

Microorganisms Identification at Strain-Level

Statistical and Deep Learning methods for analysis of Biotyper MALDI-TOF mass spectra

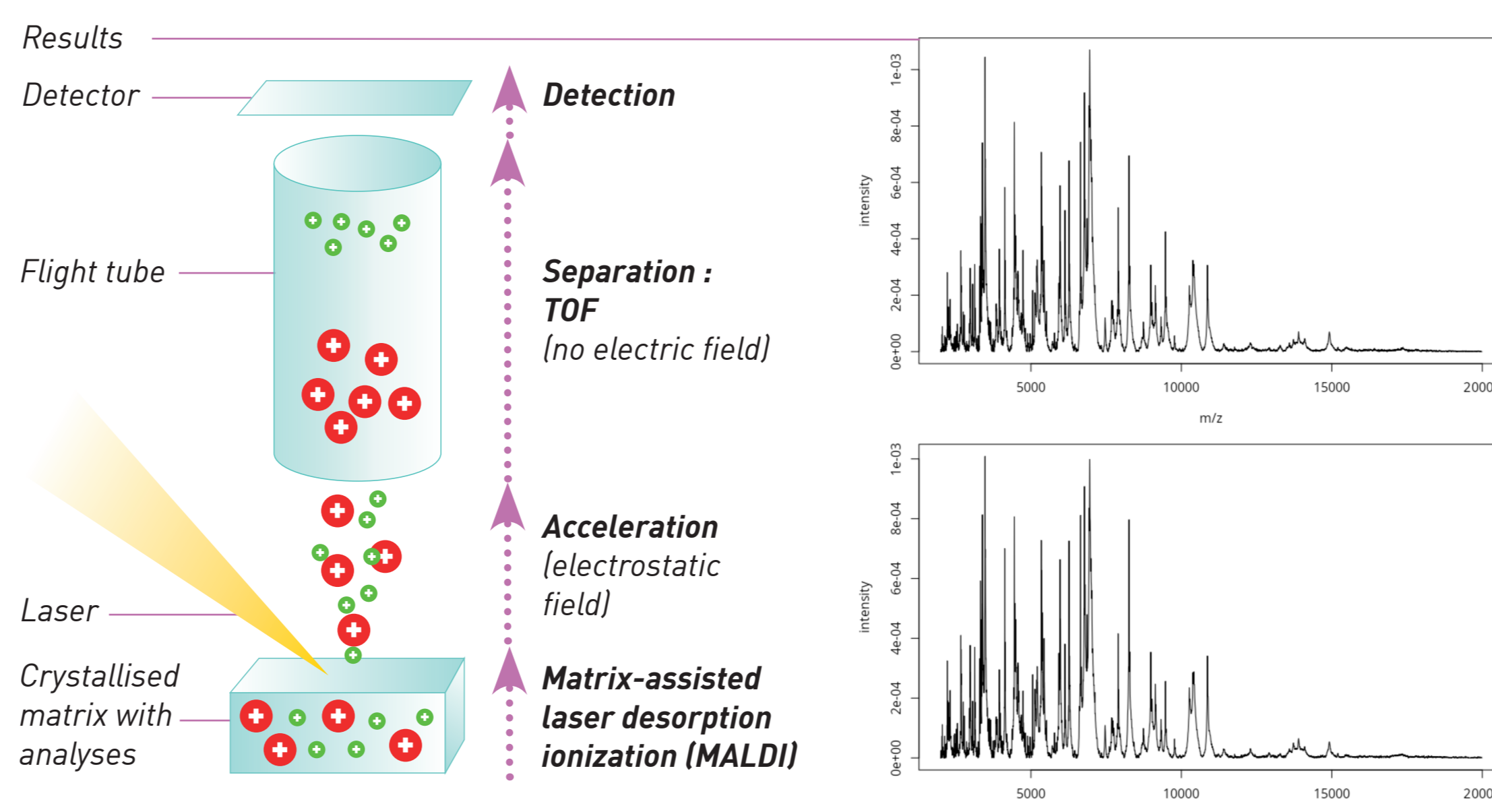
Poster 17

Marie Bourlioux^{1,2}, Paul-Marie Grollemund^{1,2}, Caroline Arous¹, Adrien Geneste¹ and Anna Grizon¹
¹UMR 0545 Fromage, Université Clermont Auvergne, INRAE, VetAgro Sup, Aurillac, France
²LMBP Laboratoire de Mathématiques Blaise Pascal, Campus universitaire des Cèzeaux, Aubière, France

Maldi-tof data and strain identification

MATRIX ASSISTED LASER DESORPTION IONIZATION TIME OF FLIGHT MASS SPECTROMETRY (MALDI-TOF MS)

- Promising alternative to conventional molecular, phenotypic or biochemical microorganism identification methods ^[1]
- Consistent identification at species-level
- Limits: insufficient discriminatory power for robust strain-level identification ^[2]



Example of two spectra representing two different strains from the species *Streptococcus thermophilus* and having similar peaks.

