Detection of chemical substances from colorimetric sensor data using probabilistic machine learning

CRIMTRACK

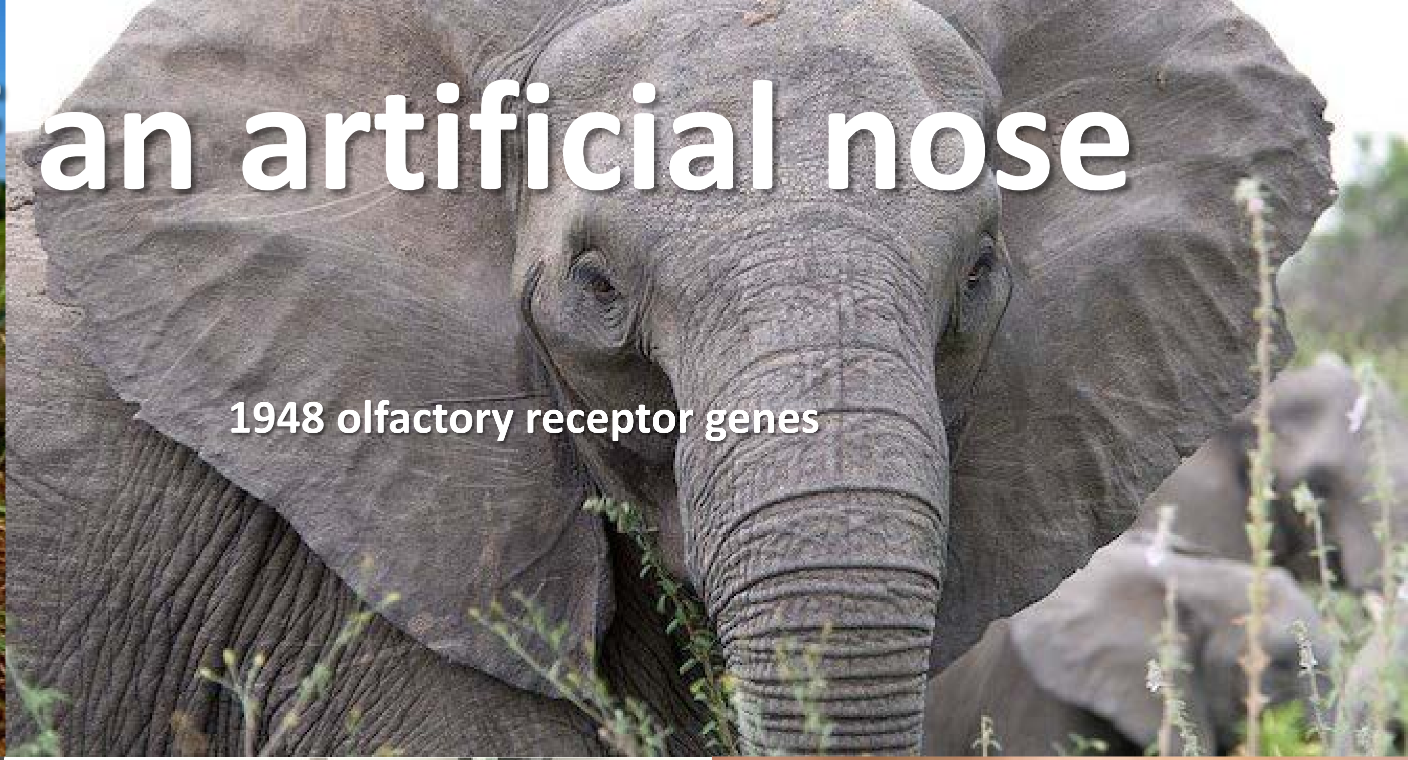
Lasse L. Mølgaard, Ole T. Buus, **Jan Larsen**

Hamid Babamoradi, Ida L. Thygesen, Milan Laustsen, Jens Kristian Munk, Eleftheria Dossi, Caroline O'Keeffe, Lina Lässig, Sol Tatlow, Lars Sandström, and Mogens H. Jakobsen

Section for Cognitive Systems, DTU Compute, Richard Petersens Plads, Building 324, 2800 Kongens Lyngby, Denmark

