Uncorking white wine liking

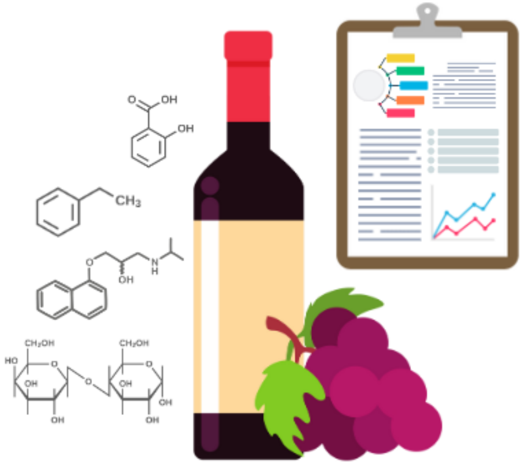
Combining analytical chemistry and chemometrics with crowd-sourced data to predict quality ratings

Frederikke Hjertholm
Industrial PhD
Department of Food Science

UNIVERSITY OF COPENHAGEN

Background

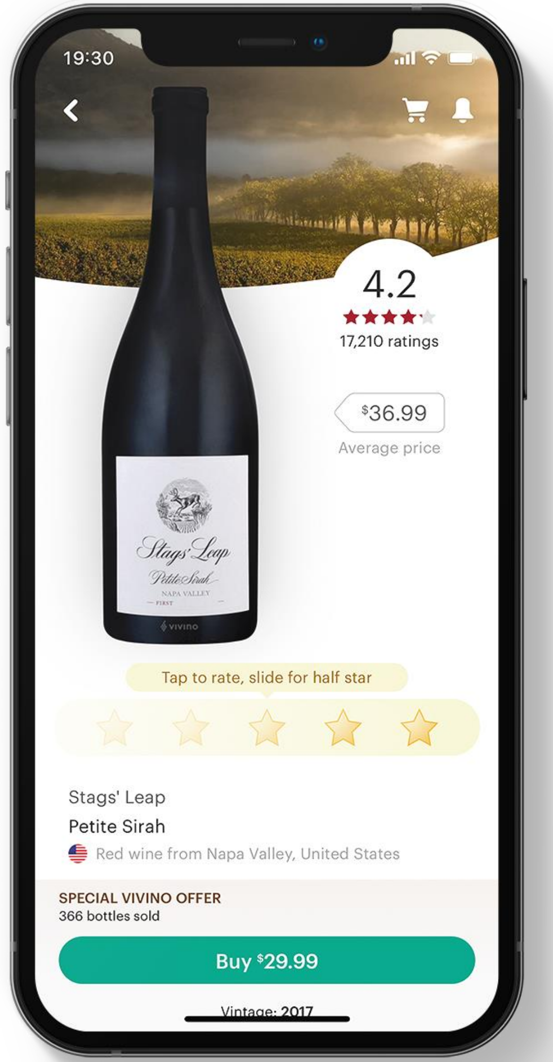
- Wine quality assessment is an important aspect of wine research, particularly regarding consumer acceptance[1,2]
  - Lack of standardization for sensory analysis
  - Consumer incorporate extrinsic (price, presentation, visuals etc.) cues when evaluation hedonic quality
- Vivino: a unique perspective on wine quality
  - Crowd sourced reviews using a 1-5 rating system



1. Sáenz-Navajas, M. P., Ballester, J., Pêcher, C., Peyron, D., & Valentin, D. (2013). Sensory drivers of intrinsic quality of red wines. Effect of culture and level of expertise. *Food Research International*, *54*(2), 1506–1518. https://doi.org/10.1016/j.foodres.2013.09.048

2. Tiwari, P., Bhardwaj, P., Somin, S., Parr, W. V., Harrison, R., & Kulasiri, D. (2022). Understanding Quality of Pinot Noir Wine: Can Modelling and Machine Learning Pave the Way? *Foods*, *11*(19). https://doi.org/10.3390/foods11193072

vivino



2.8 B
Scanned Labels

300 M
Ratings

70 M
Users

18 M
Wines

Methods: Samples

German white wine (n=89)

