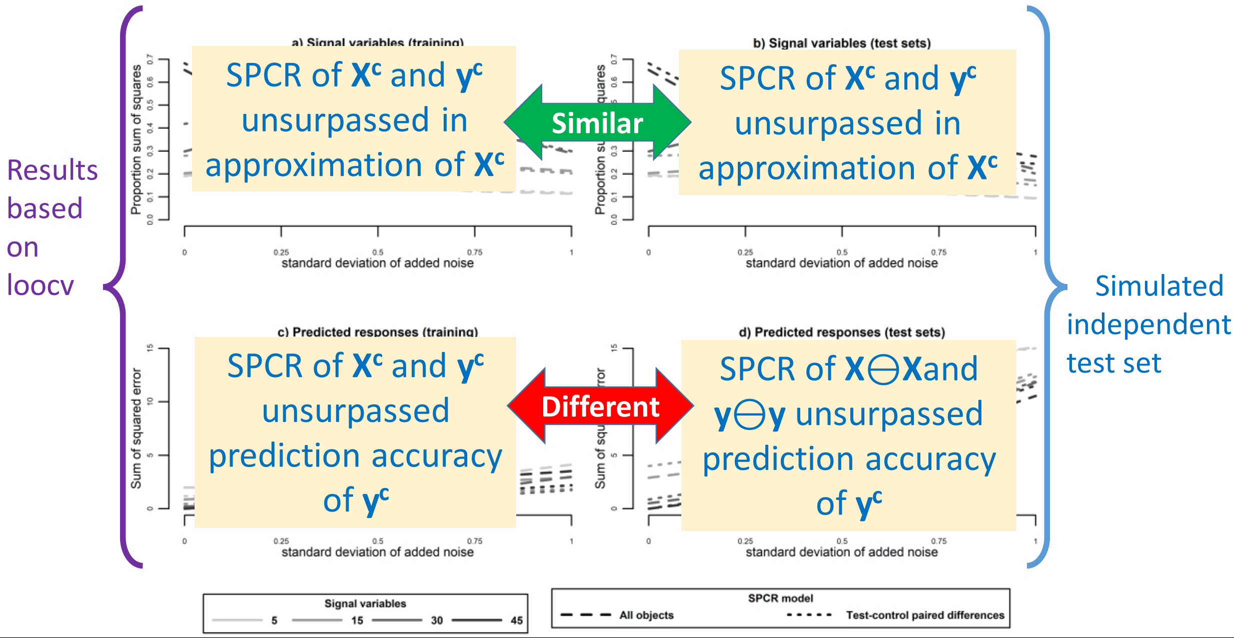
Number of “signal variables”: 5, 15, 30, 45.
Std dev of random added noise: 0, ¼, ½, ¾, 1.



Simulation study



Simulation study



Castura & Tomic (2025)

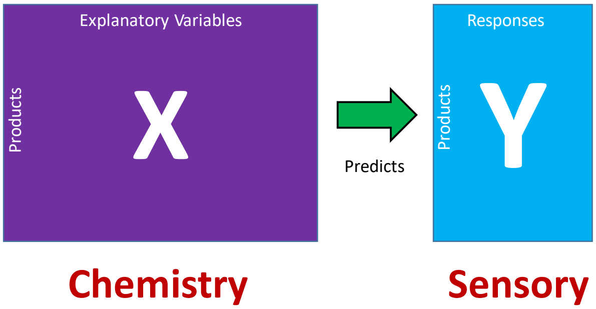
The simulation study shows the superiority of the SPCR of test-control paired differences model for the training set does not generalize to test sets simulated from the same data generating process.

We are only interested in the test-control paired differences, not the test-test paired differences, but since both types of pairs carry information about the association between variables, the relationship between predictor and response variables is modelled better by the SPCR of all objects model than by the SPCR of test-control paired differences model.

This finding shows **overfitting appears in many guises**.

For this reason, the SPCR of all objects model is recommended for most routine analyses.