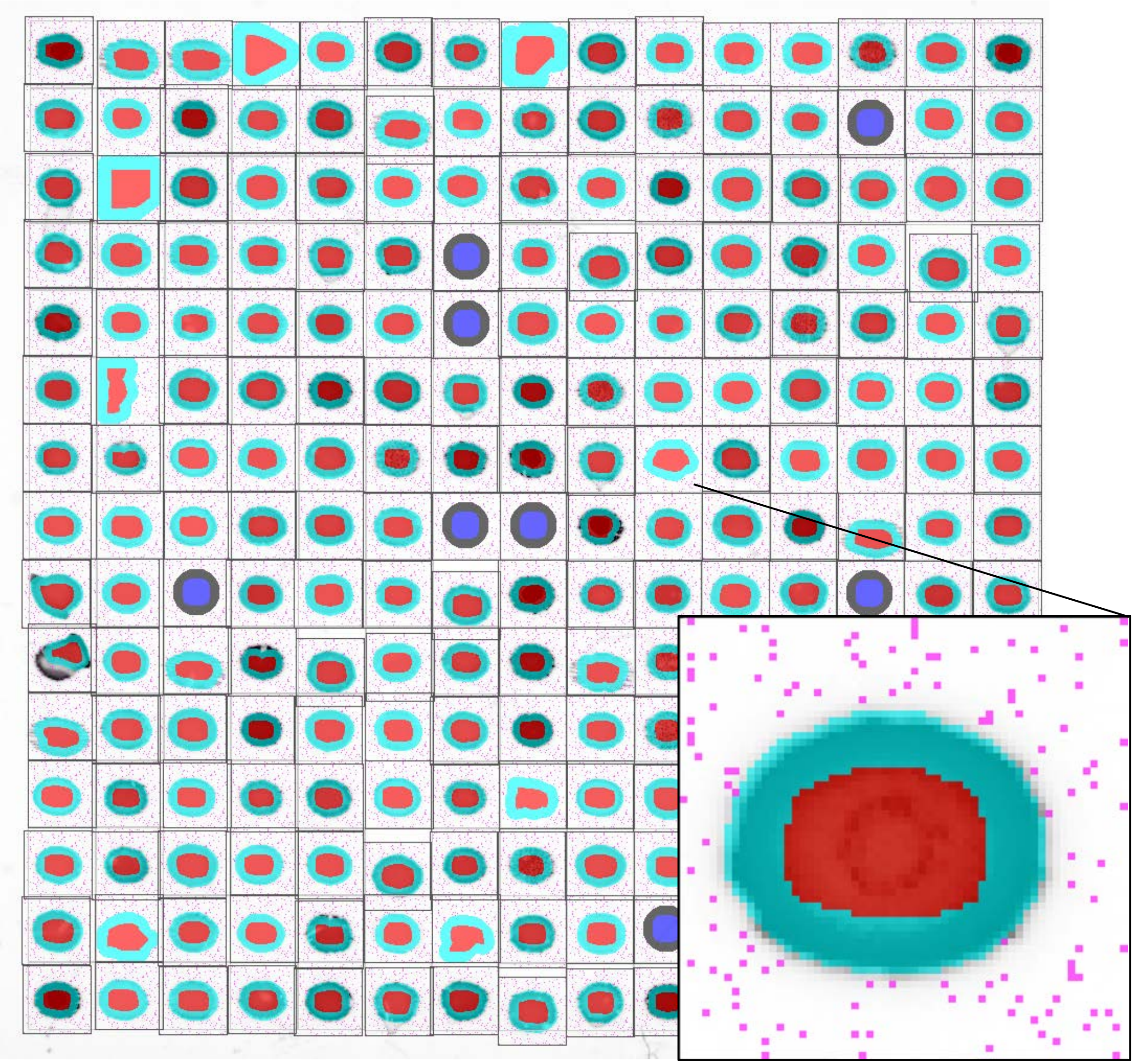
DREAM

The dream of an artificial nose



PRINCIPLE

Data conditioning and preprocessing – median RGB values for each dye

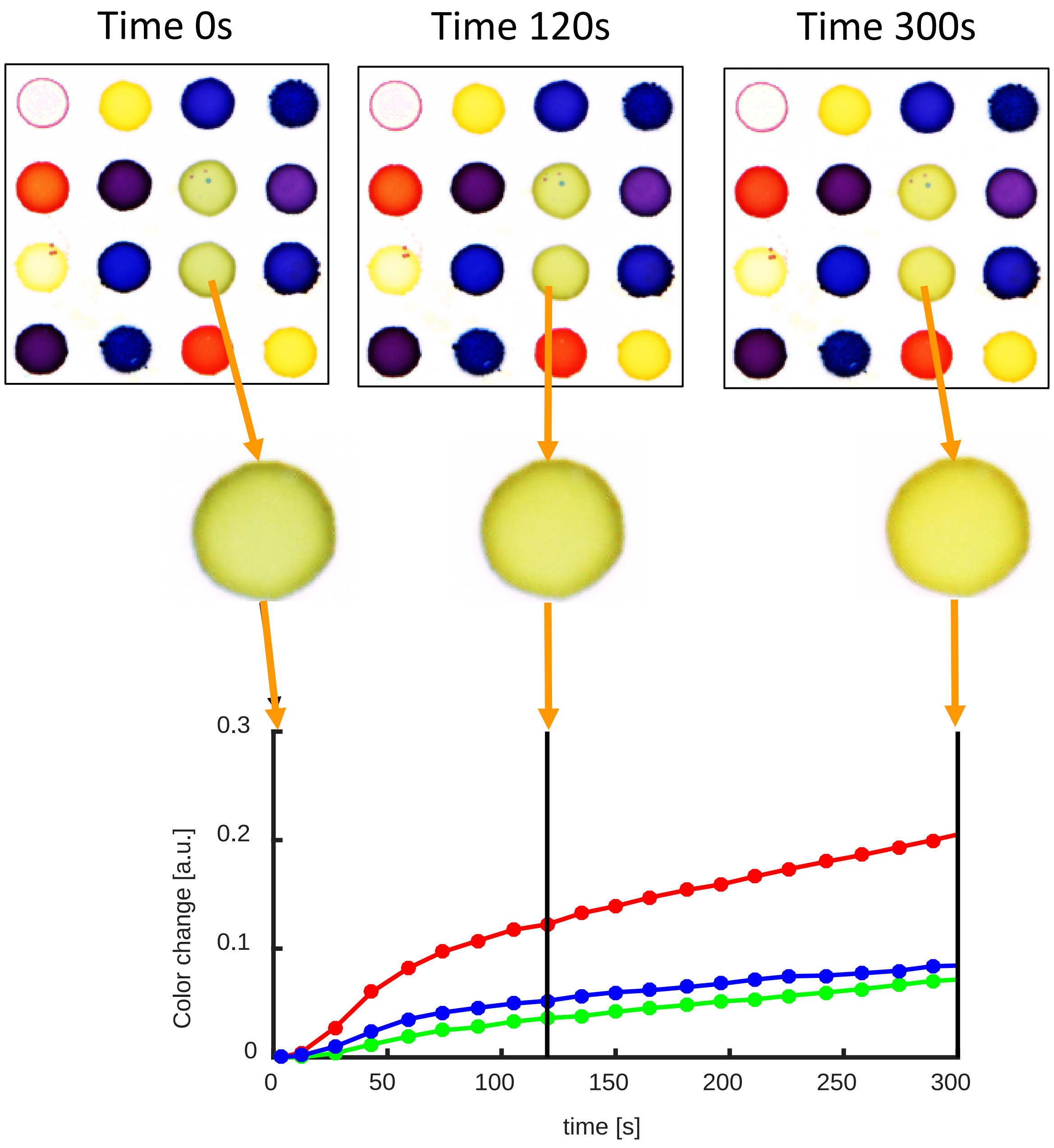


Extraction of color change

Dye spots are detected automatically.

RGB color changes are summarized as the relative color change to the pre-image at 0s.

The changes are small and requires sophisticated analysis.

