天津商業大學
TIANJIN UNIVERSITY OF COMMERCE

Development of a Low-Cost Pulse-Modulated VOC Sensor Device for ppb-Level Detection

Chen, Mengdi[1], Hou, Jing[1], Sun, Yuexia[2],*
[1]School of Mechanical Engineering, Tianjin University of Commerce, Tianjin, China
[2]Tianjin Key Lab. of Indoor Air Environmental Quality Control, School of Environmental Science and Engineering, Tianjin University, Tianjin, China
*yuexiasun@tju.edu.cn

Presenter： Chen Mengdi **Date：2026/10/7**

CONTEXT

1 **Research Background and Objectives**
Challenges in indoor VOC monitoring and our proposed solution

2 **Methods**
Hardware design, experimental setup, and data processing workflow

3 **Results**
Key findings in parameter optimization and model performance.

4 **Conclusion**
Summary of achievements and potential applications.

1 Research Background and Objectives

The Severe Challenge of Indoor Air Quality

With the acceleration of urbanization, public buildings have become the primary settings for people's daily activities. However, these densely occupied spaces commonly suffer from indoor volatile organic compound (VOC) contamination, whose hazardous effects are concealed and long-term, posing a potential threat to public health.

Health Risks

Long-term exposure to low-concentration VOCs poses potential, cumulative health risks. Epidemiological studies show that even ppb-level exposure can impair cognitive functions, trigger asthma, and is associated with leukemia and lung cancer.

Diverse Sources

VOCs in indoor environments originate from an extremely wide and complex range of sources, exhibiting an "omnipresent" characteristic:

- Construction and renovation: engineered wood products, paints and coatings, adhesives, etc.
- Living and office: furniture, carpets, printers, carbon paper, etc.
- Daily chemical products: air fresheners, cleaning agents, cosmetics, etc.

1 Research Background and Objectives

Limitations of Current Monitoring Methods and Equipment

Routine Fixed-Site Monitoring

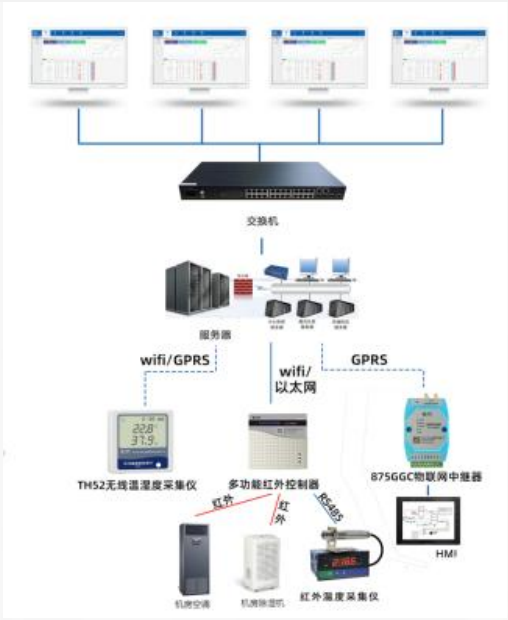






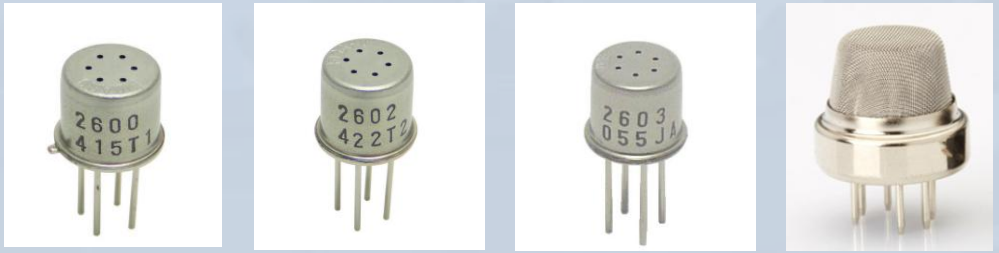
- High cost
- Poor real-time performance
- High precision