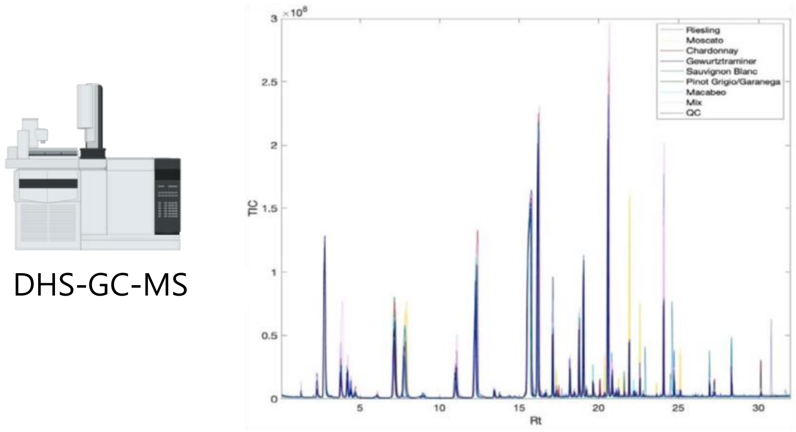
Grape varieties
Riesling (34), Pinot Blanc (9), Pinot Gris (8), Chardonnay (6), Sylvaner (5), Sauvignon Blanc (5), Scheurebe (2), Bacchus (1), Cabernet Blanc (1), Elbling (1), Gewürztraminer (1), Muskateller (1), Goldmuskateller (1), Gutedel (1), Müller-Thurgau (1), Sauvignac (1), Traminer (1) and Cuvées (10)

Region
Pfalz (58), Rheinhessen (9), Mosel (6), Baden (5), Franken (3), Rheingau (3), Württemberg (2), Saar (1), Nahe (1) and Mittelrhein (1)

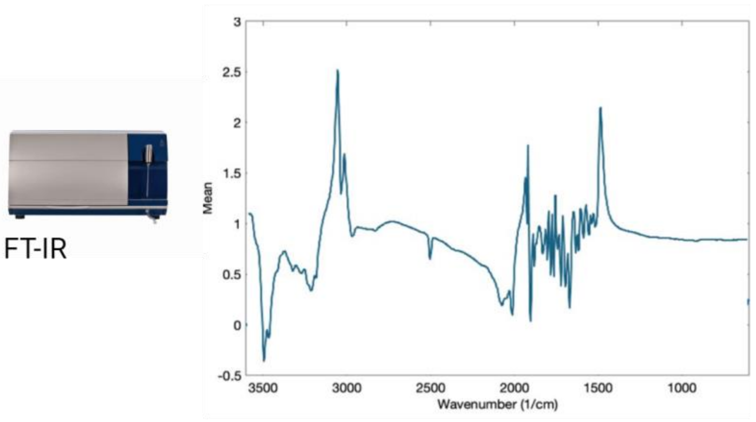
Vintage
2020 (42), 2021 (14), 2019 (13), 2017 (9), 2018 (7), 2016 (3) and 2014 (1)

Obtained from Badische Anillin & Soda Fabrik (BASF, Ludwigshafen, Germany)

Methods: Data Collection



PARAFAC2-based deconvolution and identification →



PLS regression calibration model →

Volatile Organic Compounds

Ethanol
Tert-butanol
1-Hexanol
cis-3-Hexenol
2,3-Butanediol
Acetaldehyde
Benzaldehyde
2,6-Dimethyl-5-heptenal
Hexanoic acid
Octanoic acid
Isobutyl isovalerate
Butyl acetate
Amyl acetate
Ethyl Acetate
Hexyl acetate
Ethyl nonanoate
Isoamyl caproate
2-methyl-1,3-dioxane
Styrene
1,1,6-trimethyl-2H-naphthalene
Thujene
Limonene
Myrcene
β-Cedrene
.....

Volatile Organic Compounds

$$
\begin{array}{l}
 \\
\text{Wine 1} \\ \text{Wine 2} \\ \text{Wine 3} \\ \text{Wine 4} \\ \text{Wine n}
\end{array}
\begin{array}{c}
\begin{array}{cccc} \text{VOC1} & \text{VOC2} & \text{VOC3} & \text{VOCn} \end{array} \\
\begin{bmatrix}
X_{11} & X_{21} & X_{31} & X_{n1} \\
X_{12} & X_{22} & X_{32} & X_{n2} \\
X_{13} & X_{23} & X_{33} & X_{n3} \\
X_{14} & X_{24} & X_{34} & X_{n4} \\
X_{1n} & X_{2n} & X_{3n} & X_{nn}
\end{bmatrix}
\end{array}
$$

89 x 145

Chemical parameter

$$
\begin{array}{l}
 \\
\text{Wine 1} \\ \text{Wine 2} \\ \text{Wine 3} \\ \text{Wine 4} \\ \text{Wine n}
\end{array}
\begin{array}{c}
\begin{array}{cccc} \text{C1} & \text{C2} & \text{C3} & \text{Cn} \end{array} \\
\begin{bmatrix}
X_{11} & X_{21} & X_{31} & X_{n1} \\
X_{12} & X_{22} & X_{32} & X_{n2} \\
X_{13} & X_{23} & X_{33} & X_{n3} \\
X_{14} & X_{24} & X_{34} & X_{n4} \\
X_{1n} & X_{2n} & X_{3n} & X_{nn}
\end{bmatrix}
\end{array}
$$