Objectives of the in-depth analysis of spectra

- Identification of microorganisms at strain-level
- Take full advantage of MALDI-TOF mass spectrometry for routine microbial identification

Methodology

METHODS USING FUNCTIONAL DATA STRUCTURE

MAIN STEPS OF THESE METHODS

FPLS-DA

Functional Partial Least Squares - Discriminant Analysis ^[3]

FNN

Functional Neural Networks ^[4]

$$x(t) = \beta_1 \varphi_1(t) + \dots + \beta_d \varphi_d(t)$$

$x(t)$: spectra (functional data)
 β_1, \dots, β_d : basis coefficients
 $\varphi_1(t), \dots, \varphi_d(t)$: basis functions

1

Reduction dimension

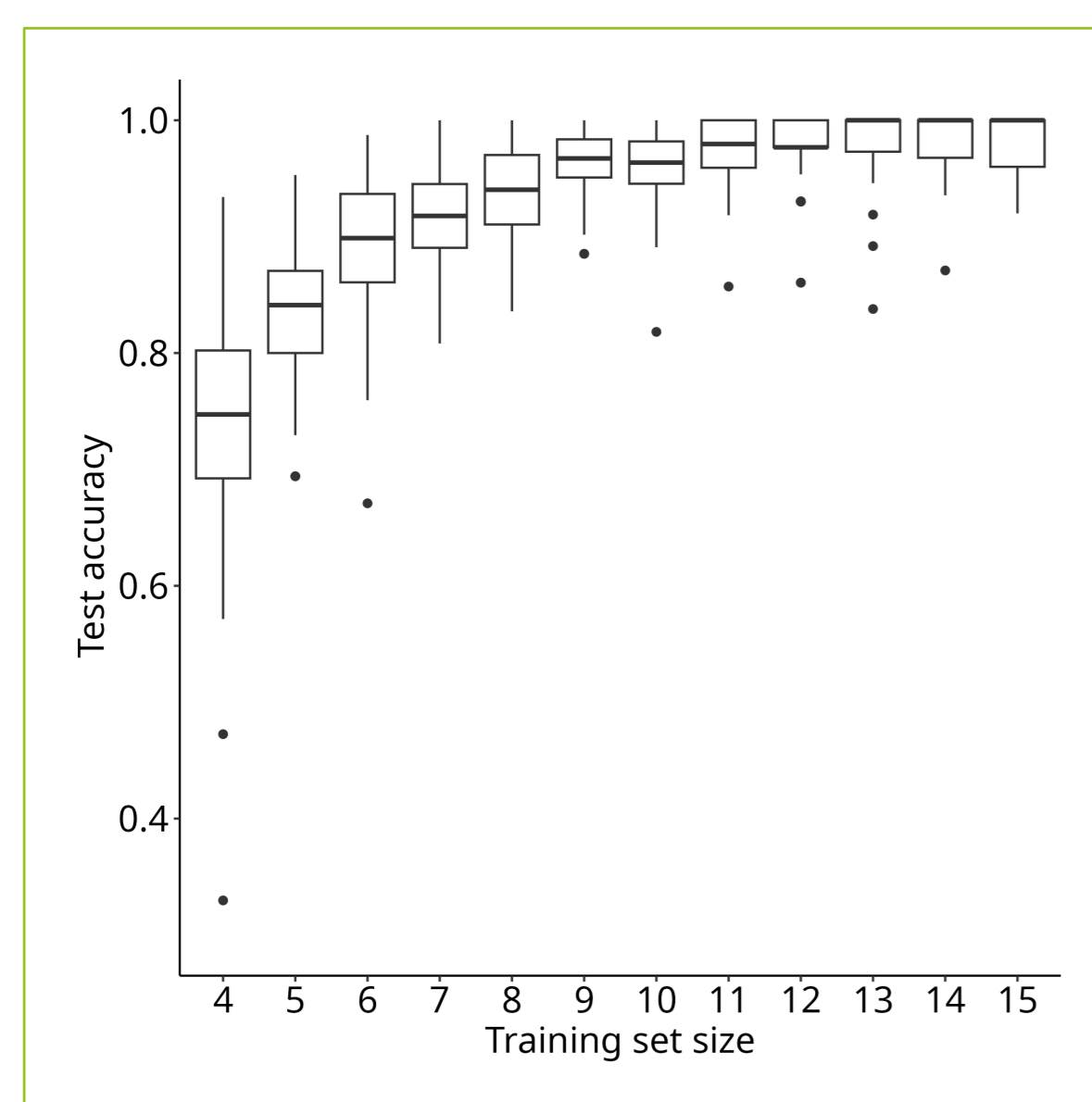
Expression of each spectra as a combination of a small number of basis functions