Real-Time Online Monitoring



- Low cost
- High real-time performance
- Low precision

MOS sensor



Advantages

- Low cost and scalability
- Compact and Portable
- Ultra-low power consumption

Limitations

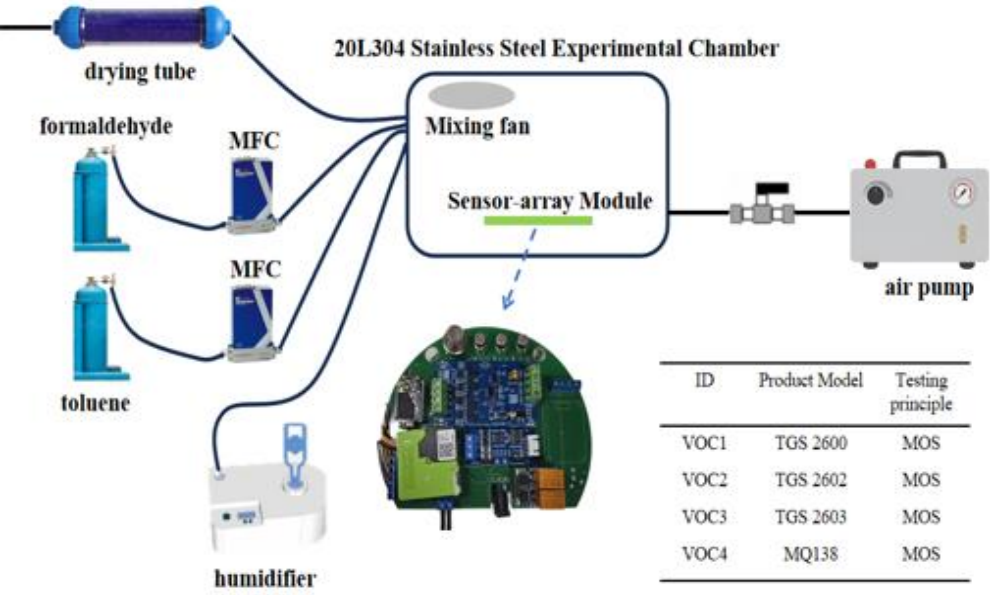
- Limited sensitivity
- Cross-sensitivity issue
- Baseline drift and instability

Objective: To develop a low-cost, high-precision VOC sensor device capable of continuous monitoring and component identification.

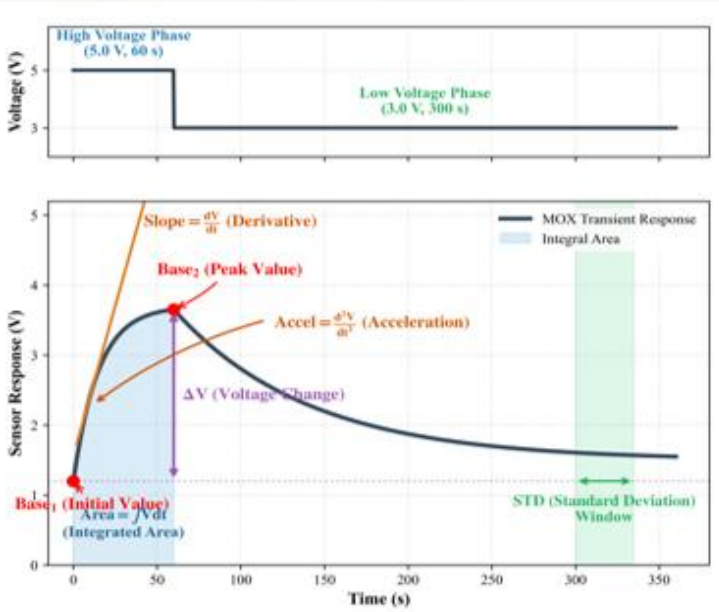
2 | Methods:Pulse Modulation + Embedded Machine Learning

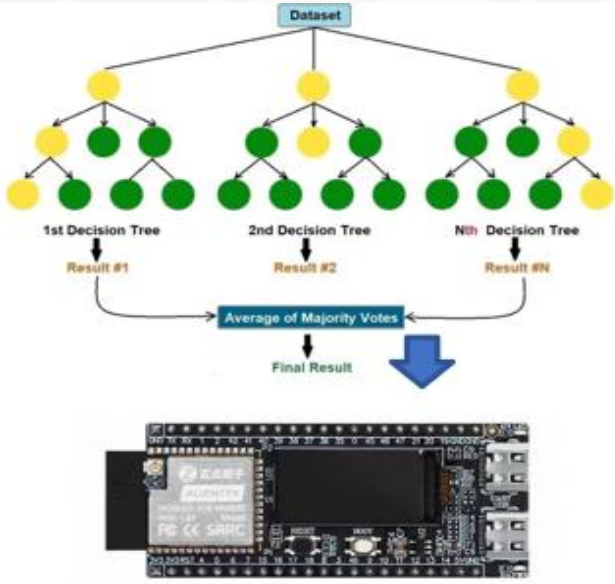
Technical workflow :

Transient signal acquisition → **Feature extraction** → **Model development** → **Embedded deployment**



