Spectroscopic Selectivity of Multivariate Analysis Techniques

Sander Vervoort [1, 2] and Marcus Wolff [1]

[1]Hamburg University of Applied Sciences, Heinrich Blasius Institute for Physical Technologies, Berliner Tor 21, 20099 Hamburg, Germany
[2]University of the West of Scotland, School of Computing, Engineering and Physical Sciences, PA1 2BE Paisley, Scotland, United Kingdom

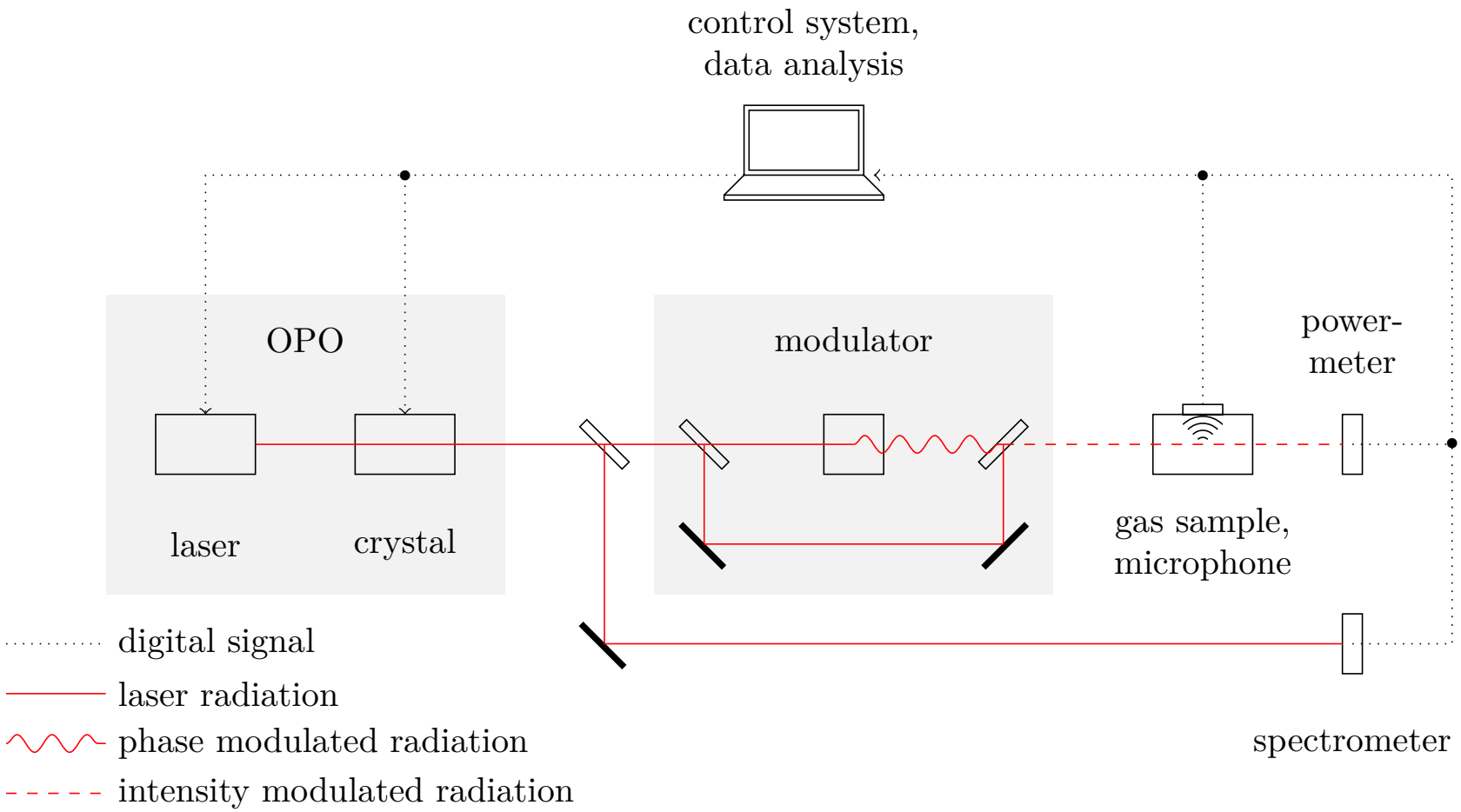
Motivation

- Determination of the individual concentrations of volatile organic compounds (VOCs) in gas mixtures
  - 2-Butanone (C_4H_8O)
  - 1-Propanol (C_3H_8O)
  - Isoprene (C_5H_8)
  - Ethylbenzene (C_8H_{10})
  - Styrene (C_8H_8)
  - Hexanal ($C_6H_{12}O$)
- Analysis of photoacoustic spectra
  - Mid-IR wavelength region between 3.2 µm and 3.5 µm
  - Signal proportional to concentration for trace gases
- Multivariate analysis techniques
  - Regression
  - Feature selection
  - Feature projection