2

Classification

Attribution of each spectra to a strain thanks to analysis of basis coefficients

Balancing accuracy and data availability



FPLS-DA

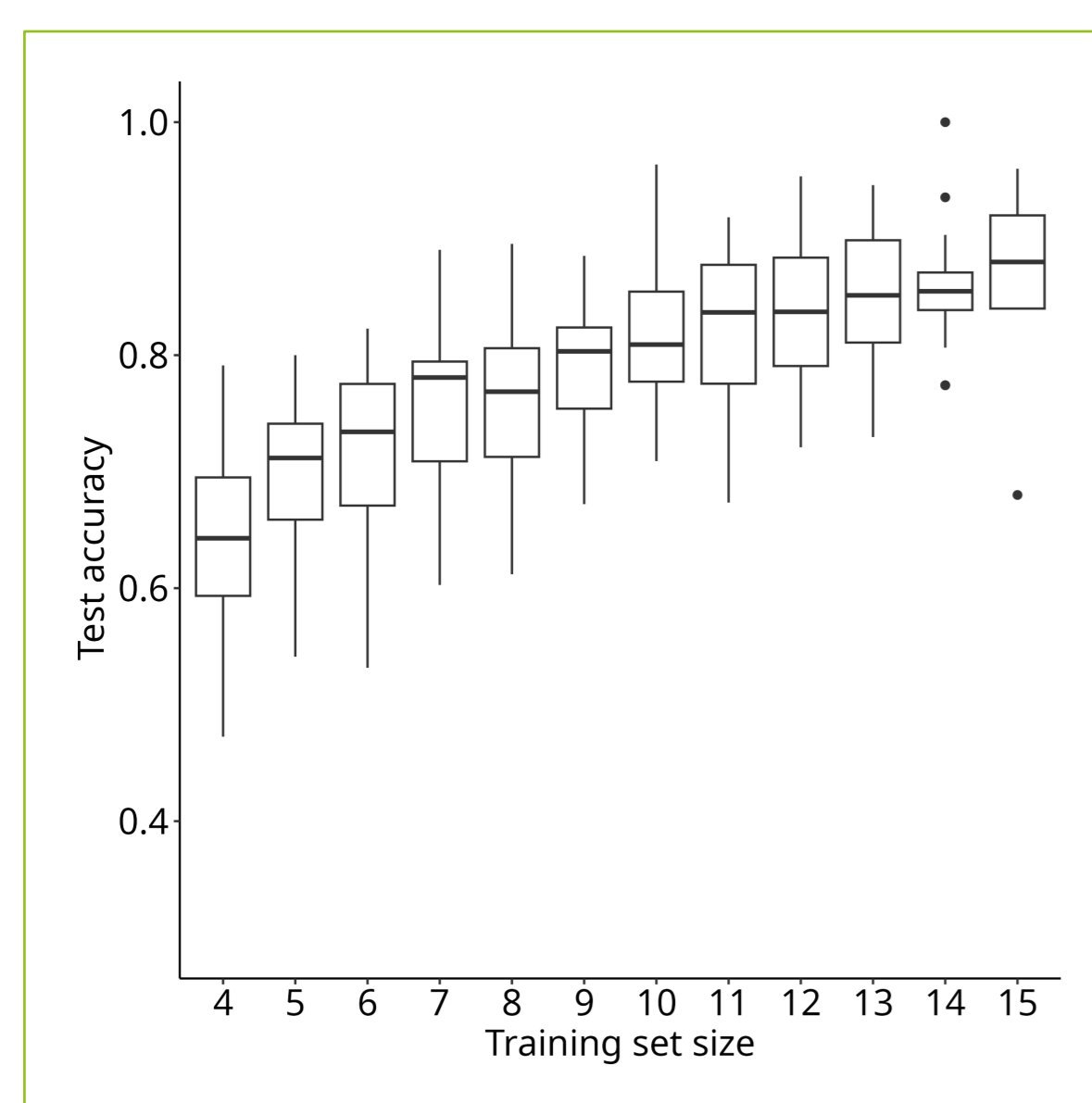
