DOI 10.5162/EUROSENSORS2025/MP2

Bayesian Inference for Reliable Gas Sensing with Metal-Oxide Sensors

Sebastian Pültz[1]
[1] Lab for Measurement Technology, Saarland University, 66123 Saarbrucken, Germany,
s.pueltz@uni-saarland.de

Summary:
In this study, a probabilistic regression approach for the quantification and separation of epistemic and aleatoric uncertainties in gas measurement with temperature-modulated MOS sensors is presented. The uncertainty of feature extraction is analyzed using Monte Carlo simulations. The subsequent regression modeling is based on Partial Least Squares Regression (PLSR) combined with Hamiltonian Monte Carlo (HMC) to capture uncertainties in the parameter estimator. The evaluation shows that most of the uncertainty is due to feature extraction, while epistemic uncertainties remain comparatively small.

Keywords: MOS sensors, TCO, Markov Chain Monte Carlo, Uncertainty

Introduction

MOS gas sensors (metal-oxide semiconductors) provide a viable basis for cost-effective and miniaturizable gas sensors. In order to improve their selectivity despite limited chemical specificity, temperature-cyclic operation has established itself as an effective method [1]. By systematically varying the surface temperature of the sensors, characteristic dynamic response patterns are generated, which are used to quantify gas concentrations by frequently extracting features [2].

This is prone to aleatoric uncertainties due to sensor noise and feature computations. At the same time, epistemic uncertainties are introduced due to limited model knowledge and training data. Classical methods such as PLSR ignore these aspects. This paper therefore presents a Bayesian regression approach with Hamiltonian Monte Carlo (HMC), which quantifies both types of uncertainties separately. In addition, the often neglected uncertainty of the extracted features relevant in temperature-cyclic operation is addressed.

This approach enables a probabilistic assessment of the prediction quality and provides uncertainty measures that go beyond standard errors, providing a crucial step towards more robust and reliable gas sensor systems.

Demonstration data set

The Dataset used for demonstration can be found in [3]. It was collected using two SGP30 gas sensors, each coated with four different MOS sensor layers. The aim is to quantify the concentrations of various gases in complex mixtures. The sensors were operated under temperature-cyclic operating conditions, with each cycle comprising several defined heating and cooling phases. The resulting time-varying response patterns enable the extraction of relevant features for regression analysis.

The complete data set contains several phases. In this work, only data collected with layer 2 of sensor 0 during the initial calibration, i.e. the first series of measurements under controlled laboratory conditions was used. This reduces complexity to validate the proposed Bayesian approach for separating epistemic and aleatoric uncertainty and ensures that external influencing factors such as ageing, or environmental changes are excluded.

Methodes

First, relevant features were extracted by segmenting the temperature-cyclic sensor signals. The mean value and the slope within certain time periods (0.5 s) are extracted as features from each cycle. This extraction reduces data dimensionality while preserving the relevant information about the gas concentrations.

A feature selection is then carried out based on the Pearson correlation to the target value. Then, a partial least squares regression (PLSR), particularly suitable for highly correlated and high-dimensional input data, is used for modeling.

Bayesian inference is applied to the PLSR to quantify the epistemic uncertainty. For this purpose, a Hamiltonian Monte Carlo (HMC) sampler is used, sampling from the posterior distribution of the model parameters and the measurement variance (σ^2), representing the aleatoric uncertainty.

A Monte Carlo simulation was performed to estimate the aleatoric uncertainty due to feature extraction. The initially extracted features were perturbed with random, normally distributed noise corresponding to the empirically estimated standard deviation per feature. For each perturbed feature matrix, a prediction was then calculated using the means of the regression coefficients from the HMC sampling distribution. The standard deviation of these predictions per data point quantifies the uncertainty resulting solely from the variances in feature extraction.

Results

Fig 1 shows the prediction of the hydrogen concentration based on MOS gas sensor data, considering the determined uncertainty components. The mean prediction (black line) closely follows the true curve, with the colored areas visualizing the uncertainty. The dark blue area represents epistemic uncertainty, caused by limited training data and model uncertainty. The light blue area shows the total uncertainty including epistemic and aleatoric components. The latter result from random influences, especially in feature extraction and the residuals. The red dots represent individual prediction samples derived from MCMC sampling. The diagram illustrates the model's ability to quantify and separate uncertainty in a structured manner.

