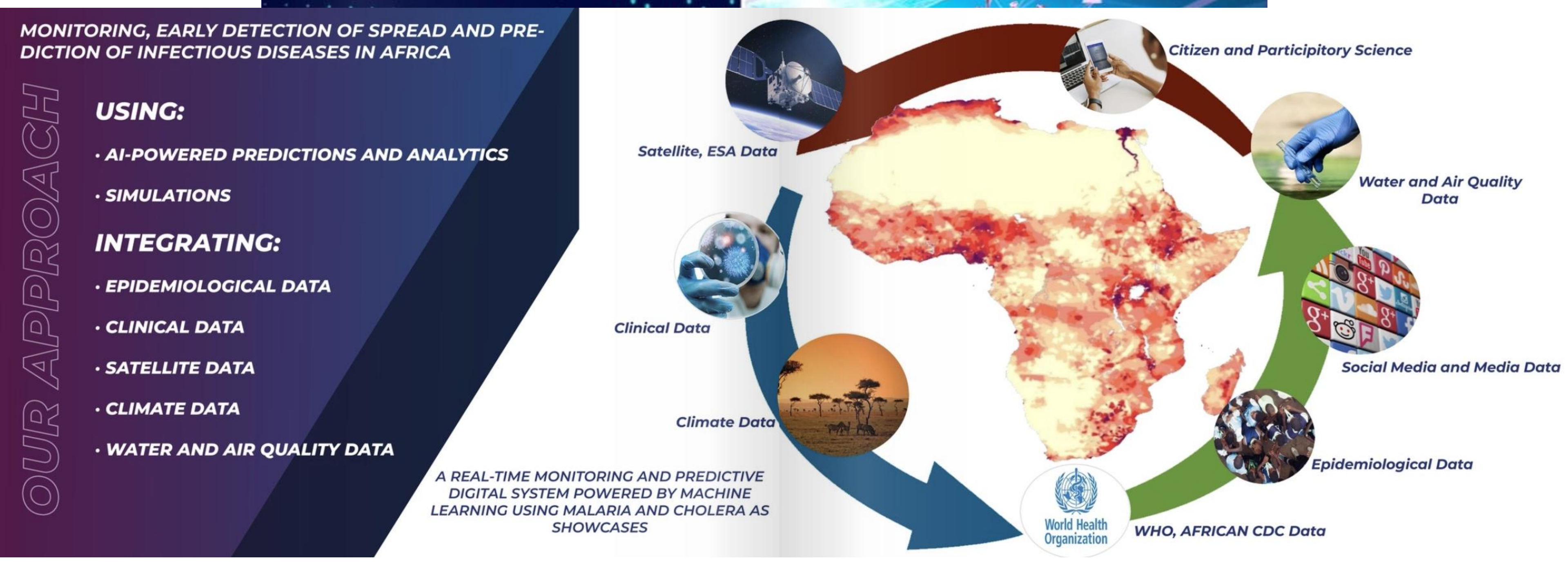
AI-powered Early Detection Systems for Global Public Health

P.Basu, S.Bhattacharya, L.Boateng, A.Cervello, V.Chabalala, L.Chen, J.Choma, S.E.Dahbi, H.Goldblatt, K.Hayasi, J.Kong, B.Lieberman, M.Malaatje, T.Mathaha, R.McKenzie, I.Mpano, Z.Movahedinia, M.Msweli, B.Mellado, C.Nyati, T.Nyeleti, B.Ogbukiri, N.Perikli, C.Rudolph, F.Stevenson, T.Wilkinson