TRADE-OFF BETWEEN MICROORGANISMS IDENTIFICATION PERFORMANCE AND AMOUNT OF DATA

? Problem

High financial and human costs to produce mass spectra

→ Objective

Minimizing the amount of data produced



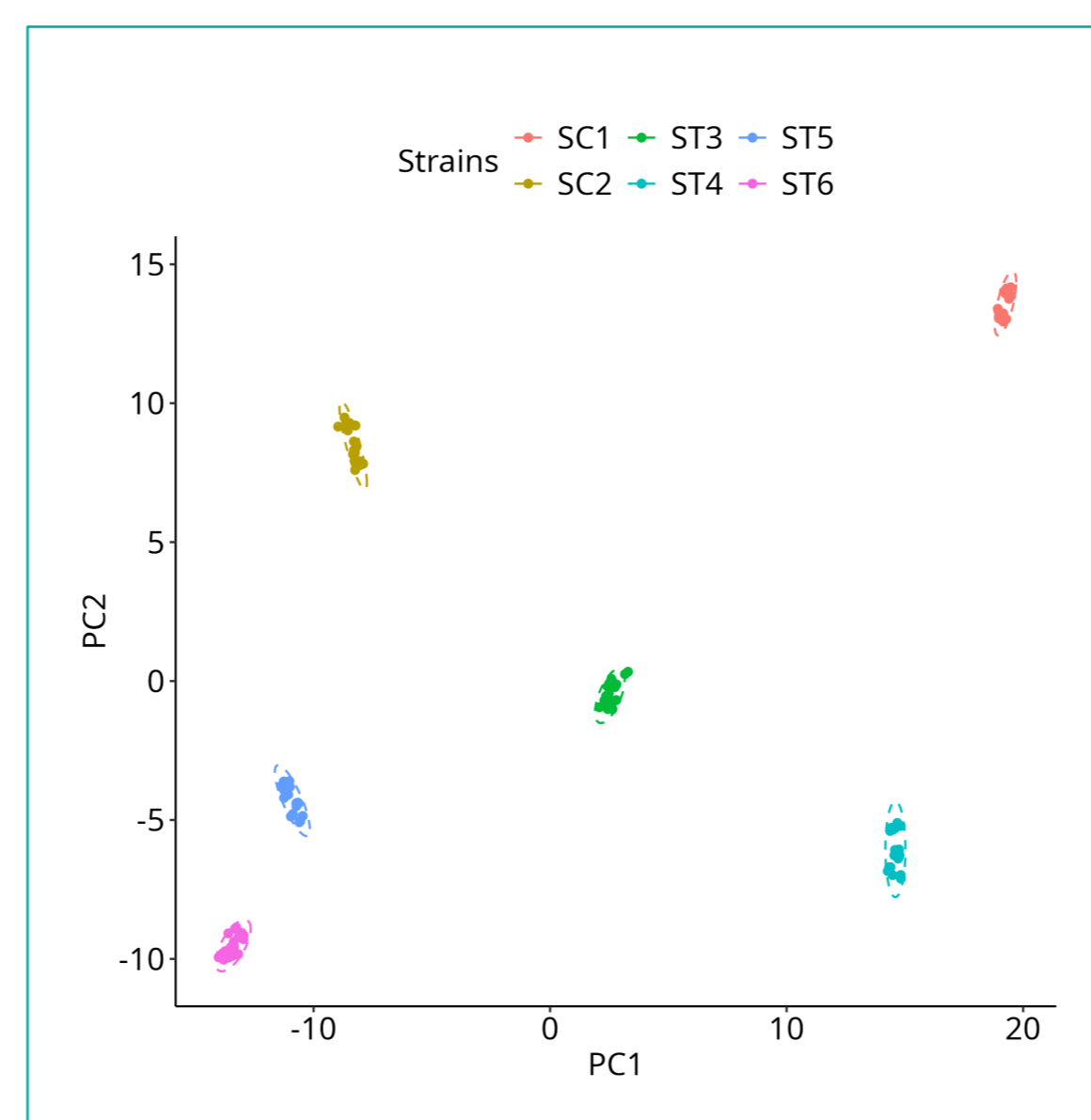
FNN

Models significant results

