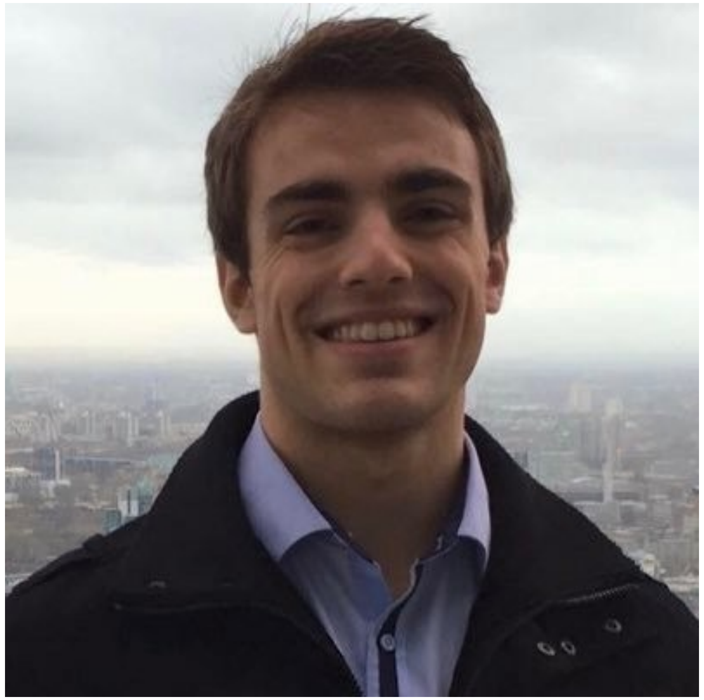
SciTinyML - ICTP workshop
Scientific Use of Machine Learning on Low Power Devices

Applications of TinyML for atmospheric monitoring

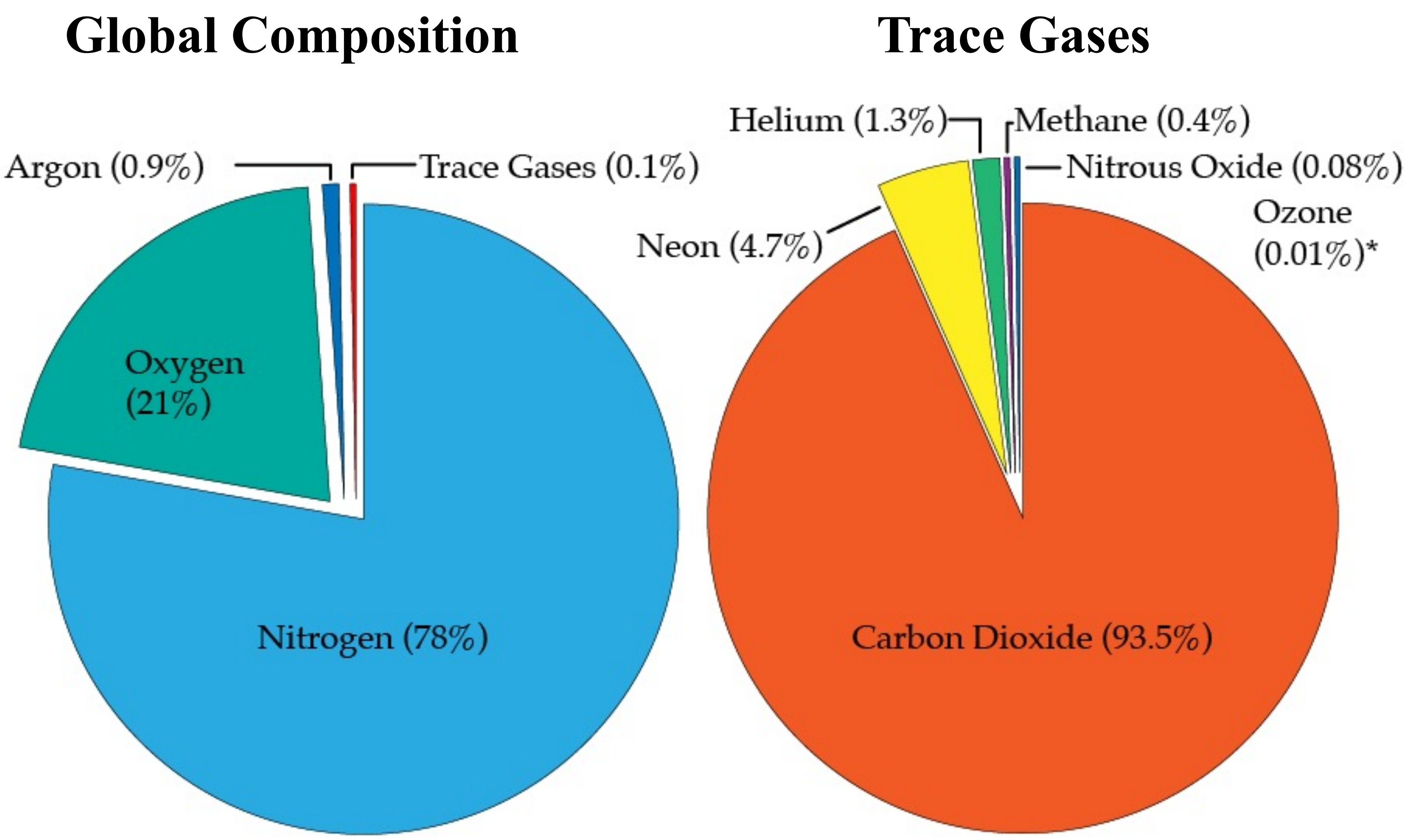
Matthew Stewart
Harvard University, USA
Email: Matthew_stewart@g.harvard.edu

About Me

- BEng/MEng in Mechanical Engineering in 2017 from Imperial College London with a year abroad at the National University of Singapore.
- Currently pursuing Ph.D. in engineering sciences and data science at Harvard University, developing intelligent chemical sensing systems to study the atmospheric chemistry of the tropical Amazonian rainforest.
- Data science blogger published in Towards Data Science, TOPBOTS, Experfy, and other sites.
- Contributor to the HarvardX TinyML series on EdX.