ID	Product Model	Testing principle
VOC1	TGS 2600	MOS
VOC2	TGS 2602	MOS
VOC3	TGS 2603	MOS
VOC4	MQ138	MOS





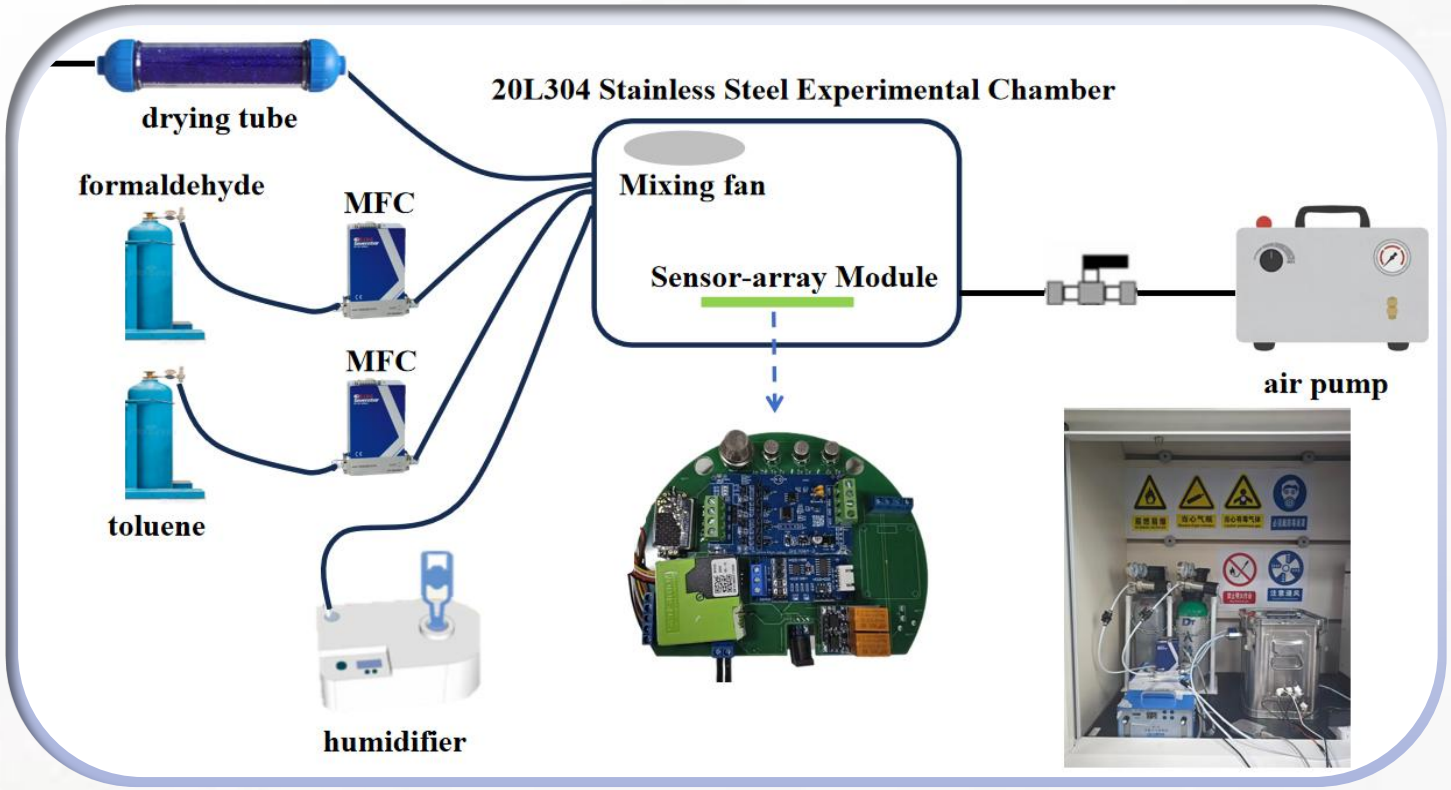
1.Temperature–voltage dual pulse modulation experiments were conducted on an array consisting of four MOS sensors.

2.Time-domain curve features are extracted from the sensor responses.

3.A Random Forest model is trained to predict gas concentrations.

4.The model is converted to C code and run on the ESP32 microcontroller for offline inference.

Controlled environment :



Test chamber
A sealed 20 L stainless steel chamber was used to ensure controlled gas concentrations.

Target gases
Toluene and formaldehyde (common indoor VOCs).

Concentration range
30 to 150 ppb (a challenging range for low-cost sensors).

Gas generation
Mass flow controllers and a humidifier were used to create the desired gas mixtures.