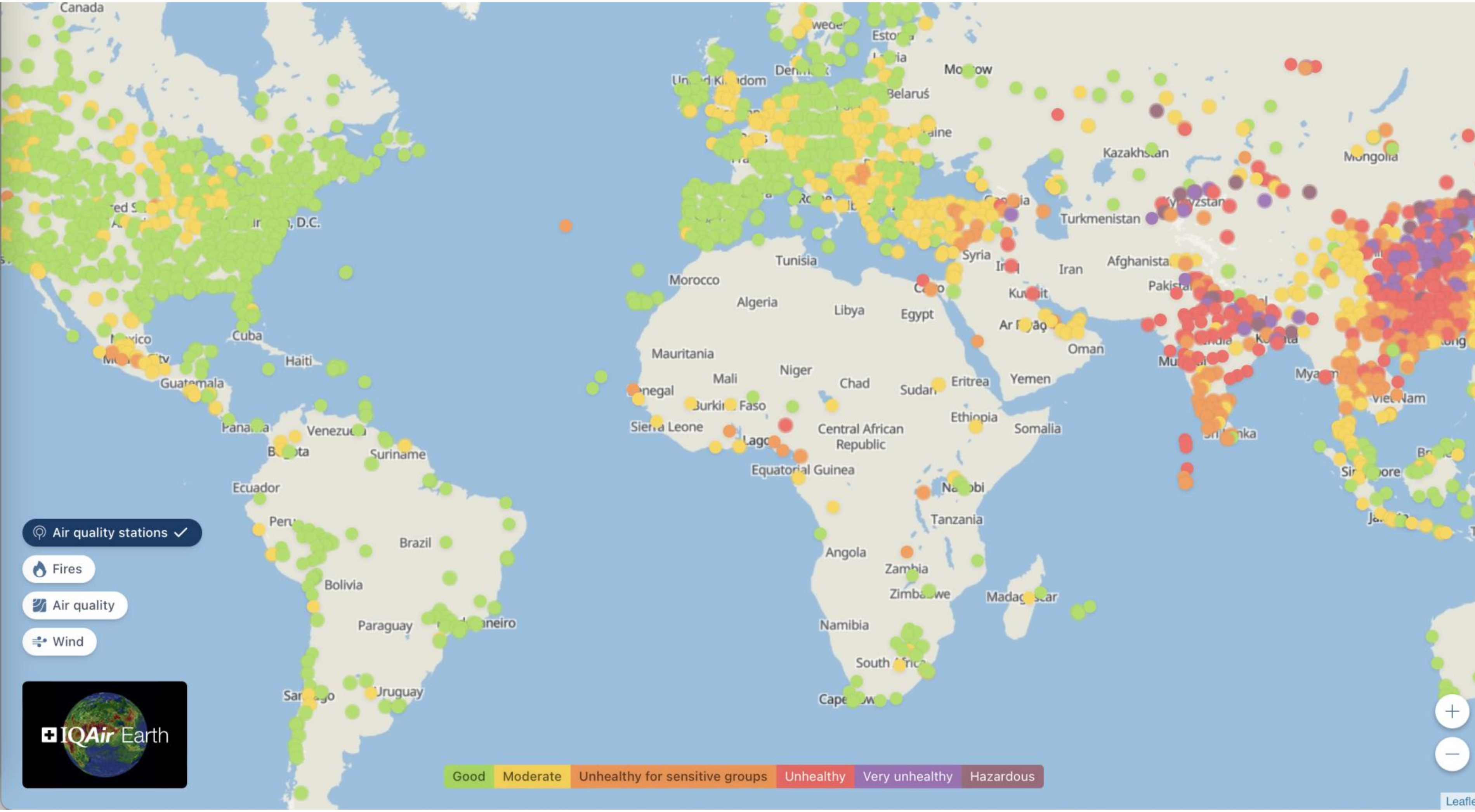
MfPH Workshop on Early Warning Systems for Emerging and Re-emerging Diseases – 23/01/23


AI-POWERED
EARLY DETECTION AND
PREDICTION OF PANDEMICS


MONITORING, EARLY DETECTION OF SPREAD AND PREDICTION OF INFECTIOUS DISEASES IN AFRICA
OUR APPROACH
USING:
• AI-POWERED PREDICTIONS AND ANALYTICS
• SIMULATIONS
INTEGRATING:
• EPIDEMIOLOGICAL DATA
• CLINICAL DATA
• SATELLITE DATA
• CLIMATE DATA
• WATER AND AIR QUALITY DATA
A REAL-TIME MONITORING AND PREDICTIVE DIGITAL SYSTEM POWERED BY MACHINE LEARNING USING MALARIA AND CHOLERA AS SHOWCASES
Satellite, ESA Data
Citizen and Participitory Science
Water and Air Quality Data
Social Media and Media Data
Epidemiological Data
WHO, AFRICAN CDC Data
World Health Organization
Climate Data
Clinical Data