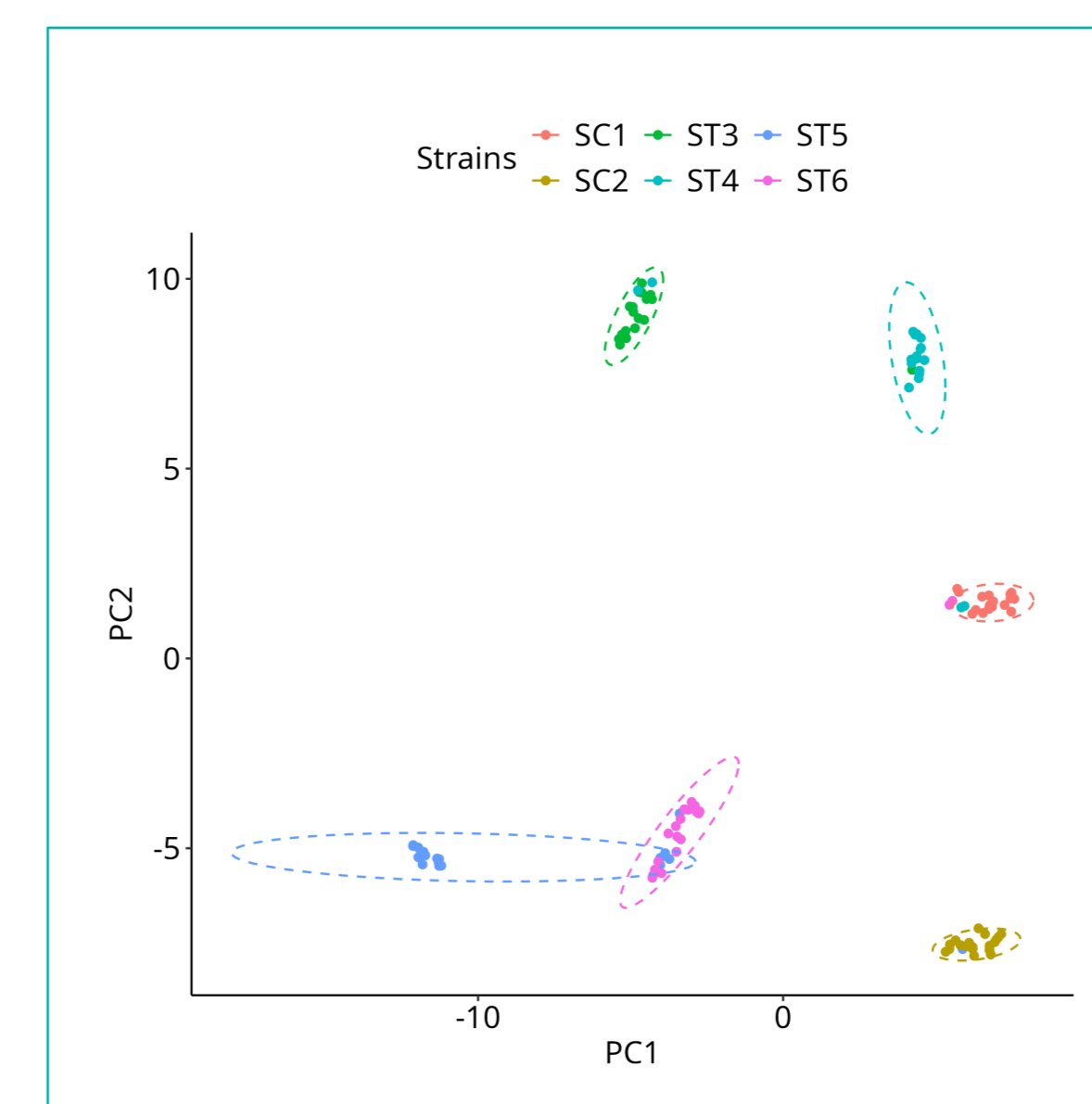
- More replicates needed with FNN than FPLS-DA to get same accuracy
- Accuracy of 90% reached with FPLS-DA applied on training dataset of 7 biological replicates per strain