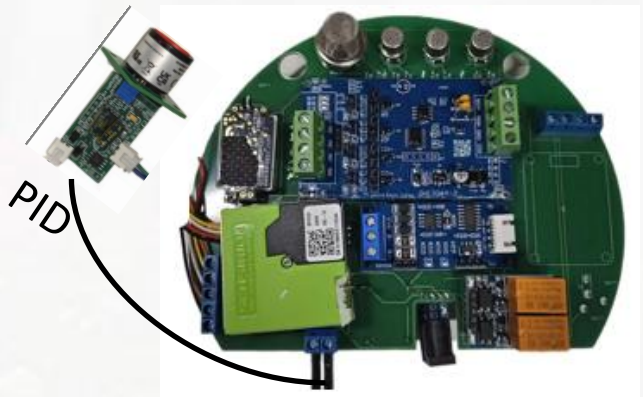
Experimental procedure :Control variable method and data acquisition

01. Environmental Initialization and Controlled Variables

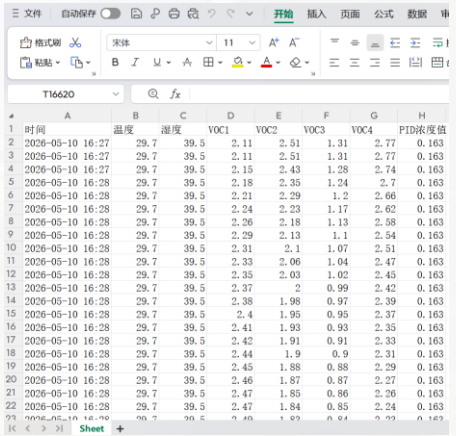
- Sensor warm-up: 120 min preheating for hardware stabilization
- Environment setting: Humidity controlled at target level (e.g., 30%); 30 min stabilization
- Gas injection: Formaldehyde or toluene standard gases (30–150 ppb) introduced for testing
- Program setting: The ESP32 drives the sensors with alternating 5 V / 3 V signals for temperature–voltage dual pulse modulation.

02. Multi-source synchronous monitoring and data storage

- Array coordination: Synchronous acquisition from 4-channel MOS array and PID reference sensor for multi-dimensional sensing
- High-frequency sampling: 1 Hz sampling rate to capture rapid gas response dynamics
- Real-time transmission: Data sent to PC database via serial port, ensuring complete data integrity