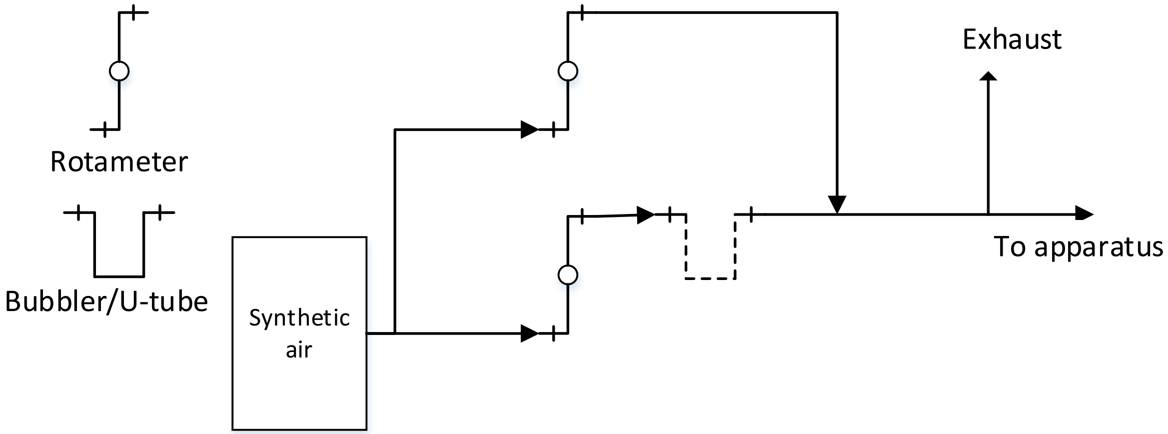


EXPERIMENTS

Experimental setup - two target analytes are investigated

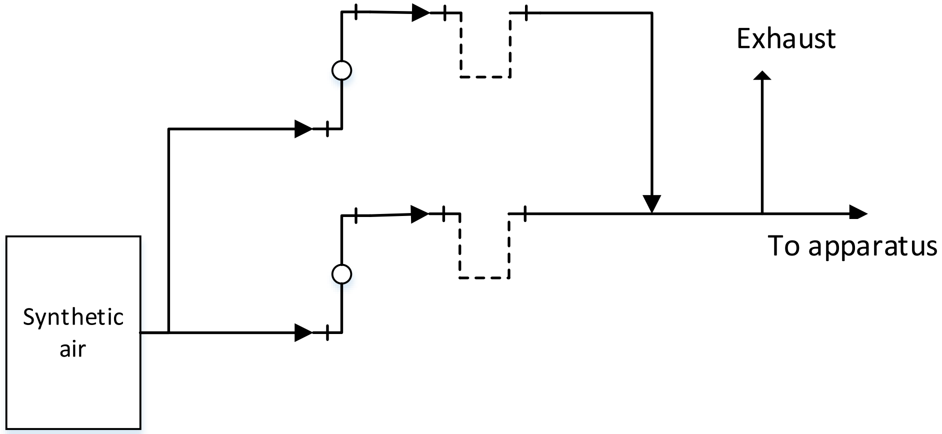
H_2O_2 - explosives precursor

– Generate different mixtures of synthetic air and analyte air samples
– Ratios between the target analyte and clean air: 0.1, 0.4, 0.7, and 1