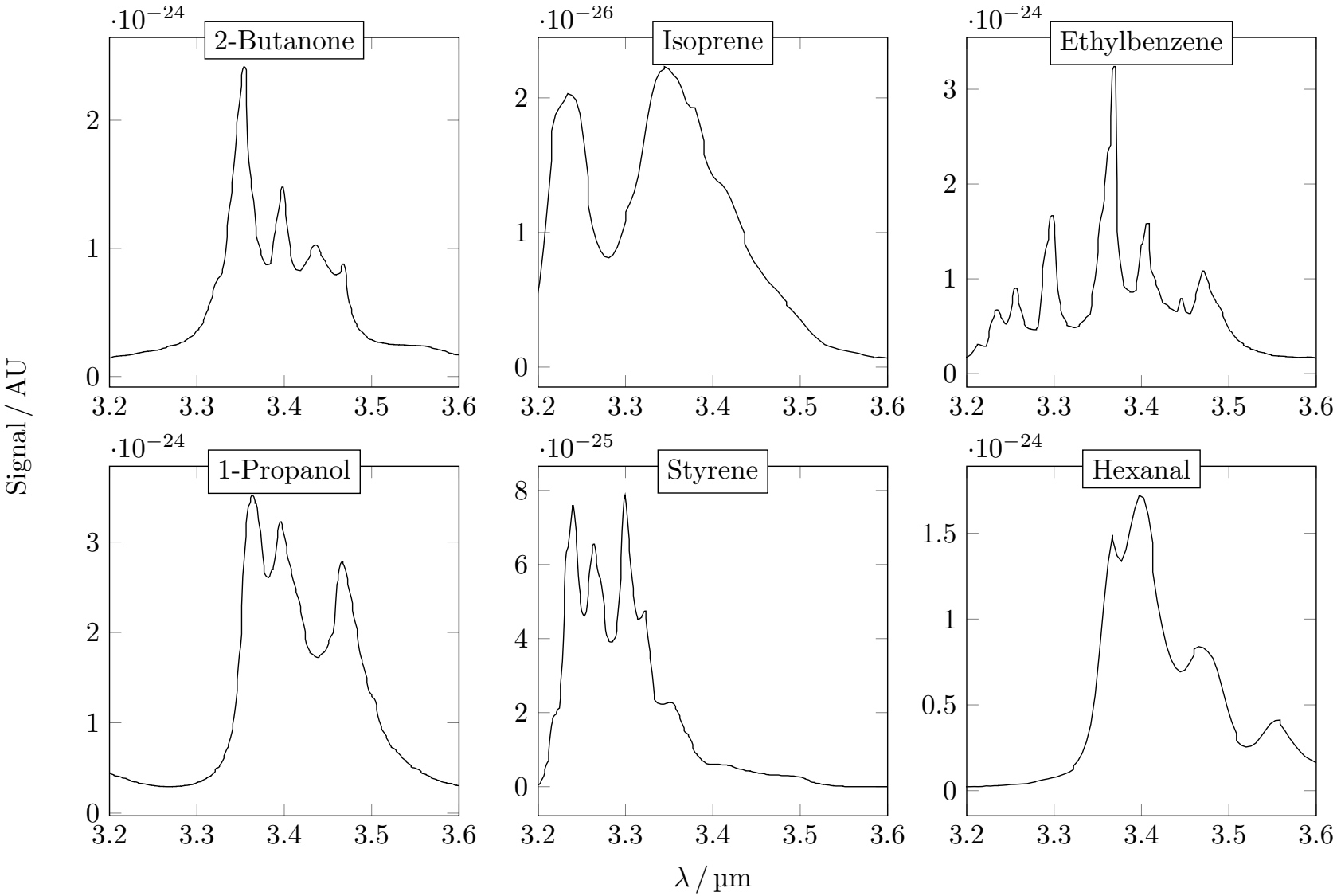
Experiment

- Optical parametric oscillator as radiation source 3.2 µm to 3.5 µm
- Electro-optic modulator for frequency stability
- Modulation frequency of 2.81 kHz
- Digital mems microphone for low signal to noise ratio