Atmospheric Composition

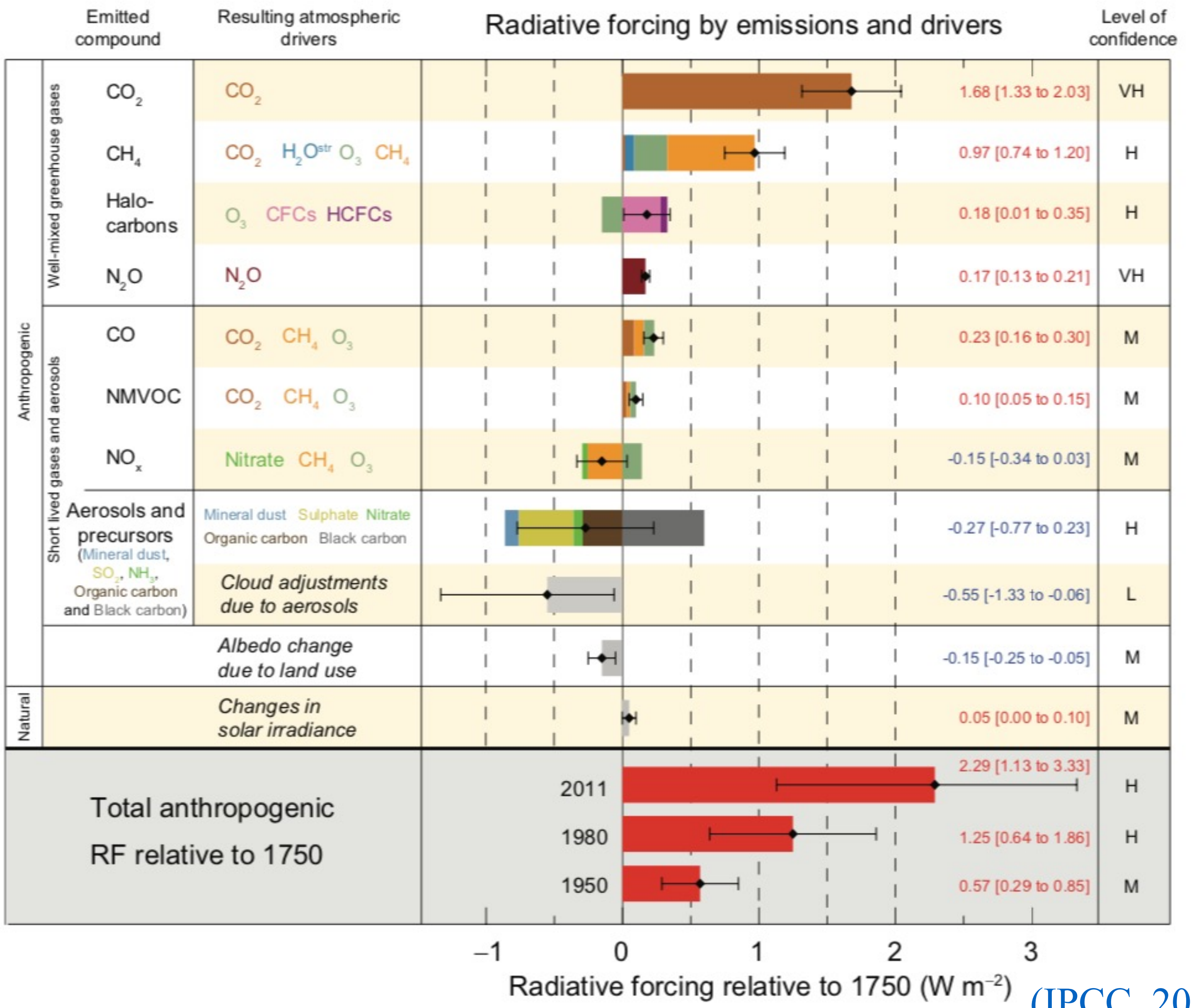


Despite their relatively small abundance, trace gases (1) dominate atmospheric reactivity, (2) strongly influence the climate, as well as public health, and (3) are a major focus of atmospheric research.

Important Atmospheric Constituents – Climate Change

- Recent IPCC reports (AR4-AR6) highlights the importance of certain trace gas emissions on global temperatures
- Most closely monitored emissions are carbon dioxide, methane, CFCs, carbon monoxide, and nitrogen oxides
- Aerosols also play an important factor but are still poorly understood

What about non-methane volatile organic compounds (NMVOCs)?



(IPCC, 2014)

Important Atmospheric Constituents – NMVOCs

The dominant emission source of volatile organic compounds (VOCs) to the atmosphere is from biogenic sources, such as tropical (e.g., Amazon/Congo/Borneo) and boreal (e.g., Canada/Russia) forests, but also industry.

VOCs lead to ozone and aerosol formation via atmospheric cycling.

It is estimated that more than 10,000 VOC species are emitted to the atmosphere (Goldstein and Galbally, 2007).