Castura & Tomic (2025)



Partial least squares regression (PLSR)

Partial least squares regression (PLSR)

We want to ***predict*** multivariate Y from the multivariate X.

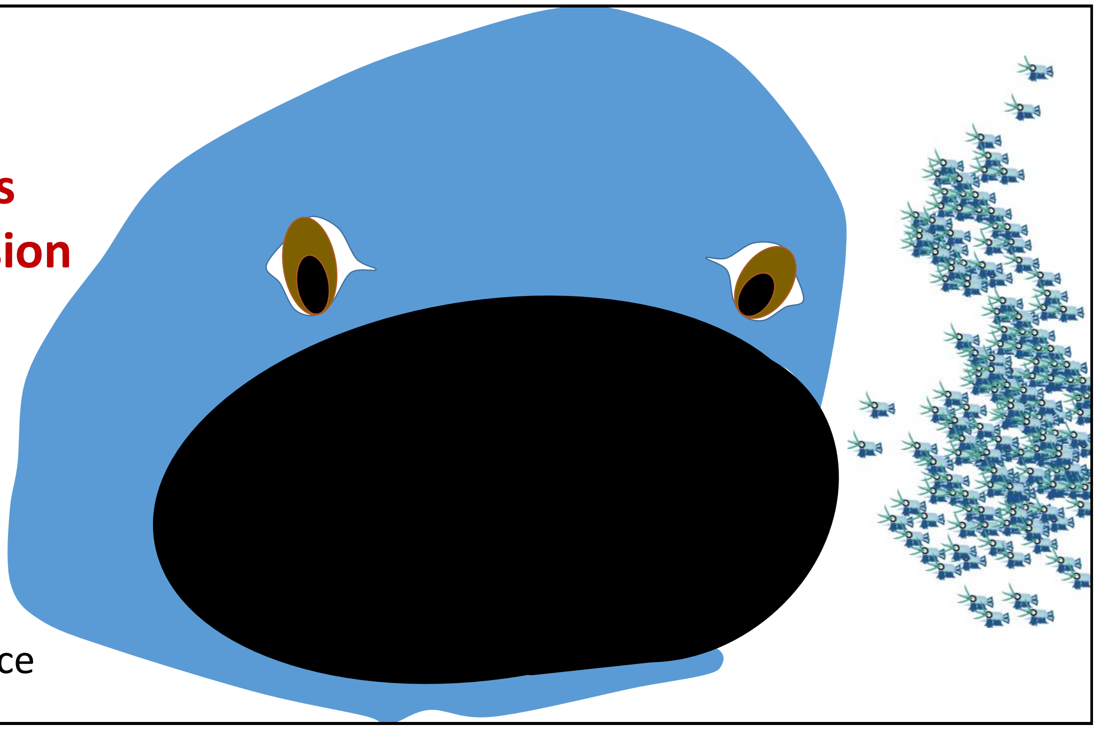
Successive PLS components extract ***covariation*** between X and Y maximally.

Partial least squares regression (PLSR)

We want to ***predict*** multivariate Y from the multivariate X.

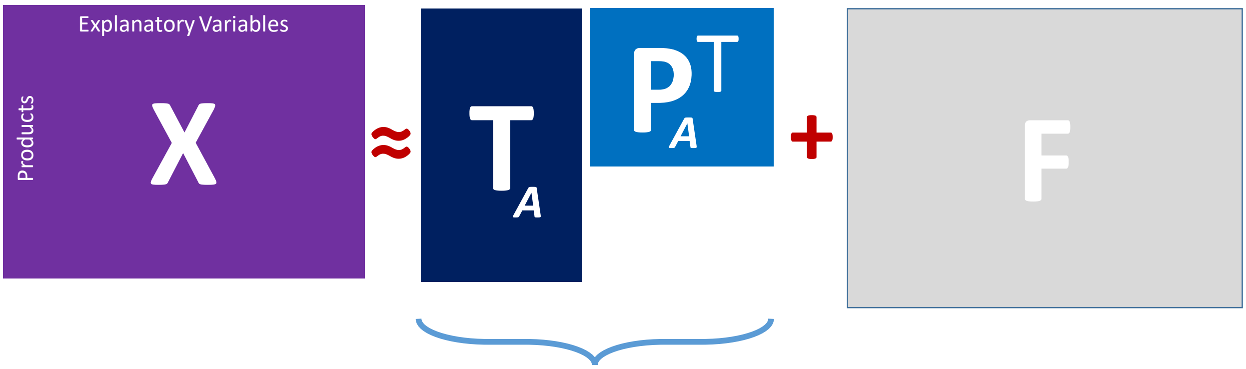
Successive PLS components extract ***covariation*** between X and Y maximally.