Experiment

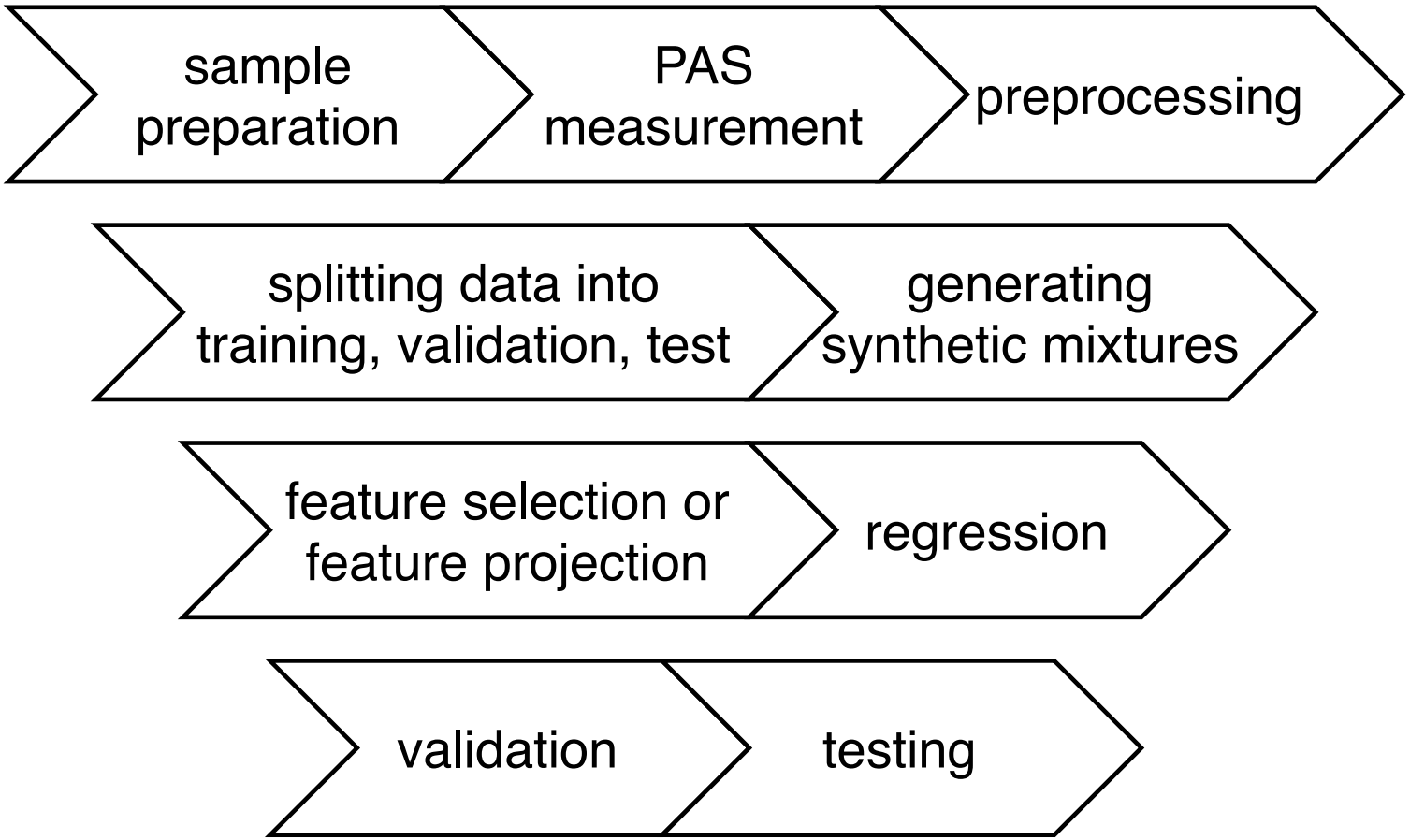

control system,
data analysis
OPO
laser
crystal
modulator
power-
meter
gas sample,
microphone
spectrometer
digital signal
laser radiation
phase modulated radiation