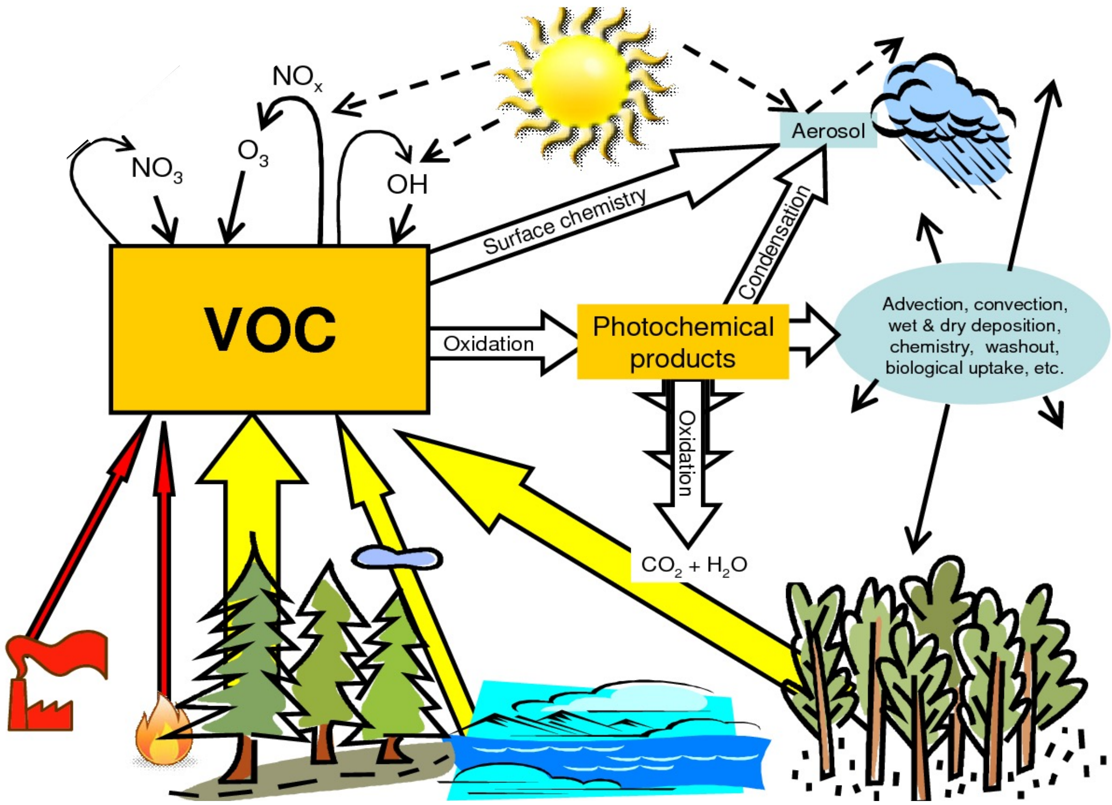
Consequences

Chemistry: source and sinks, local or long transport

Climate: light scattering & absorption, effect on clouds

Health: asthma, mortality, lung cancer

Vegetation: reduction in light required for photosynthesis and increase in leaf temperature due to changed surface optical properties



Important Atmospheric Constituents – Sulfur + Nitrogen Oxides

Two important gases are nitrogen oxides and sulfur dioxide, the main constituents of acid rain (rain with an acidic pH range).

Predominantly emitted through lightning strikes, volcanic eruptions, and by the combustion of sulfur-containing coal.

Acid rain is bad for human health, crops, and can cause weathering and erosion of objects out in the open.

Most countries have regulations that have largely eliminated the problem of acid rain, but it is still prevalent in some regions of the world.

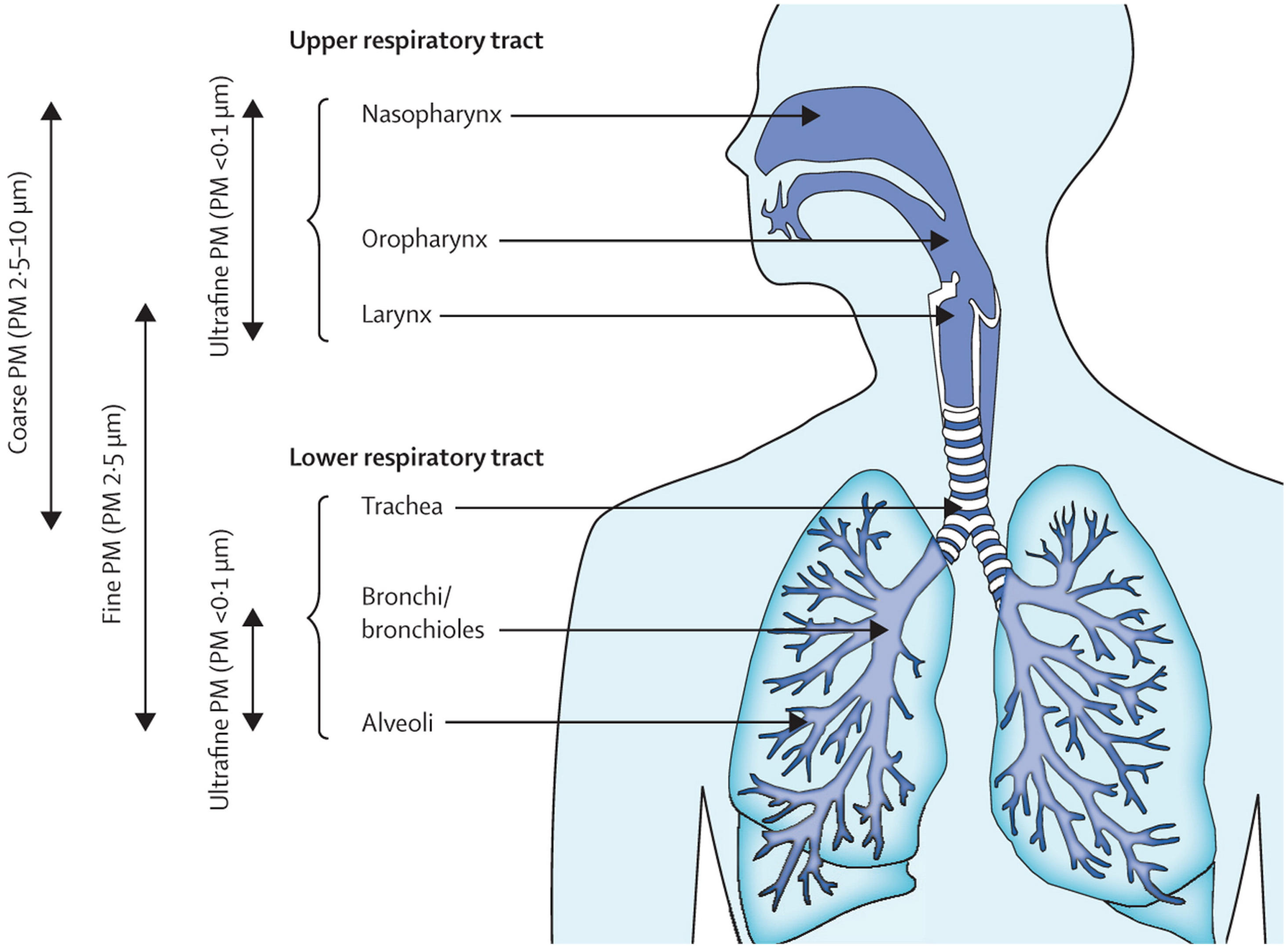
Important Atmospheric Constituents – Aerosols

Aerosols come in two forms: primary and secondary.