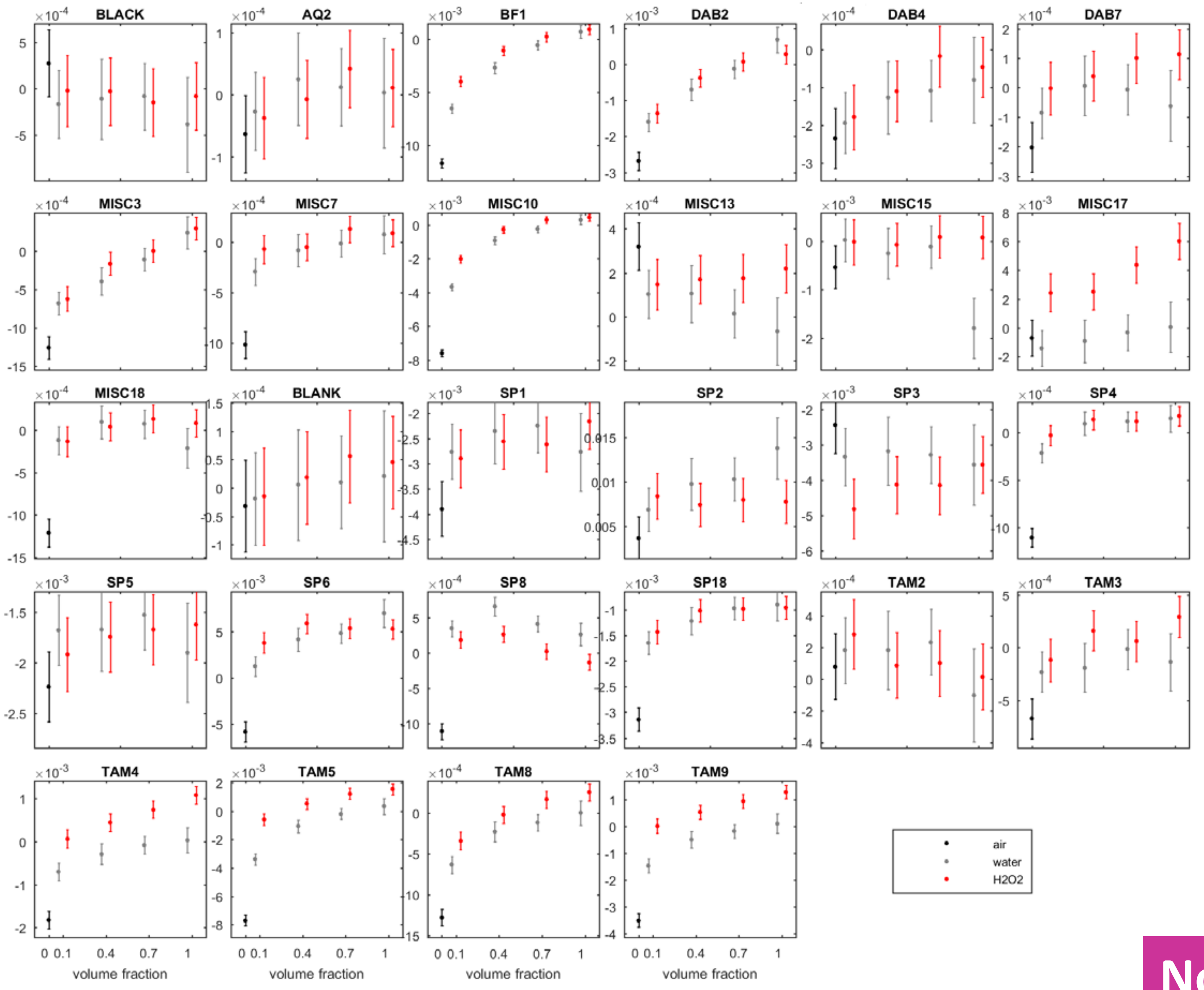


Phenylacetone (BMK) - illegal drug precursor

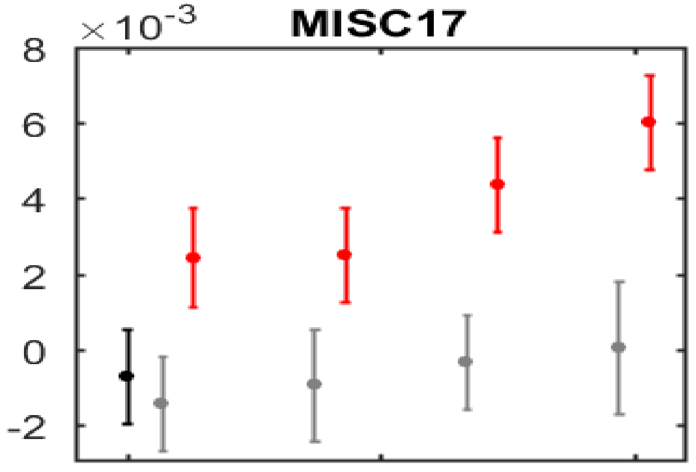
– Compare colorimetric response with naturally occurring confounders, i.e., acetone, diesel, gasoline, ethanol, water, and sea water.
– Clean samples of each substance obtained as well as mixtures of BMK with each confounder was measured.