89 x 18

Chemical parameters

Ethanol
G
T
Volatile Acidity
pH
Malic Acid
Lactic Acid
Sorbic Acid
Tartaric Acid
Citric Acid
CO2
Absorbance A420 nm
Absorbance 520 nm
Absorbance A620
Total Polyphenols
Glycerol
Reducing Sugar
Density

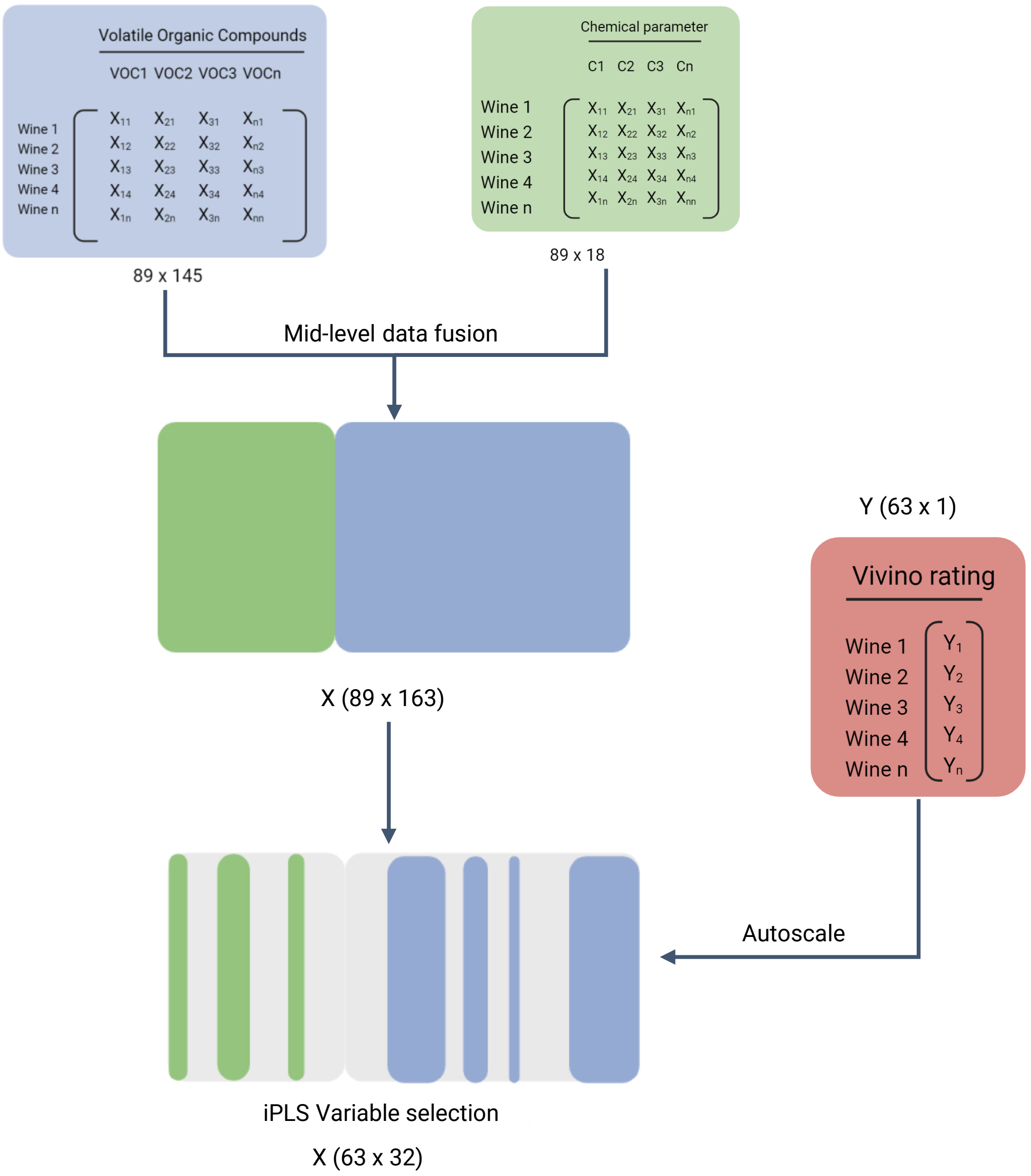
VIVINO

Vivino rating

$$
\begin{array}{l}
\text{Wine 1} \\ \text{Wine 2} \\ \text{Wine 3} \\ \text{Wine 4} \\ \text{Wine n}
\end{array}
\begin{bmatrix}
Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_n
\end{bmatrix}
$$