intensity modulated radiation

Photoacoustic spectra of single VOCs



Multivariate Analysis

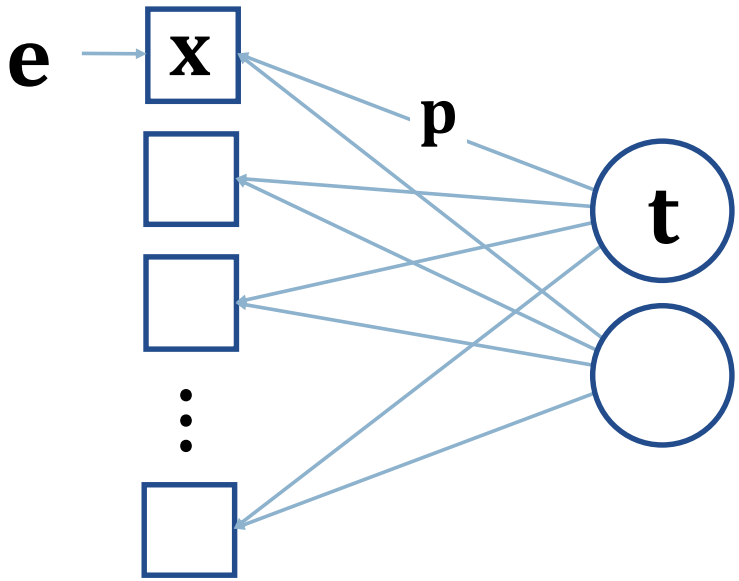
Procedure from data acquisition to multivariate data analysis and testing of gas samples