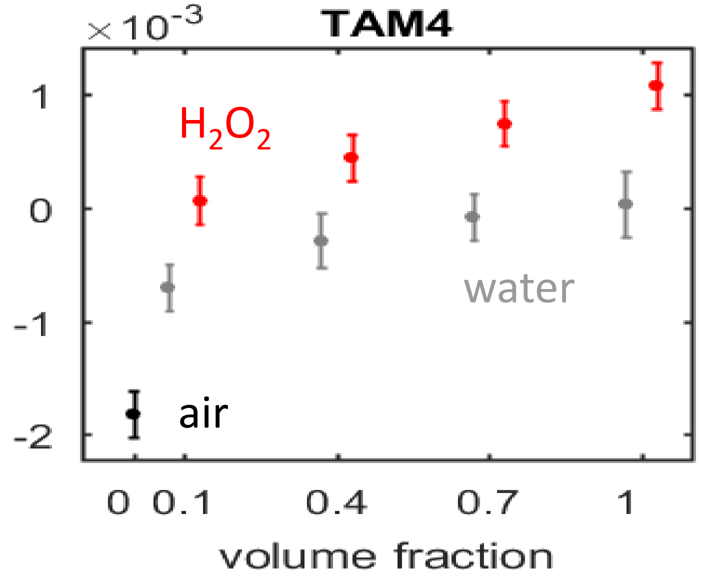


H_2O_2 dye color changes for dilution levels



Select good dyes

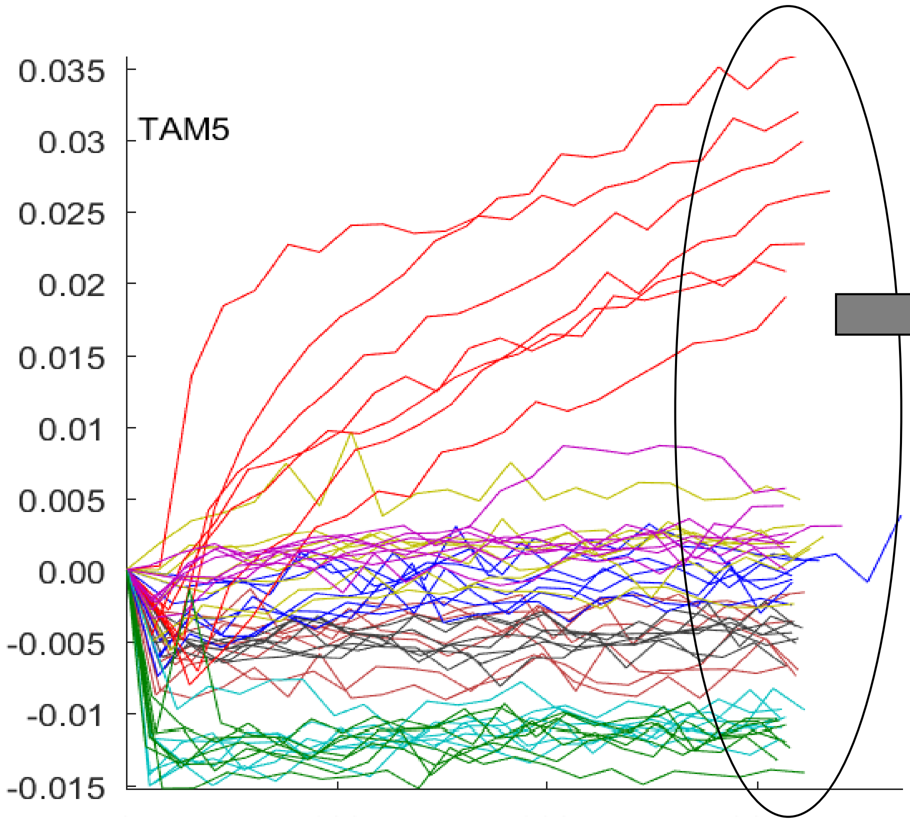




No single dye is sensitive enough!

DETECTION

Data processing pipeline

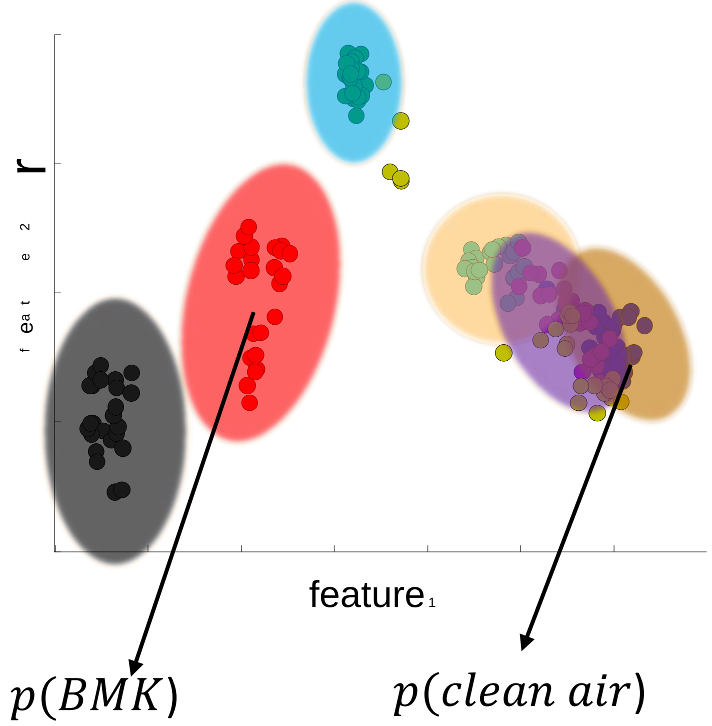


Extract color change (BMK experiments)

$$\begin{bmatrix} x_{BF1-r} \\ x_{BF1-g} \\ x_{BF1-1} \\ x_{DAB2-r} \\ x_{DAB2-g} \\ x_{DAB2-b} \\ \vdots \\ x_{TAM9-r} \\ x_{TAM9-g} \\ x_{TAM9-b} \end{bmatrix}$$