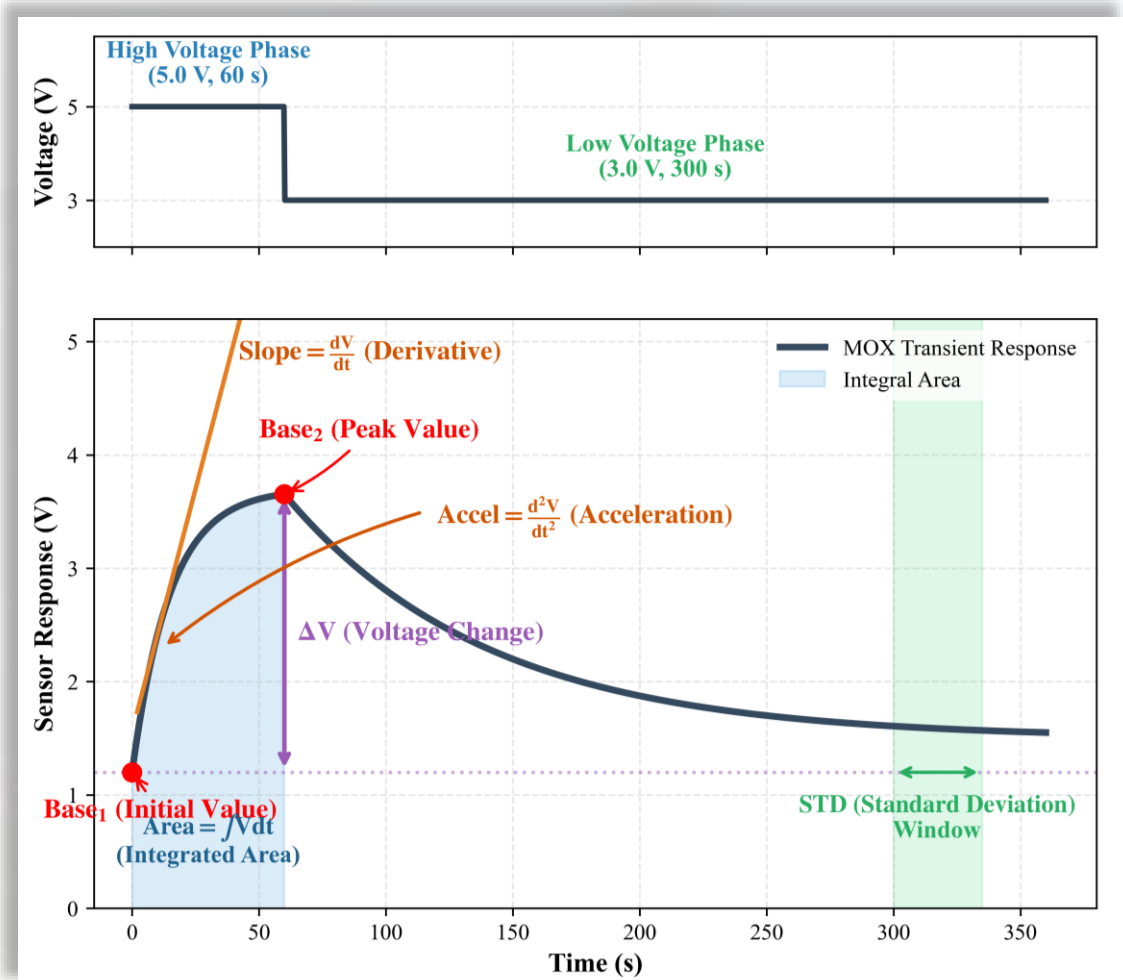






时间	温度	湿度	VOC1	VOC2	VOC3	VOC4	PID浓度值
2026-05-10 16:27	29.7	39.5	2.11	2.51	1.31	2.77	0.163
2026-05-10 16:27	29.7	39.5	2.11	2.51	1.31	2.77	0.163
2026-05-10 16:27	29.7	39.5	2.15	2.43	1.28	2.74	0.163
2026-05-10 16:28	29.7	39.5	2.18	2.35	1.24	2.7	0.163
2026-05-10 16:28	29.7	39.5	2.21	2.29	1.2	2.66	0.163
2026-05-10 16:28	29.7	39.5	2.24	2.23	1.17	2.62	0.163
2026-05-10 16:28	29.7	39.5	2.26	2.18	1.13	2.58	0.163
2026-05-10 16:28	29.7	39.5	2.29	2.13	1.1	2.54	0.163
2026-05-10 16:28	29.7	39.5	2.31	2.1	1.07	2.51	0.163
2026-05-10 16:28	29.7	39.5	2.33	2.06	1.04	2.47	0.163
2026-05-10 16:28	29.7	39.5	2.35	2.03	1.02	2.45	0.163
2026-05-10 16:28	29.7	39.5	2.37	2	0.99	2.42	0.163
2026-05-10 16:28	29.7	39.5	2.38	1.98	0.97	2.39	0.163
2026-05-10 16:28	29.7	39.5	2.4	1.95	0.95	2.37	0.163
2026-05-10 16:28	29.7	39.5	2.41	1.93	0.93	2.35	0.163
2026-05-10 16:28	29.7	39.5	2.42	1.91	0.91	2.33	0.163
2026-05-10 16:28	29.7	39.5	2.44	1.9	0.9	2.31	0.163
2026-05-10 16:28	29.7	39.5	2.45	1.88	0.88	2.29	0.163
2026-05-10 16:28	29.7	39.5	2.46	1.87	0.87	2.27	0.163
2026-05-10 16:28	29.7	39.5	2.47	1.85	0.86	2.26	0.163
2026-05-10 16:28	29.7	39.5	2.47	1.84	0.85	2.24	0.163

Feature extraction



Feature extraction：

For each sensor channel, we extract six key features from the transient response curve: Base, Delta, STD, Area, Slope, and Accel.

Feature vector construction：

Sensor data aggregation:

