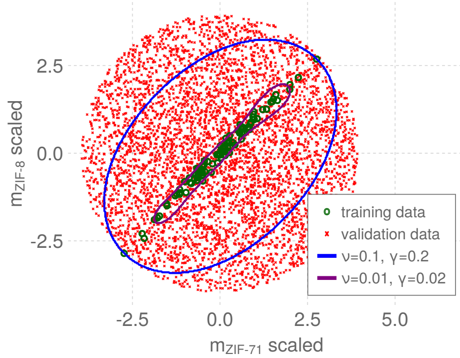
Supplementary Information (SI) for Sensors & Diagnostics.

Supporting Information: Computationally predicting the performance of gas sensor arrays for anomaly detection

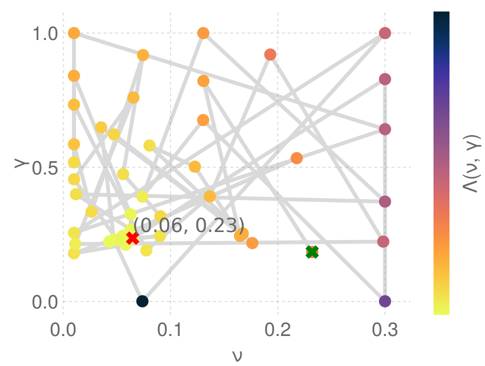
Paul Morris[1] and Cory M. Simon[1]

[1]School of Chemical, Biological, and Environmental Engineering. Oregon State University. Corvallis, OR. USA.
[1]morripau@oregonstate.edu, Cory.Simon@oregonstate.edu

August 14, 2024



(a) Estimating the area inside the SVDD decision boundary.



(b) Sequentially acquired hyperparameters (ν_i, γ_i) by Bayesian optimization.

Figure S1: Tuning the (ν, γ) hyperparameters of the SVDD using Bayesian optimization (BO). (a) To estimate the volume inside a SVDD decision boundary obtained from a given (ν, γ), we generate 5000 synthetic response vectors (red) inside a sphere encompassing the training response vectors (green) and count the fraction that fall inside of the decision boundary. (b) The bounded hyperparameter space is shown $(\nu, \gamma) \in [0.001, 0.3] \times [0.001, 1.0]$ and the sequence (ν_i, γ_i) of 50 acquired hyperparameters by BO colored by the corresponding value of the objective, $\Lambda(\nu_i, \gamma_i)$. The initial hyperparameter pair (ν_0, γ_0), chosen randomly, is marked by a green x. The optimal (ν, γ), acquired on the 32$^{\text{nd}}$ step, is marked by a red x.