Primary aerosols are produced via wind erosion, sandstorms, sea salt, etc., and are generally larger in size.

Secondary aerosols are formed from the oxidation of gases, which lowers their vapor pressure and makes them more susceptible to condensation. These particles are generally smaller in size and act as nuclei for cloud droplets.

Smaller particles can penetrate deep into human lungs, while larger particles are deposited in the upper respiratory tract.



(Guarnieri and Balmes, 2014)

Examples of Aerosols

Aerosol haze from industrial pollutants in Beijing

Visibility limited by aerosol scattering in southwest Poland

How are gases and aerosols measured?

There are two ways we can study gases and aerosols: top-down and bottom-up.

- Global/regional scale → top-down measurement → satellite observations
- Local/hyperlocal scale → bottom-up measurement → singular devices

Top-down

- Analysis performed at the global, national, or regional scale
- Typically utilizes satellite data
- Involves large-scale data processing (~1 TB)
- Spatial granularity of data is low, typically 10-30 km
- Temporal granularity is typically low: most commonly daily, weekly, or monthly values
- Often column-based measurements instead of ground-level measurements

Bottom-up

- Analysis performed at the local or hyperlocal scale
- Typically uses singular devices or sensor systems
- Much smaller data sizes, processing can often be performed on small devices like microcontrollers
- Spatial granularity is fixed unless the device is moved or is part of a larger network of devices
- Temporal granularity can be as high as allowed by the sensing mechanism utilized
- Typically get in-situ, ground-level measurements

How are gases and aerosols measured?

There are two ways we can study gases and aerosols: top-down and bottom-up.

- Global/regional scale → top-down measurement → satellite observations
- Local/hyperlocal scale → bottom-up measurement → singular devices

Top-down

- Analysis performed at the global, national, or regional scale
- Typically utilizes satellite data
- Involves large-scale data processing (~1 TB)
- Spatial granularity of data is low, typically 10-50 km
- Temporal granularity is typically low: most commonly daily, weekly, or monthly values
- Often column-based measurements instead of ground-level measurements

Traditional ML methods favored

Bottom-up

- Analysis performed at the local or hyperlocal scale
- Typically uses singular devices or sensor systems
- Much smaller data sizes, processing can often be performed on small devices like microcontrollers
- Spatial granularity is fixed unless the device is moved or is part of a larger network of devices
- Temporal granularity can be as high as allowed by the sensing mechanism utilized
- Typically get in-situ, ground-level measurements

TinyML methods favored

Gold-standard instrumentation vs. low-cost sensor

Alphasense OX-A431 Ozone Sensor
Cost: (~$100)
Measurement Range: 0-20 ppm
Measurement Accuracy: +/- 15 ppb
Weight: 6 g
Power Consumption: < 1 mW

Eco Sensors model UV-100 ozone analyzer
Cost: (~$4500)
Measurement Range: 0.01-999 ppm
Measurement Accuracy: +/- 2%
Weight: 2.1 kg
Power Consumption: 4.8 W

Challenges integrating TinyML with gas sensors

(1) Not all sensors are created equal

Electrochemical sensors generally have low power consumption, but some low-cost sensors have higher power requirements, especially those with optical/heating sensing mechanisms.

Alphasense OX-A431 Ozone Sensor
Power Consumption: ~1 mW

Alphasense PID-AH1/2 Sensor
Power Consumption: ~100 mW

Alphasense OPC-R2 Particle Monitor
Power Consumption: ~300 mW

Alphasense CH-A3 Combustible Gas Pellistor
Methane Sensor
Power Consumption: 190 mW

Alphasense IRC-A1 CO2 Sensor
Power Consumption: 300 mW

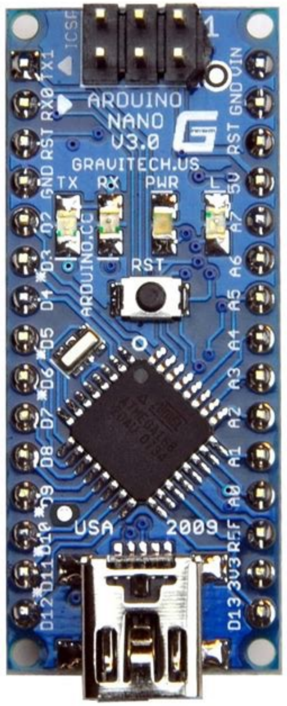
Alphasense VOC Metal Oxide Sensors
Power Consumption: ~500 mW

Challenges integrating TinyML with gas sensors

(2) Combining sensors rapidly increases power consumption

TinyML is typically defined with on-device power consumption on the order of 1 mW – by combining multiple sensors our device can quickly reach power consumption values on the order of ~1W