Dimensionality reduction by feature projection

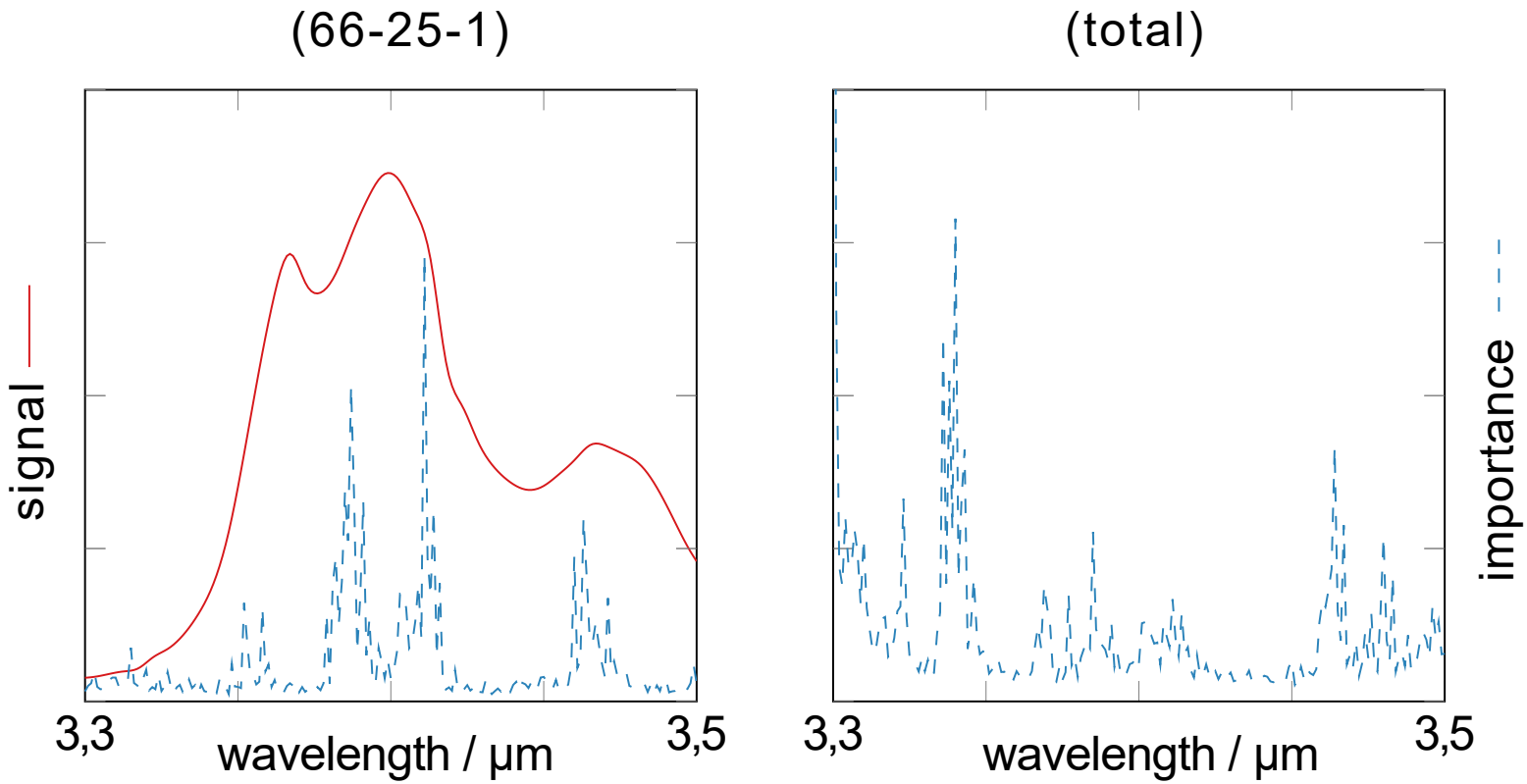
Principal component analysis (PCA) and partial least squares regression (PLSR)

- Features are decomposed into linear combinations $\mathbf{x} = \mathbf{t}\mathbf{p}^{\mathsf{T}} + \mathbf{e}$
- Scores ($\mathbf{t}$) serve as new features



Dimensionality reduction by feature selection

- Photoacoustic spectra of Hexanal CAS 66-25-1 solid red
- Feature importance by random forest algorithm (RFA) dashed blue



Accuracy of multivariate models

- Feature selection by RFA does not decrease accuracy
- Feature projection does increase accuracy
- PLSR with highest accuracy

	MAE / ppm	s / ppm
MLR	6.8	9.2
RFA + MLR	6.8	9.2
PCA + MLR	5.9	7.9
PLSR	5.8	7.8

Table: Error of multivariate regression models

Outlook

- Improving the evaluation of the multivariate models by cross-validation and additional test data
- Investigation of the influence of further VOCs on the accuracy
- Further investigations on the feature selection

References

- *Laserspectroscopy*
  Demtröder, W. Laserspektroskopie: Grundlagen und Techniken; 4th ed.; Springer-Verlag: Berlin Heidelberg, 2000; ISBN 978-3-662-08266-9.
- *Multivariate analysis techniques*
  Kessler, W. Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik: ein Lehrbuch; 1st ed.; WILEY-VCH: Weinheim, 2008; ISBN 978-3-527-31262-7.
- *Setup and prior study*
  Saalberg, Y.; Wolff, M. Multivariate Analysis as a Tool to Identify Concentrations from Strongly Overlapping Gas Spectra. Sensors 2018, 18, 1562, doi:10.3390/s18051562.