Identification performances

PERFORMANCES ANALYSED ON SPECTRA OF 6 STRAINS OF STREPTOCOCCUS THERMOPHILUS



FPLS-DA

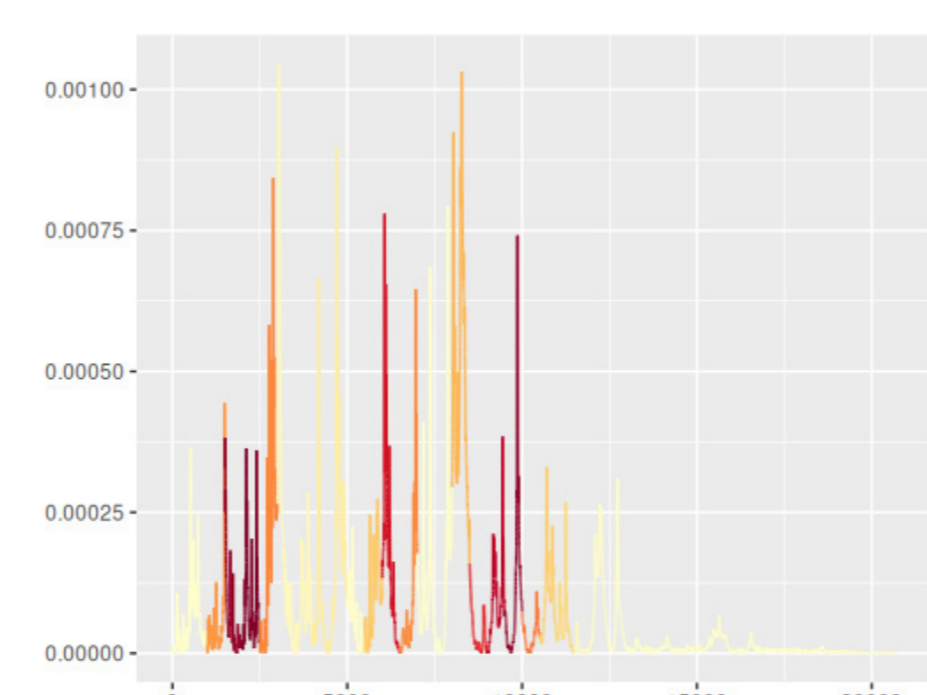


FNN

Observations

- More accurate classification with FPLS-DA rather than FNN
- Visualisation of strains that are more difficult to distinguish for each model

MASS SPECTRA FEATURES INFLUENCING NEURAL NETWORK DECISION



eXplainable Artificial Intelligence (XAI)

- Highlight spectra areas that contribute most to FNN decision making process
- Make understandable results of black-box models (such as FNN)
- Interpretation of microorganism identification

CONCLUSION AND PERSPECTIVES

- Proof of concept: Identification at strain-level within a collection of strains

- Compromise could be reached between limited data acquisition and robust strain-level identification performance

- Development of an automated microorganism-identification workflow
 - Standardized sample processing
 - Reproducible MALDI-TOF spectral preprocessing
 - Functional classification methods

Contact

Paul-Marie Grollemund
 paul_marie.grollemund@uca.fr

Acknowledgments

Research supported by The University Institute of Technology IUT Clermont-Auvergne (<https://iut.uca.fr/>)

REFERENCES

- ^[1]SK Urwyler and J Glaubitz. "Advantage of MALDI-TOF-MS over biochemical-based phenotyping for microbial identification illustrated on industrial applications". In: *Letters in Applied Microbiology* 62.2 (2016), pp. 130-137. doi: <https://doi.org/10.1111/lam.12526>.
- ^[2]Bahareh Mehrmouz et al. "The Use of MALDI-TOF Mass Spectrometry Technology in Molecular Analysis of Microbial Pathogenesis". In: (2022). doi: <https://doi.org/10.33263/BRIAC135.450>.

- ^[3]Cristian Preda, Gilbert Saporta, and Caroline Lévêder. "PLS classification of functional data". In: *Computational Statistics* 22.2 (2007), pp. 223-235.
- ^[4]Florian Heinrichs, Mavin Heim, and Corinna Weber. "Functional Neural Networks: Shift invariant models for functional data with applications to EEG classification". In: *International Conference on Machine Learning*. PMLR, 2023, pp. 12866-12881.



FAMOUS
 Fermentation Microbiologie industrielle
 Université Clermont Auvergne