4 sensor channels × 6 features = 24 features

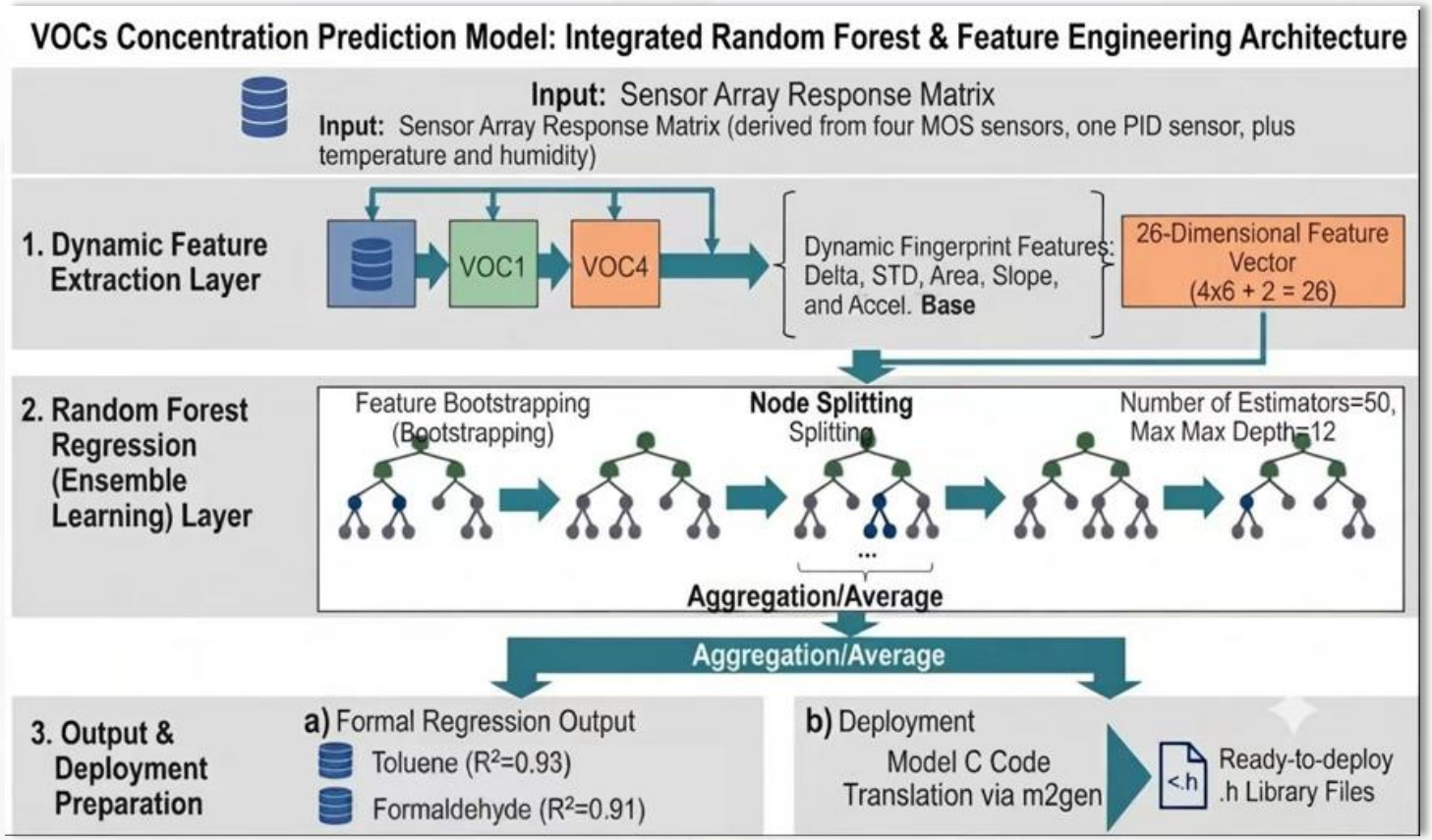
Environmental parameter compensation:

+ 2 environmental parameters (temperature and humidity)

Total: 26-dimensional feature vector

The 26-dimensional feature vector from each sensing cycle serves as the input to the machine learning model.

Model training and deployment



01. Model selection

Random Forest (RF) is a powerful ensemble learning algorithm selected for its high accuracy and capability to handle high-dimensional data.

02. Training data

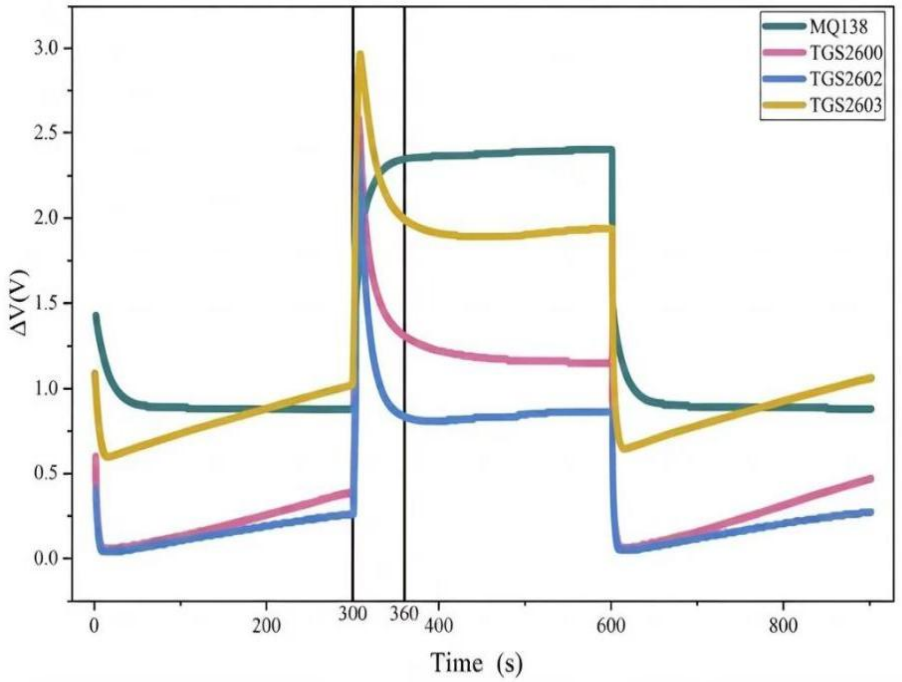
A total of 216 sensing cycles were collected for model training and validation.