Arduino Nano



Current limit: 500 mA

Interface

Alphasense PID-AH1/2 Sensor
Power Consumption: ~100 mW

Alphasense OPC-R2 Particle Monitor
Power Consumption: ~300 mW

Alphasense IRC-A1 CO2 Sensor
Power Consumption: 300 mW

Alphasense VOC Metal Oxide Sensors
Power Consumption: ~500 mW

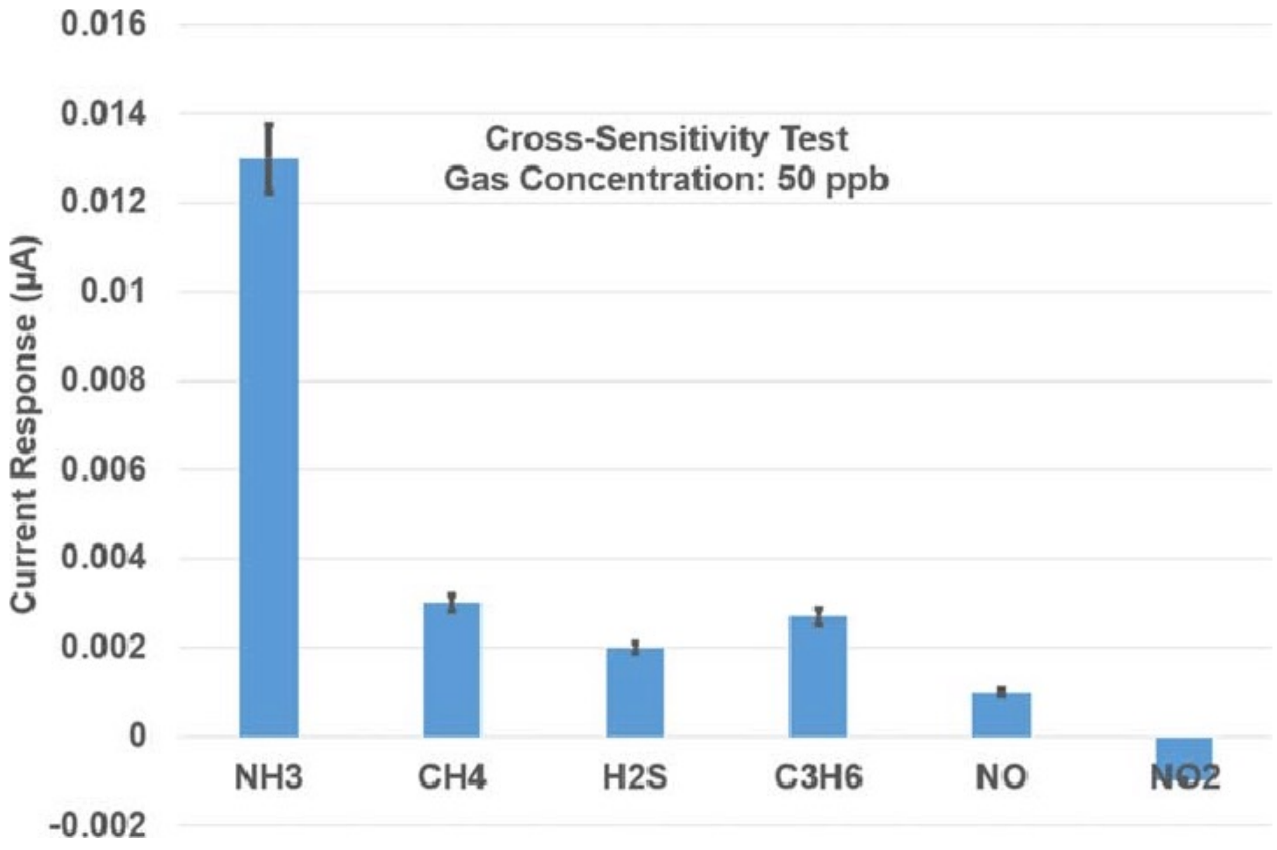
Challenges integrating TinyML with gas sensors

(3) Cross-sensitivity of gases

Cross sensitivity occurs when a sensor shows readings for a gas that is not the target gas. This interfering gas causes a reaction in the sensor — therefore showing a change in readings — even if the target gas is not present.

Examples:

- Hydrogen sulfide sensor responds to increase in hydrogen gas
- PID sensor signal may drop in the presence of high methane concentrations due to absorption of photons



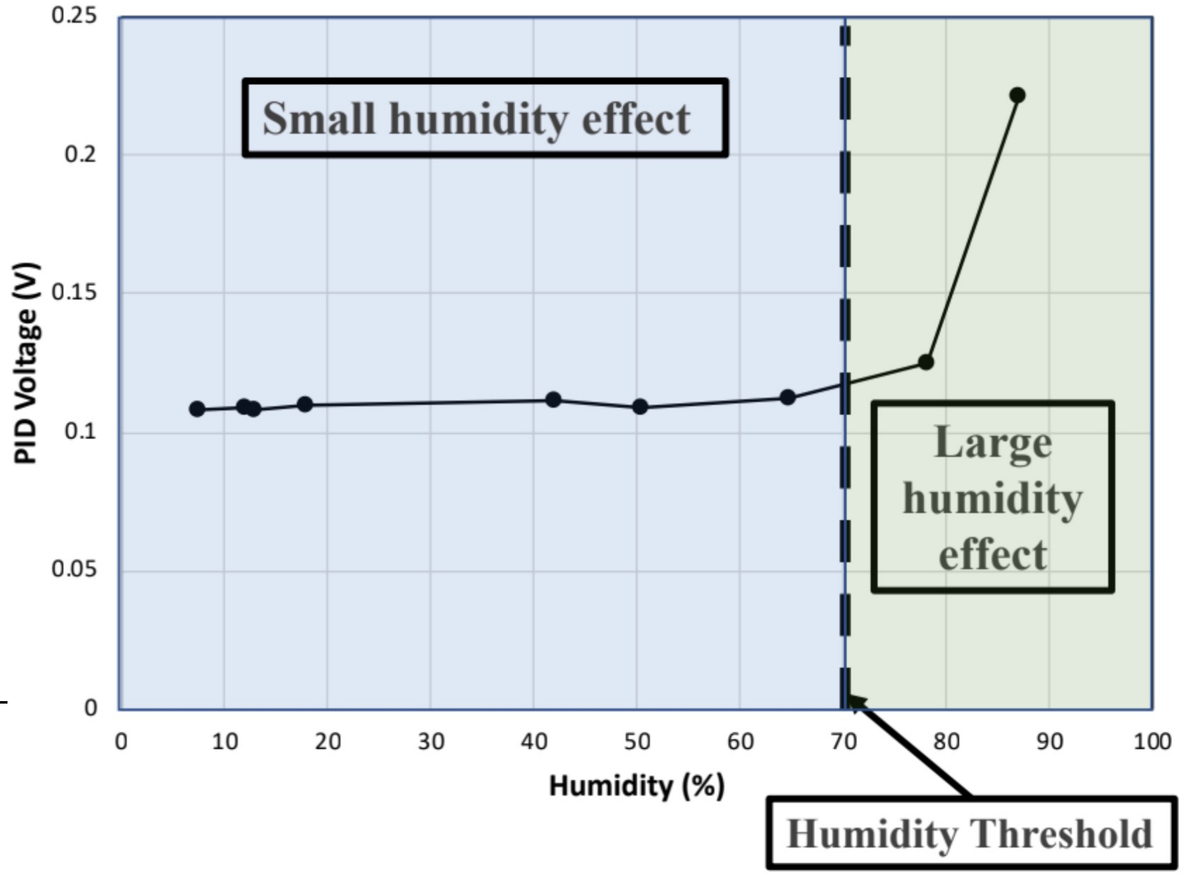
Cross-sensitivity could be likely be eliminated using TinyML!