25 dyes x 3 color channels = 75 dim vector

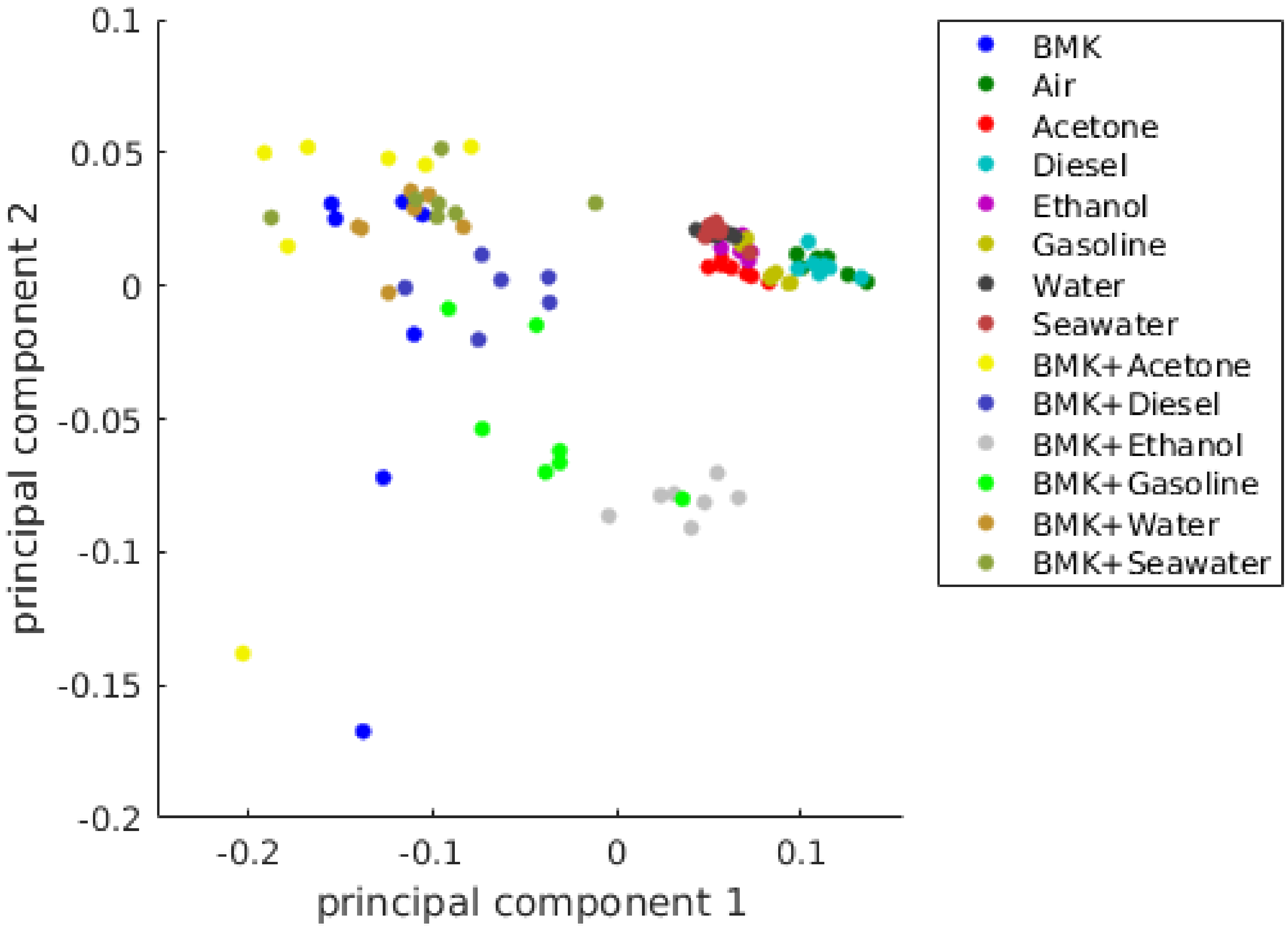
Adjust color change for nuisance factors (GLM-model)



Train classifier to find ”fingerprint” of target analyte

Includes choice of most useful dyes

BMK samples PCA visualization



Target
Confounders

Mixtures=target + confounders

Classification of BMK – 10 fold cross-validation

	KNN	logist. reg.	RF (no PCA)
BMK vs. confounders	97.3%	53.2%	100%
All BMK samples vs. confounders	100%	71%	100%
Train with clean BMK and confounders/ test on BMK mixtures	53%		86%

Classification of H_20_2 – 10 fold cross-validation

	KNN	logist. reg.	RF (no PCA)
H_2O_2	94.3%	75%	95%

KNN

		H_2O_2	confounder
True class	H_2O_2	36	4
	confounder	1	46

log.reg.