Partial least squares regression



Covariance

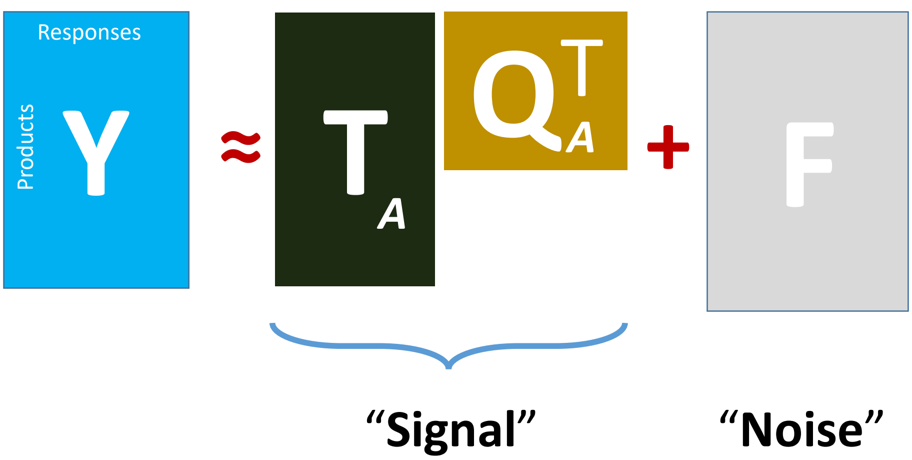
Explaining and predicting data relationships



“Signal” “Noise”

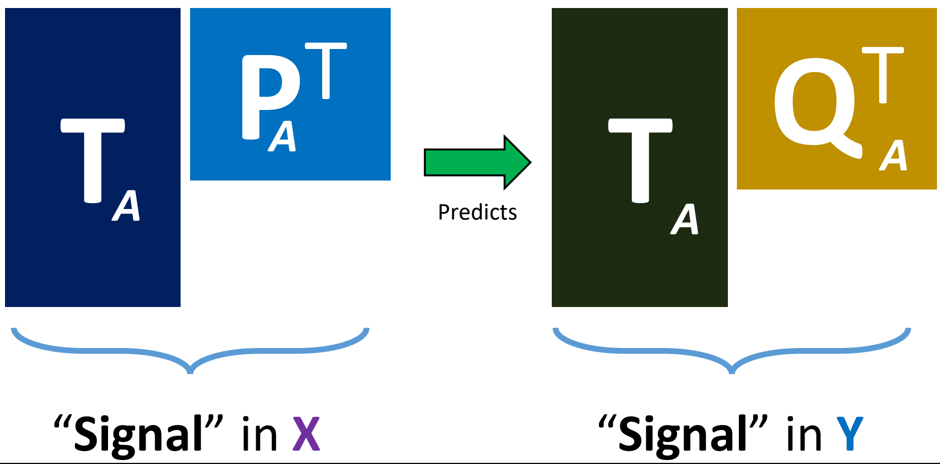
Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In Krishnaiaah, P.R. (ed.). *Multivariate Analysis*.