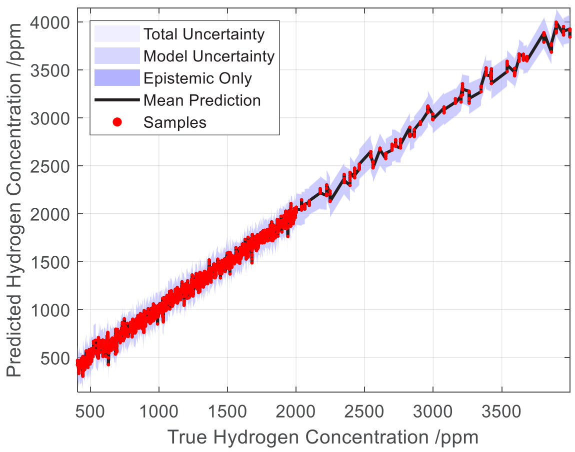

Fig. 1. Prediction of the hydrogen concentration with uncertainty quantification. Shown are the mean prediction (black), epistemic uncertainty (shaded blue), model-induced uncertainty (light blue) and the total bounds (gray). The red dots mark observed calibration values. The uncertainties result from Hamiltonian Monte Carlo (epistemic), feature Monte Carlo (aleatory, sensor-related) and residual model uncertainty.

Fig 2 provides the relative contributions of the different uncertainty components to the overall uncertainty of the predicted hydrogen concentration. The epistemic uncertainty (blue), the aleatoric uncertainty from the feature extraction (red) and the aleatoric uncertainty from the model residuals (yellow) are shown separately. It can be clearly seen that most of the uncertainty is caused by the feature extraction, while the epistemic uncertainty only accounts for a small part. The residual uncertainty is constantly in the lower range. This illustration demonstrates that sensor operation and feature calculation are significant factors influencing the reliability of the prediction.

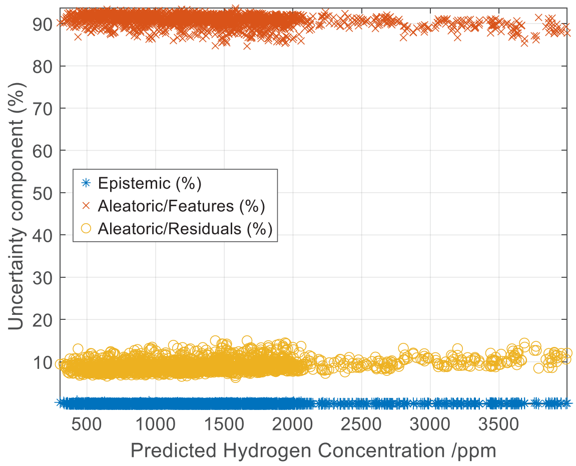

Fig. 2. Prediction of the hydrogen concentration with uncertainty quantification. Shown are the mean prediction (black), epistemic uncertainty (shaded blue), model-induced

Discussion

The presented method successfully separates epistemic and aleatoric uncertainties in gas measurement with temperature-modulated MOS sensors. The analysis shows: The main cause of the overall uncertainty is the feature extraction and to a lesser extent the inherent uncertainty of the raw data. Subsequent work will therefore also deal with more advanced feature extraction, for example within the framework of a convolutional neural network.

References

[1] C. Schultealbert, T. Baur, A. Schütze, and T. Sauerwald, 'Facile Quantification and Identification Techniques for Reducing Gases over a Wide Concentration Range Using a MOS Sensor in Temperature-Cycled Operation', *Sensors*, vol. 18, no. 3, Art. no. 3, Mar. 2018, doi: 10.3390/s18030744.

[2] T. Baur, A. Schütze, and T. Sauerwald, 'Detektion von kurzen Gaspulsen für die Spurengasanalytik: Detection of short gas pulses for trace gas analysis', *tm - Technisches Messen*, vol. 84, no. s1, pp. 88–92, Sep. 2017, doi: 10.1515/teme-2017-0035.

[3] J. Amann, T. Baur, and C. Schultealbert, 'Measuring Hydrogen in Indoor Air with a Selective Metal Oxide Semiconductor Sensor: Dataset'. Zenodo, Mar. 10, 2021. doi: 10.5281/zenodo.4593853.