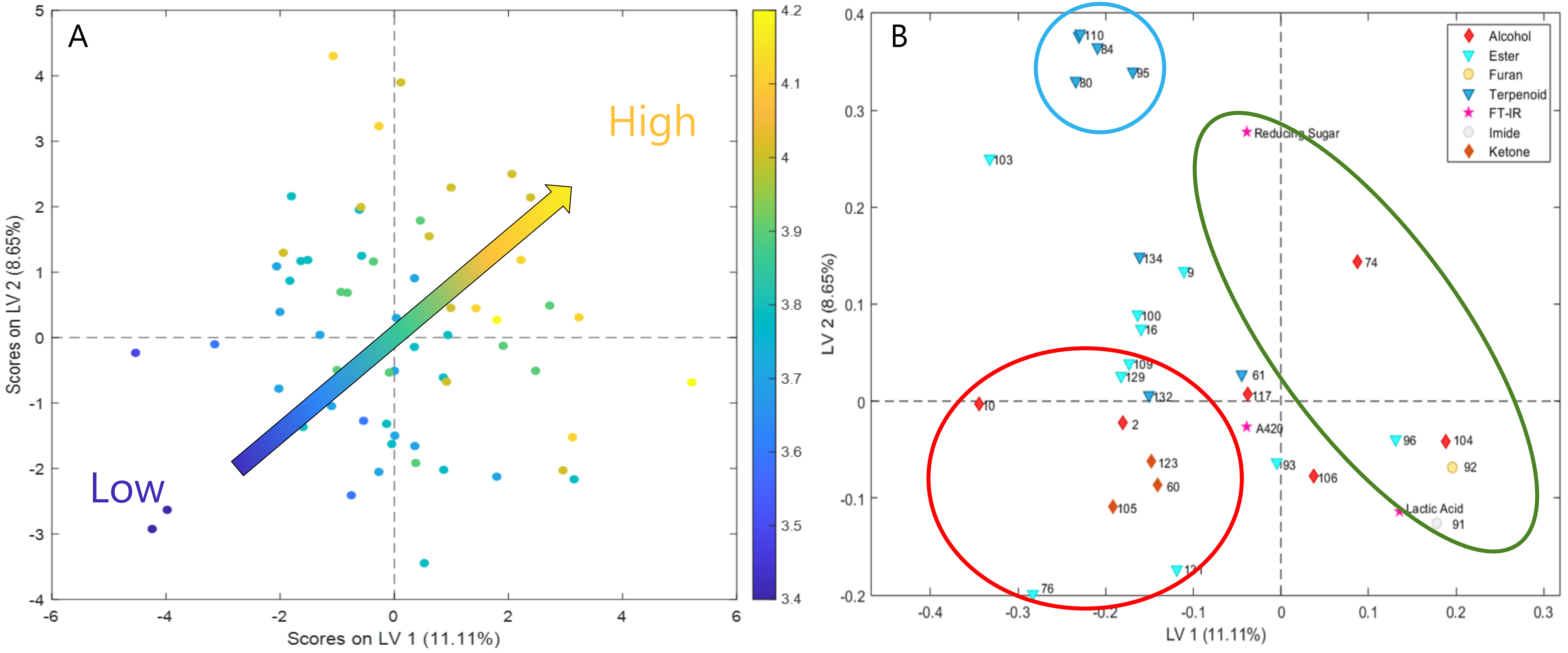
Model Development

- Merge data blocks
  - Combined calibration matrix
- Auto scale prediction matrix
  - Range from 3.1-4.2
- Perform variable selection
  - RMSECV as selection criteria
  - Contiguous block CV
- Build model!