Challenges integrating TinyML with gas sensors

(4) Noise from changes in environmental variables

Changes in temperature, humidity, and pressure can influence sensor measurements depending on the sensing mechanism used.

- Pressure can be assumed constant for most atmospheric monitoring purposes
- Temperature → maybe correction factor
- Humidity → correction factor



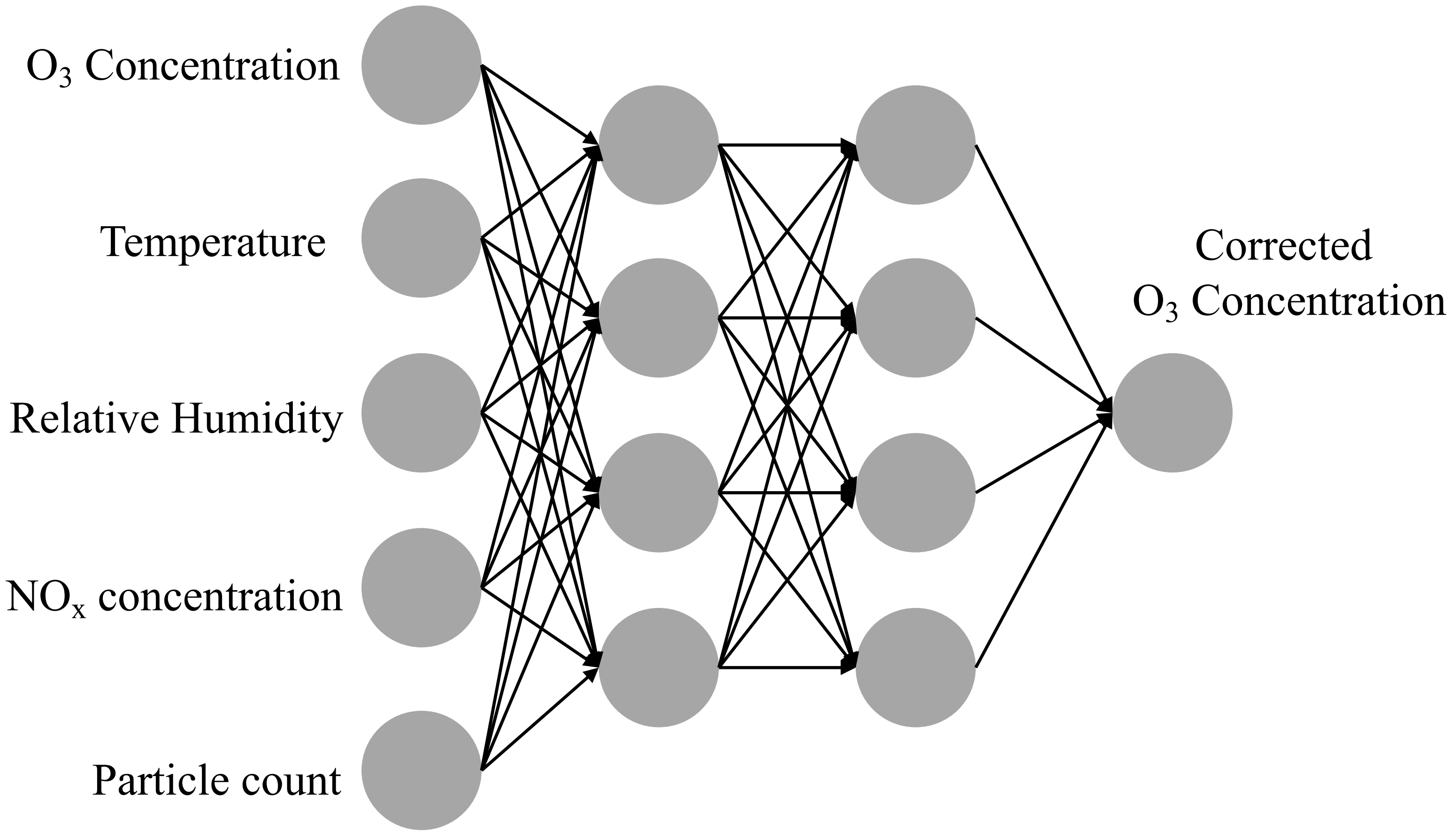
It is standard practice to implement a temp/RH sensor onboard to introduce the correction factor – this could be learned and continuously retrained on-device using TinyML

Applications of TinyML for Atmospheric Monitoring

(1) Real-time sensor correction

Data corrections can be performed in real-time by passing sensor data through the inference pipeline of a trained model

We could also train our model to predict the concentration of a gold-standard device using the low-cost sensor and other peripheral sensors