		H_2O_2	confounder
True class	H_2O_2	39	1
	confounder	20	27

RF

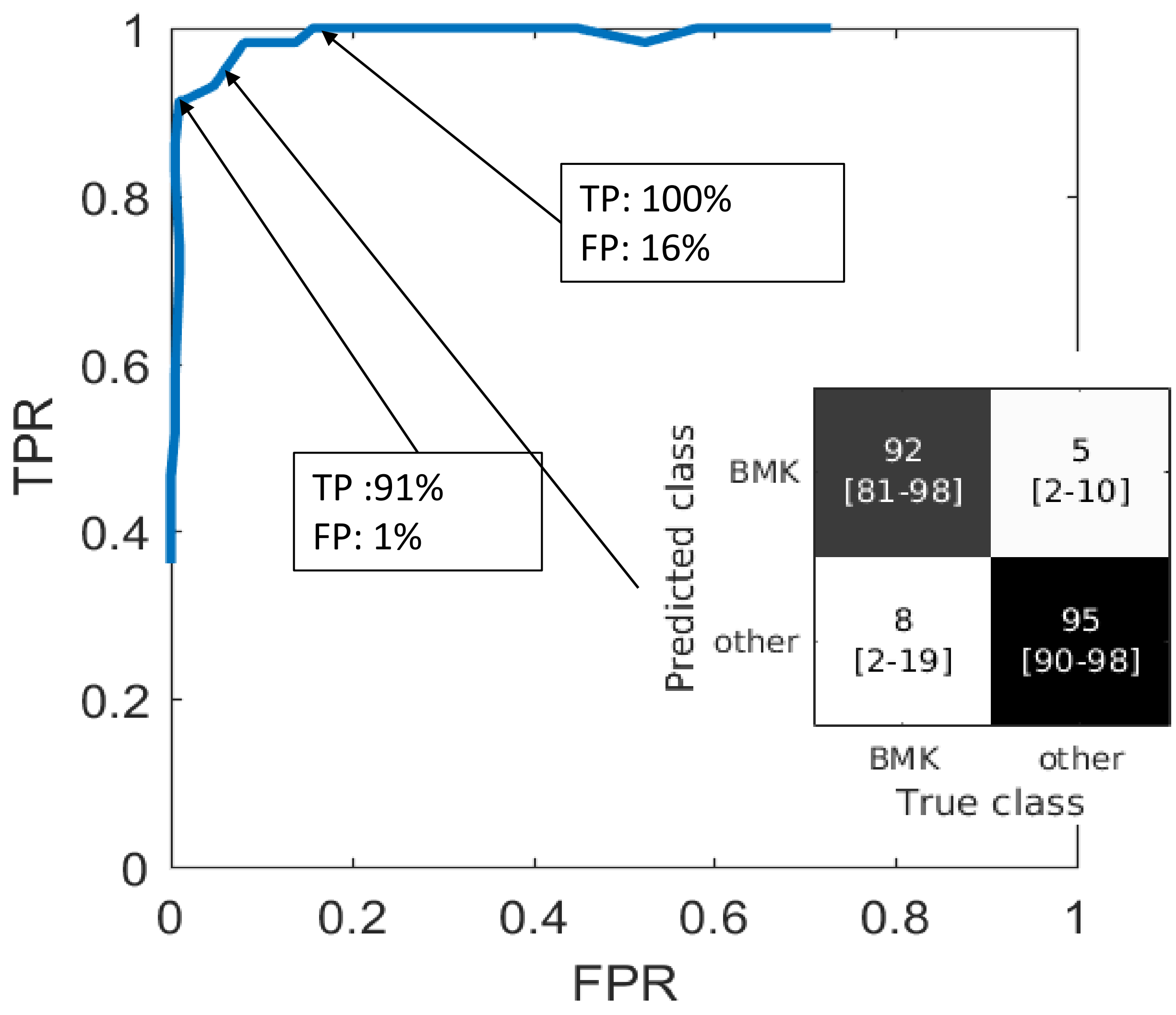
		H_2O_2	confounder
True class	H_2O_2	38	2
	confounder	2	45

USE-CASE RESULTS

Detecting 100 µl BMK in a box

- BMK end-user campaign focused on discriminating BMK from ambient air, water vapor, and acetone.

- Training on 176 samples. Expected performance evaluated on hold-out test set.



Detecting 50 mg HMTD in a box

- HMTD end-user campaign focused on discriminating HMTD from ambient air, water vapor, and H_2O_2.

- Training on 350 samples. Expected performance evaluated using hold-out test set.