PLS model

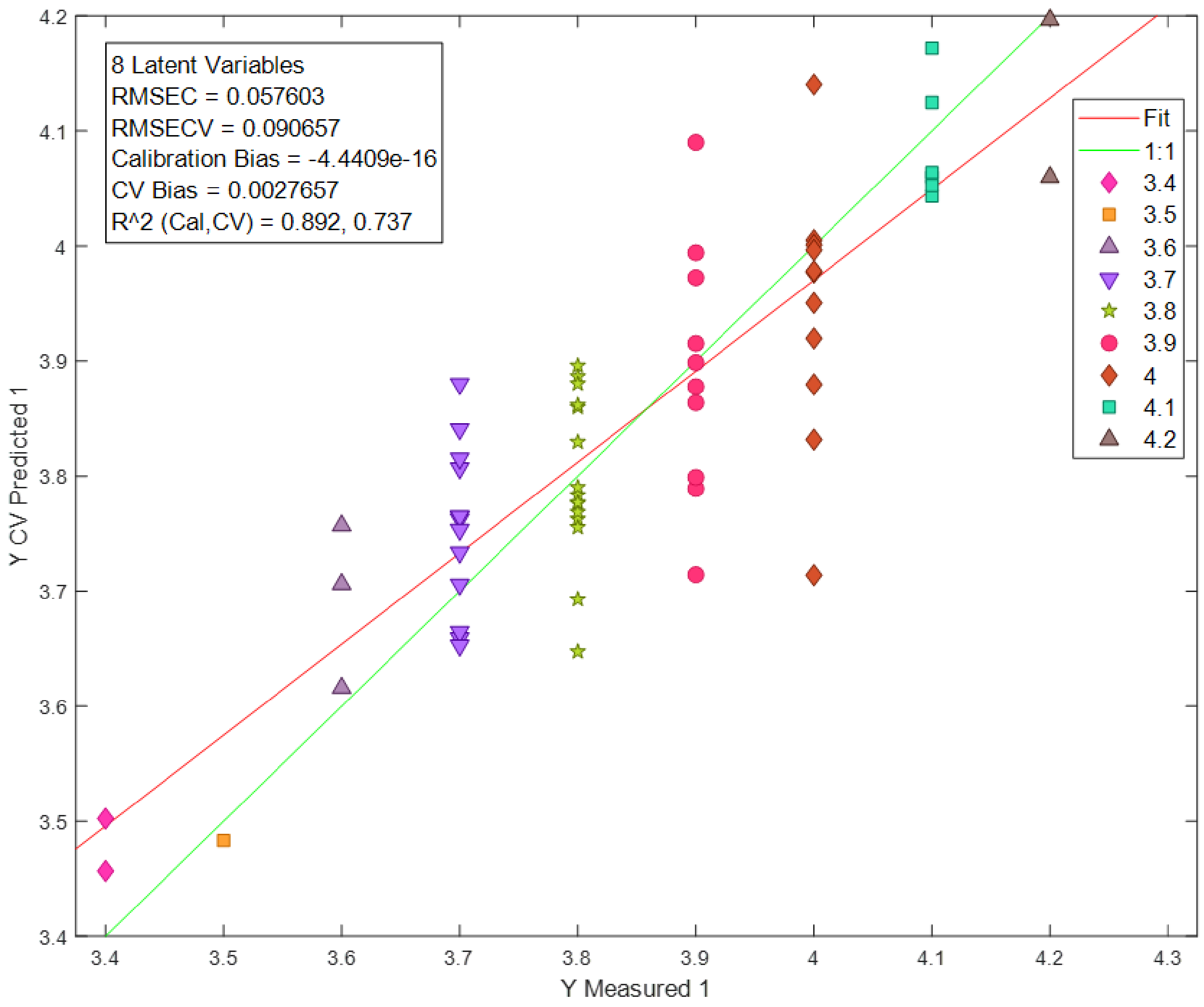
PLS model of 8 Latent variables



Scores coloured according to observed Vivino ratings

Predicting Vivino ratings

- Prediction range: 3.4 – 4.2
  - Corresponds with observations
- R^2 (CV) = 0.74
  - Good correlation between chemistry and rating
- RSMECV = 0.09
  - Lowest possible RMSE value considering contribution from ratings and model



Conclusion & Limitations

- A Partial Least Squares (PLS) model can predict Vivino quality ratings using 32 chemical variables
- Chemical properties are correlated with consumer perceived quality ratings

Limitations

- Limited prediction range
- Include validation set
- Missing statistic information on extracted ratings
  - Standard deviation and mean

Acknowledgement

Rabea Goetz

Paul-Albert Anselm Schneide

Rasmus Bro

Mikael Agerlin Petersen

Beatriz Quintanilla Casas

Funding

Industrial PhD research grant from the Innovation Fund, Denmark.

Danish Data Science Academy Postdoctoral fellowship, funded by the Novo Nordisk Foundation (NNF21SA0069429)

Thank you to Badische Anilin- und Soda-Fabrik (BASF) for providing wine samples.

EINAR WILLUMSEN When taste and speed matter

UNIVERSITY OF COPENHAGEN

Thank you for your attention!

Frederikke Hjertholm
Industrial PhD
fni@ewflavours.com

EINAR WILLUMSEN
When taste and speed matter

UNIVERSITY OF COPENHAGEN