Explaining and predicting data relationships

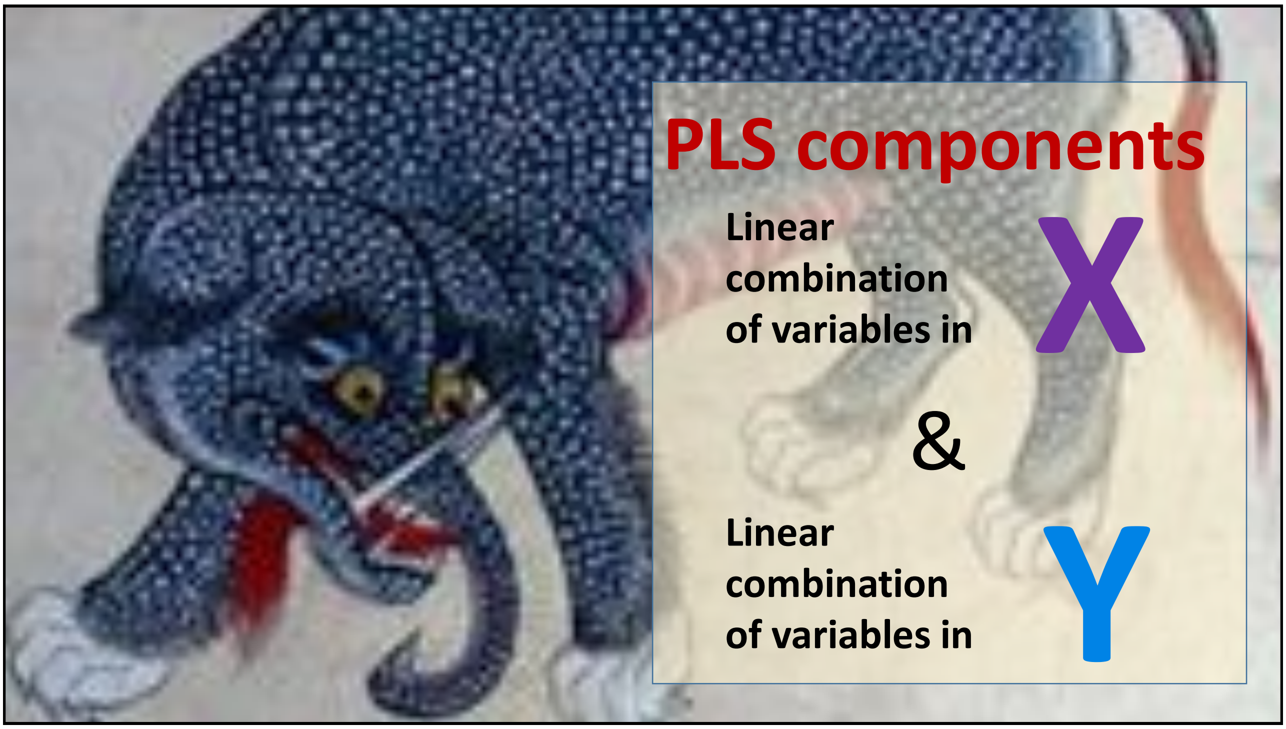


Wold (1966)

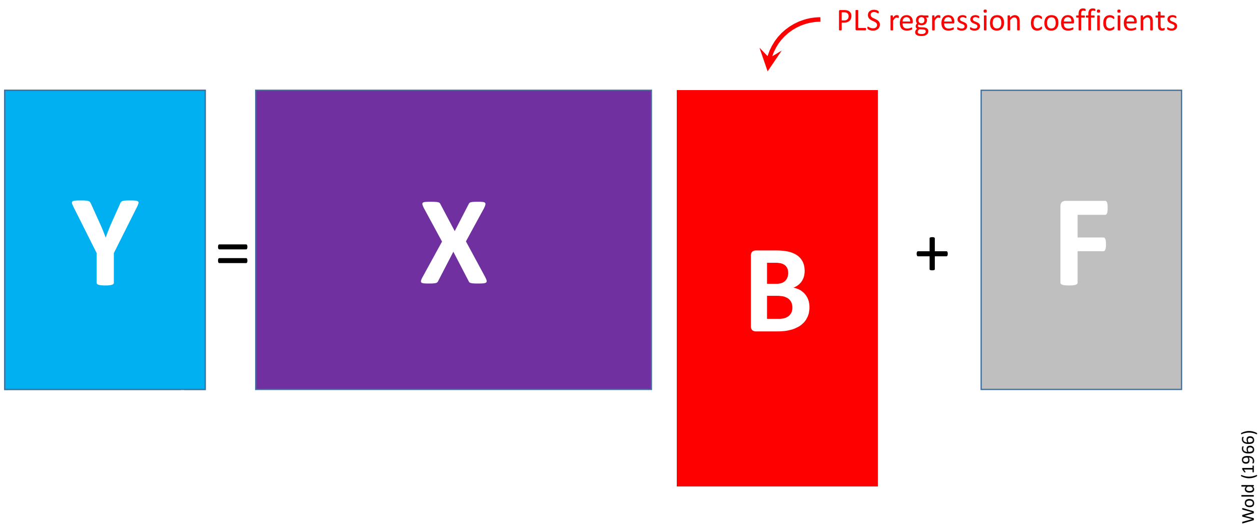
Partial least squares regression (PLSR)