03. Embedded deployment

The trained RF model was converted to C code and deployed on the ESP32 for real-time offline inference, with no internet connection required.

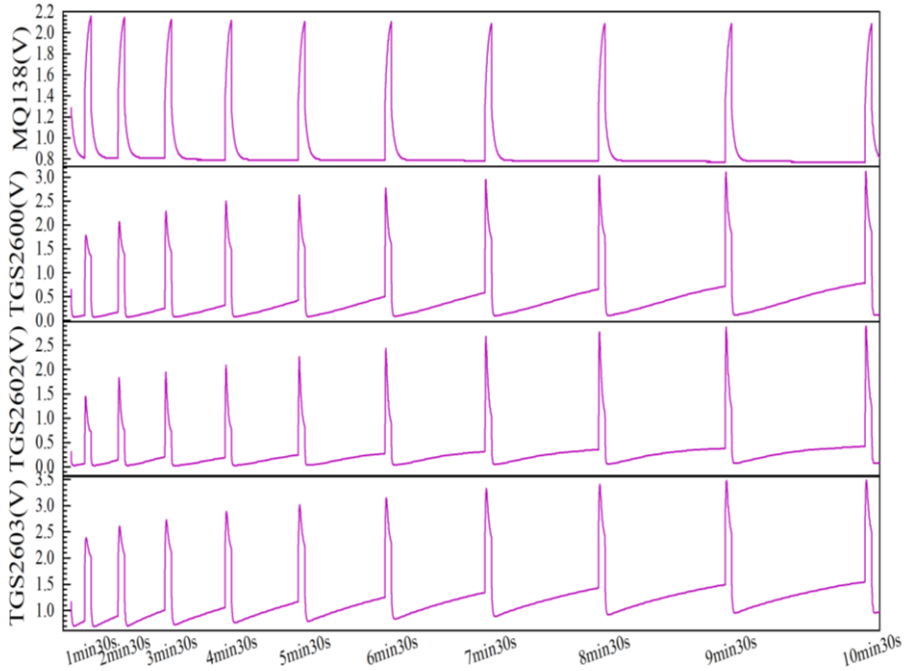
The pulse parameters were optimized to maximize the information content of the transient responses.

Heating duration t_H



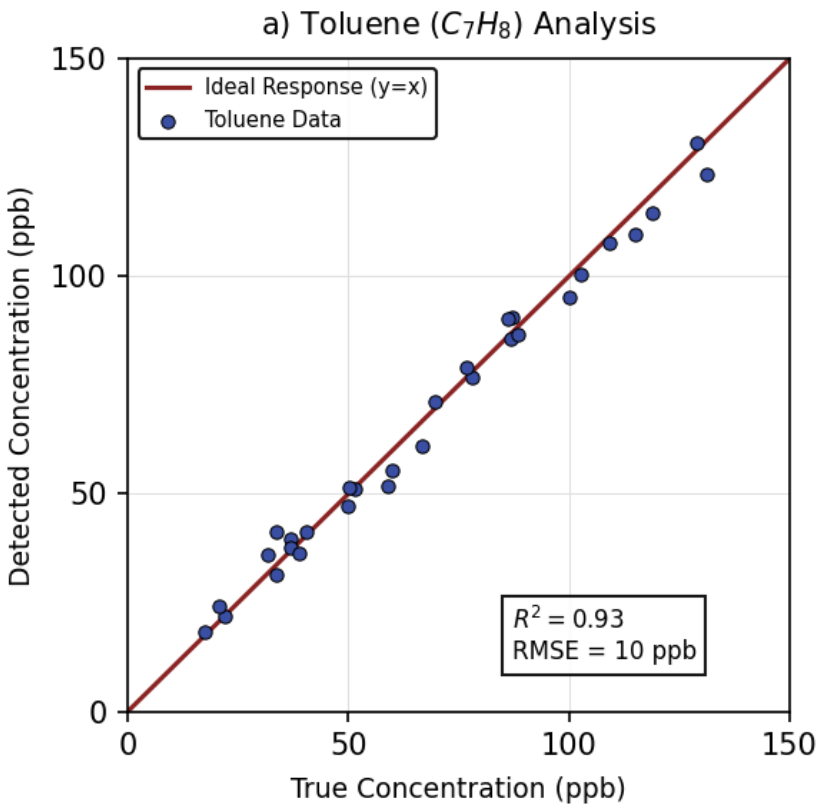
Different heating durations were tested, and 60 s provided the best balance between signal intensity and response time.