Applications of TinyML for Atmospheric Monitoring

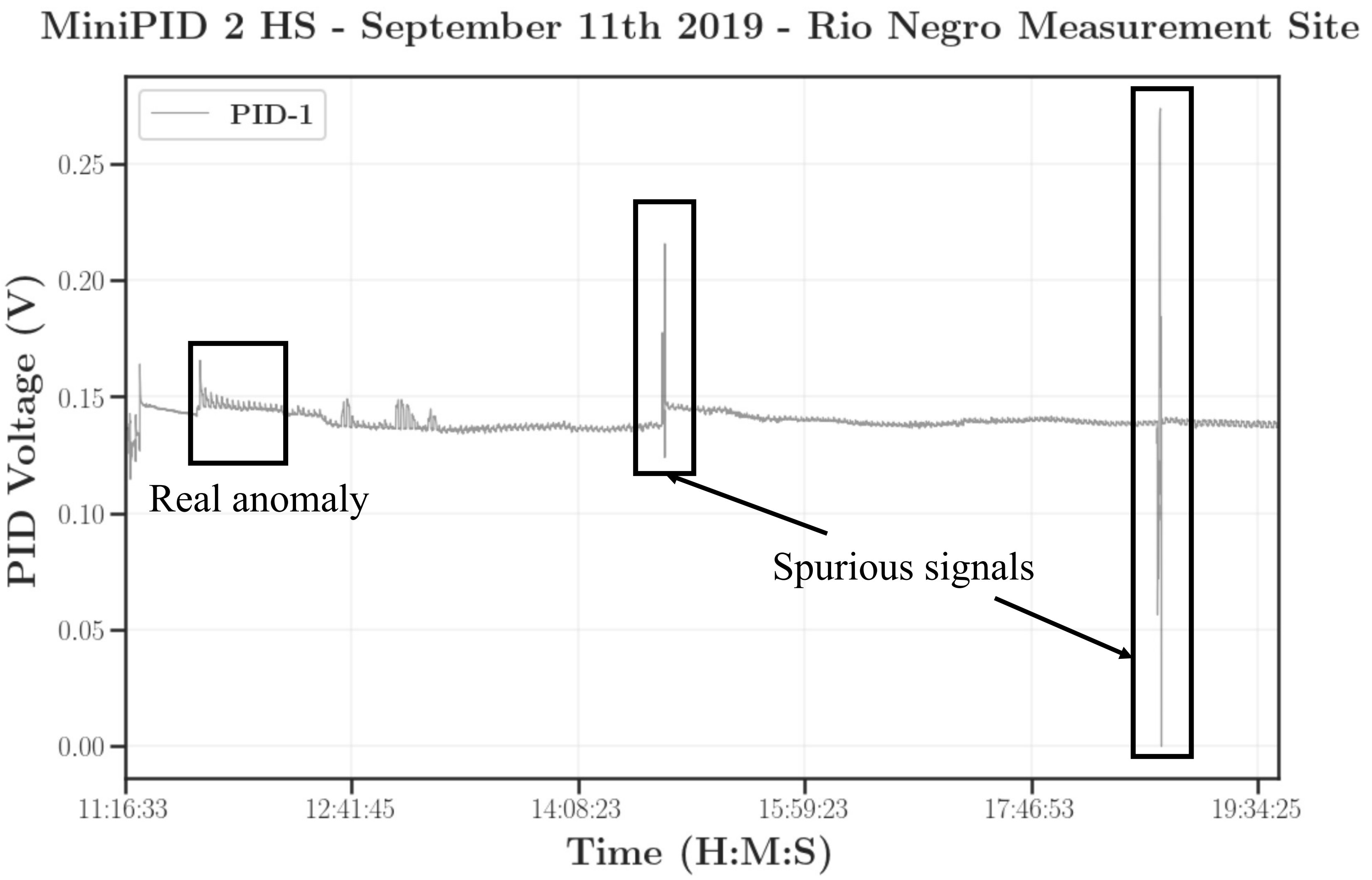
(2) Gas Anomaly Detection

Anomaly detection algorithms can be used to determine real anomalies as opposed to spurious signals.

Pydetect offers 4 anomaly detection algorithms:

- MeanDetector
- VarianceDetector
- MeanVarianceDetector
- GESDDetector

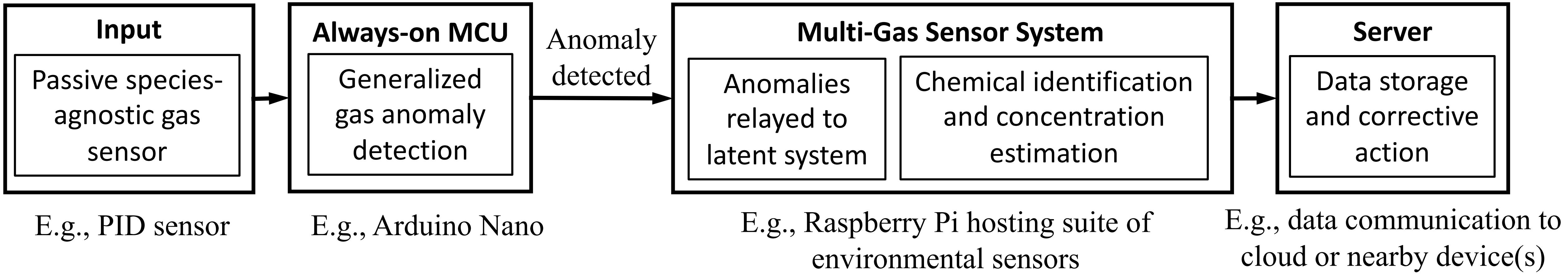
Deep learning methods can also be used for anomaly detections



Applications of TinyML for Atmospheric Monitoring

(3) Multi-Gas Anomaly Detection

- Gas sensor system set up in a cascade architecture
- Upstream segment consists of a passive species-agnostic gas sensor for low-power anomaly detection
- Downstream segment consists of an active multi-gas detection system for chemical identification and concentration estimation