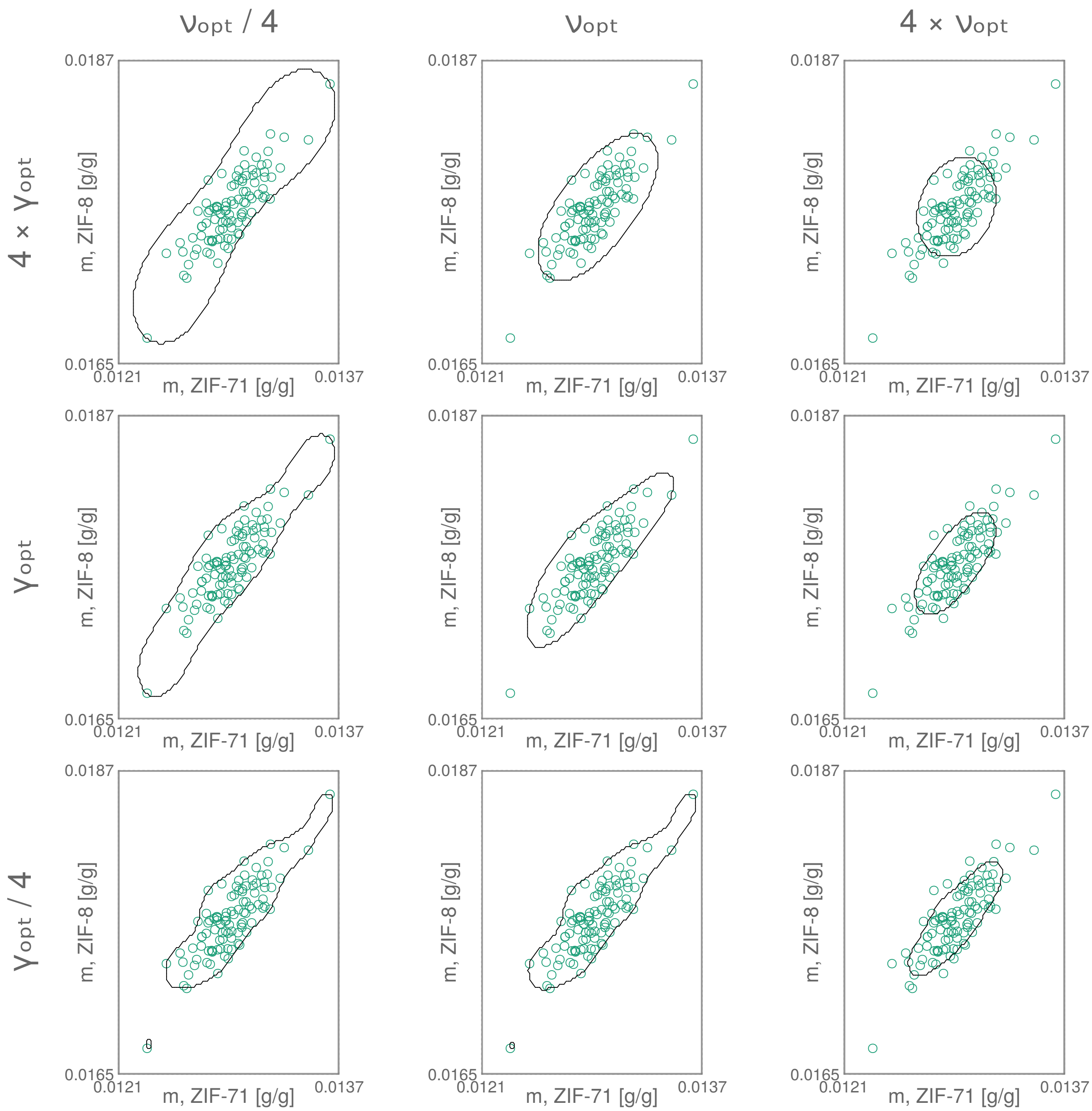
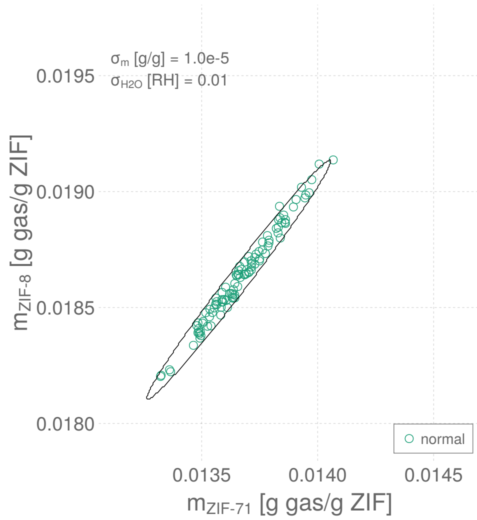
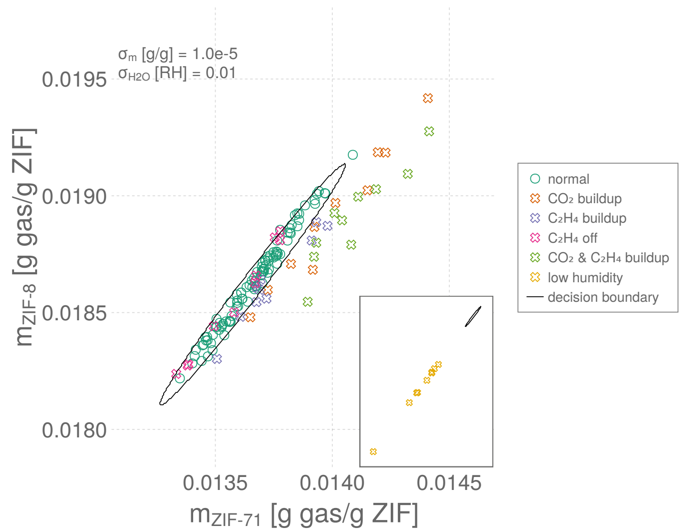


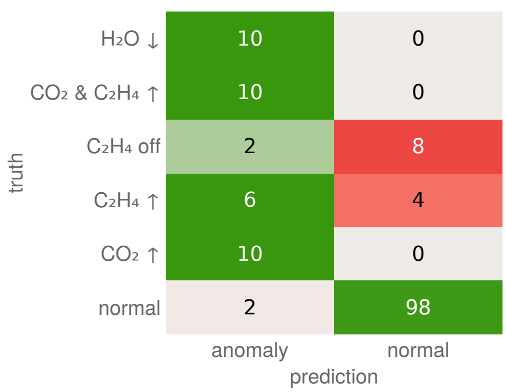
Figure S2: Showing the impact on the trained SVDD decision boundary by decreasing and increasing ν and γ. The SVDD decision boundary made with the optimal pair of (ν, γ) hyperparameters chosen by Bayesian selection is the center column, ν_{opt}, and center row, γ_{opt}. ν values used for training the SVDD increase from the left column to the right column, while γ values used increase from the bottom row to the top row.



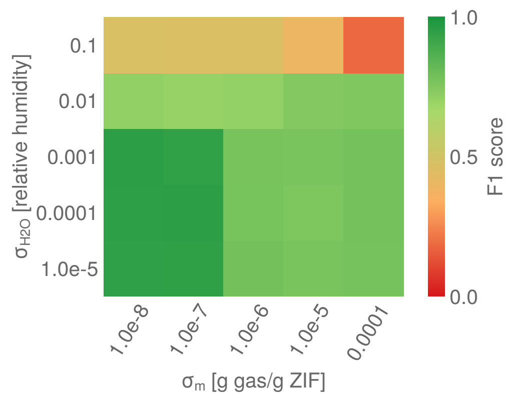
(a) Elliptic envelope decision boundary over training data.



(b) Elliptic envelope decision boundary over test data.



(c) Elliptic envelope test data confusion matrix.



(d) Average F1-score heatmap over a range of sensor error and humidity variances.

Figure S3: Elliptic envelope (EE) anomaly detector. (a) A data set of sensor array response vectors collected under normal conditions used to train the EE anomaly detector and (b) the test set to evaluate the performance of the EE. We found a contamination hyperparameter value of 0.025 to be optimal using the same method for the SVDD. (c) Confusion matrix showing the performance of the trained EE anomaly detector for discriminating normal from anomalous conditions in the test data. (d) A heat map showing average F1-score performance metric on test data over 100 instances of EE anomaly detectors trained and tested on different realizations of gas compositions and their associated sensor responses for a range of variances in humidity and measurement noise.