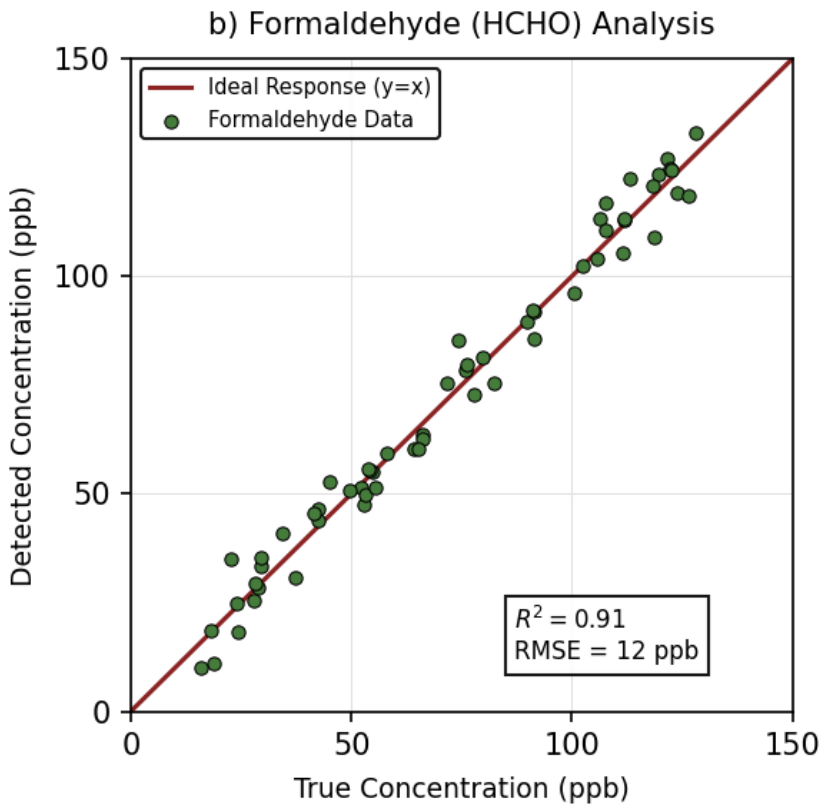
Recovery duration t_L



A recovery time of 300 s was chosen to ensure that the sensor baseline largely recovered before the next pulse cycle.

Achieving High-Precision Prediction and Component Identification





Toluene Prediction

Coefficient of determination (R^2)

R^2=0.93

Root mean square error (RMSE)

RMSE=10ppb

Formaldehyde Prediction

Coefficient of determination (R^2)

R^2=0.91

Root mean square error (RMSE)

RMSE=12ppb

The high R^2 values and low RMSE values indicate that the model can accurately predict VOC concentrations within the tested range.

Successful deployment 01

The trained Random Forest model has been successfully deployed on the ESP32 microcontroller, achieving an average single prediction time of **2.74 ms** over repeated measurements.

Real-time performance 02

The device can process sensor data and provide real-time concentration predictions.

Offline capability 03

All computations are performed locally on the device, providing convenience and high efficiency.

Core advantage: The device bridges the gap between high-precision environmental sensing and autonomous, low-power operation, making it well-suited for a wide range of complex scenarios.

4 | Conclusion: Summary of Key Achievements



01. Low-cost device:
A cost-effective VOC sensor device was developed using off-the-shelf components (ESP32 and MOS sensors).

02. High sensitivity:
Using pulse-modulated transient responses, reliable ppb-level quantification of toluene and formaldehyde was demonstrated.

03. Accurate model:
High prediction accuracy ($R^2 > 0.90$) was achieved using the Random Forest model.

04.Standalone operation
The model was successfully deployed on the ESP32, enabling real-time and offline monitoring.


Thank you for your attention!