Applications of TinyML for Atmospheric Monitoring

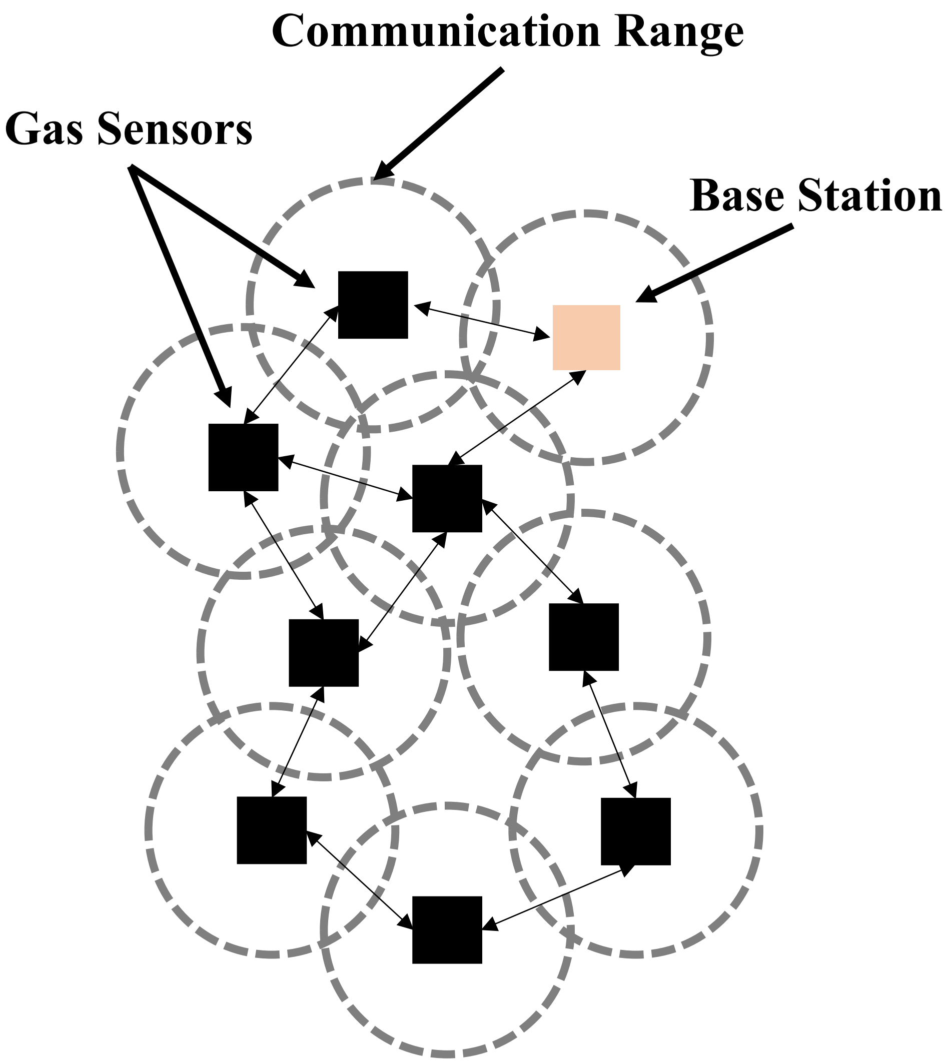
(4) Intelligent Gas Sensor Networks

Applications

- Emissions compliance in industrial facilities
- Air quality monitoring in urban areas – e.g., inner cities, highways, indoor facilities
- Monitoring of forest fires in remote locations

Features

- Only communicate anomalous data
- Device/sensor configuration may be either homogeneous or heterogeneous
- Can be disconnected from a power grid for a finite period for remote monitoring

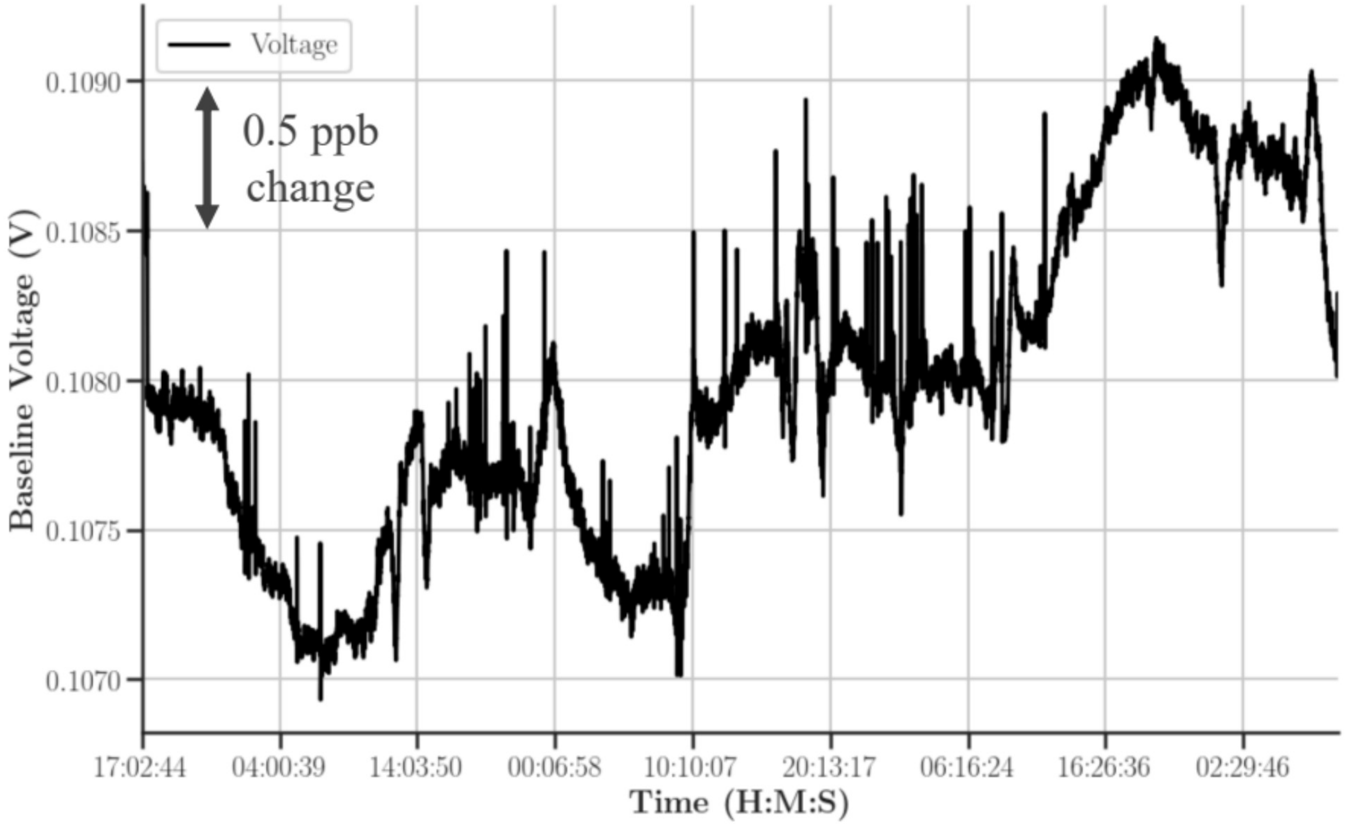


Applications of TinyML for Atmospheric Monitoring

(5) Continuous retraining to accommodate sensor drift (Tentative application)

Any deployed machine learning model will need to be retrained over time to accommodate sensor drift – on-device training is currently difficult to achieve

- Similar to treatment of model drift in traditional MLOps
- More challenging due to limited available resources on device
- Drift may be short- or long-term, long-term drift associated with sensor aging, short term with environmental changes



SciTinyML - ICTP workshop
Scientific Use of Machine Learning on Low Power Devices

Applications of TinyML for atmospheric monitoring

Matthew Stewart
Harvard University, USA
Email: Matthew_stewart@g.harvard.edu

TINY ML

HARVARD
John A. Paulson
School of Engineering
and Applied Sciences