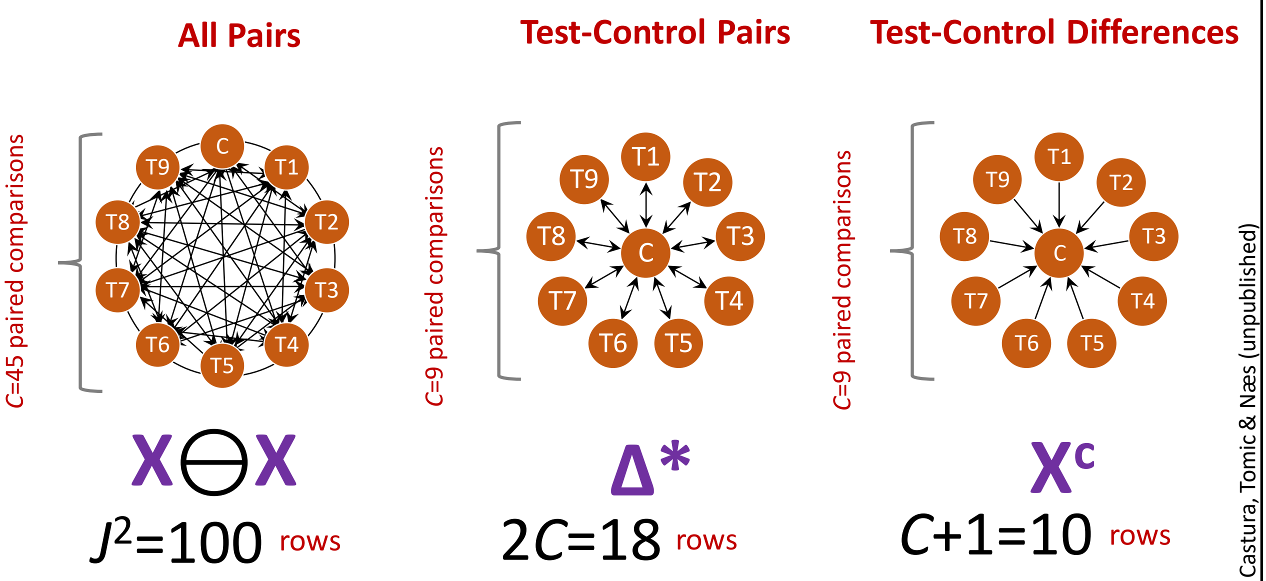
Wold (1966)



Investigating data relationships



Paired comparisons after PLSR



For further information, contact

jcastura@compusense.com

John Castura

Compusense

Acknowledgements

Oliver Tomic

Paula Varela

Tormod Næs

Véronique Cariou

Norwegian University of Life Sciences

Nofima

Oniris

Exploring the relationship between sensory and instrumental data with component-based methods

KoSFoST International Symposium and Annual Meeting 2025

Pioneering Future Connection in FoodTech

Gwangju, Korea · 2-4 July 2025

John Castura

Dr. Philos., M.Sc.

Research Fellow

Compusense