https://www.iqair.com/air-quality-map

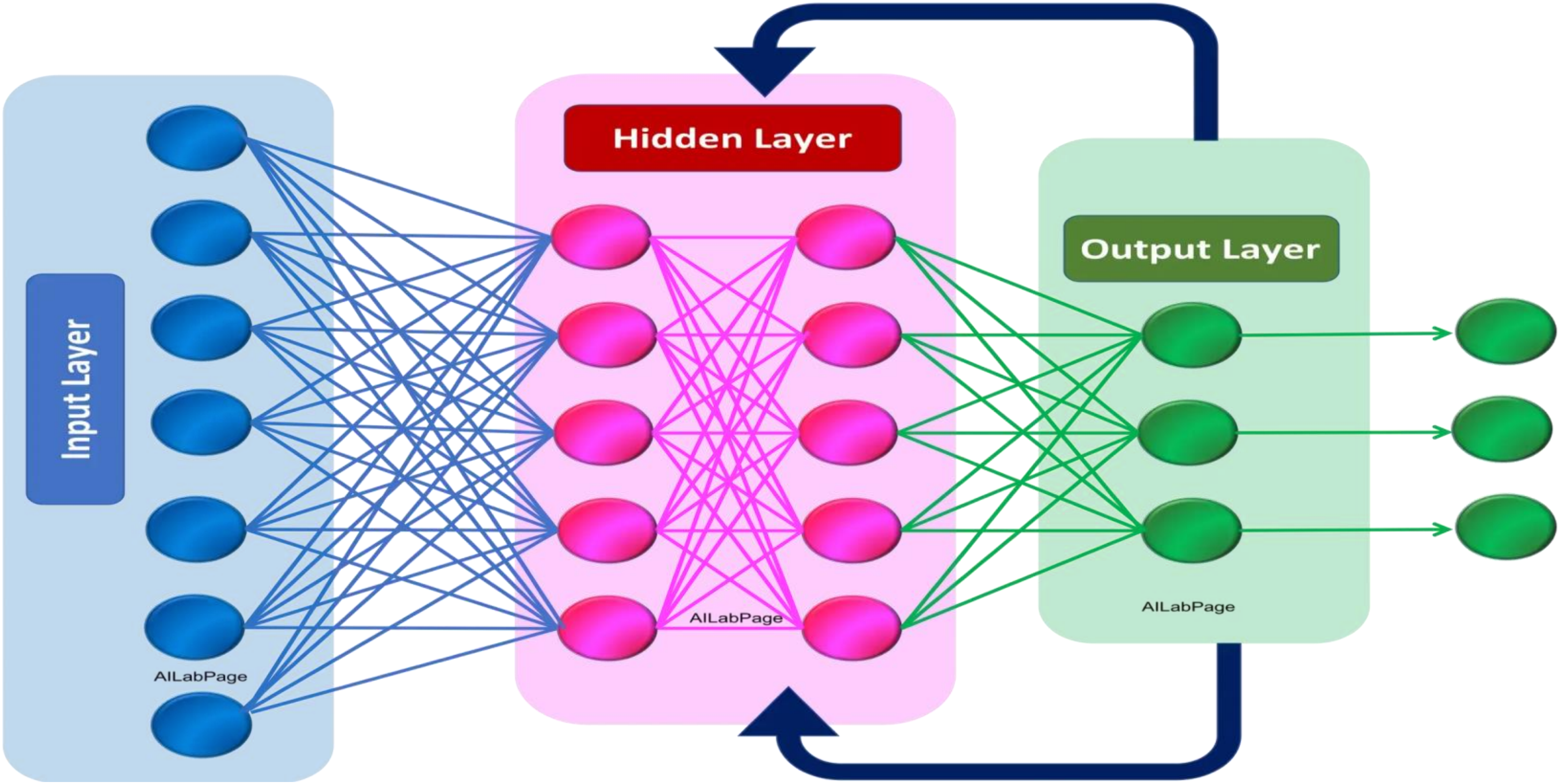
Air quality monitoring stations in the World



An example of the strong imbalance in monitoring between the Global North and South

Early Detection with Machine Learning

Recurrent Neural Networks



The recurrent structure of RNNs enables the following characteristics:

- Specialized for processing a sequence of values $x^{(1)}, \ldots, x^{(\tau)}$
  - Each value $x^{(i)}$ is processed with the **same network *A* that preserves past information**

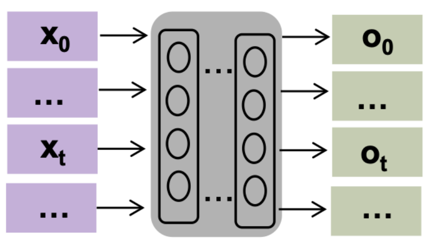
Foundation of Temporal Recurrent Neural Networks

Goal

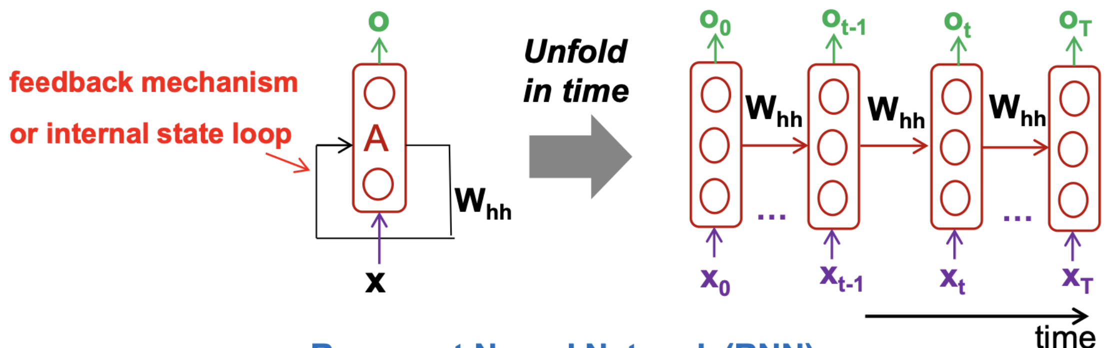
- model long term dependencies
- connect previous information to the present task
- model sequence of events with loops, allowing information to persist

Feed Forward NNets can **not** take **time dependencies** into account.

Sequential data needs a **Feedback Mechanism**.



FF-net / CNN



Recurrent Neural Network (RNN)

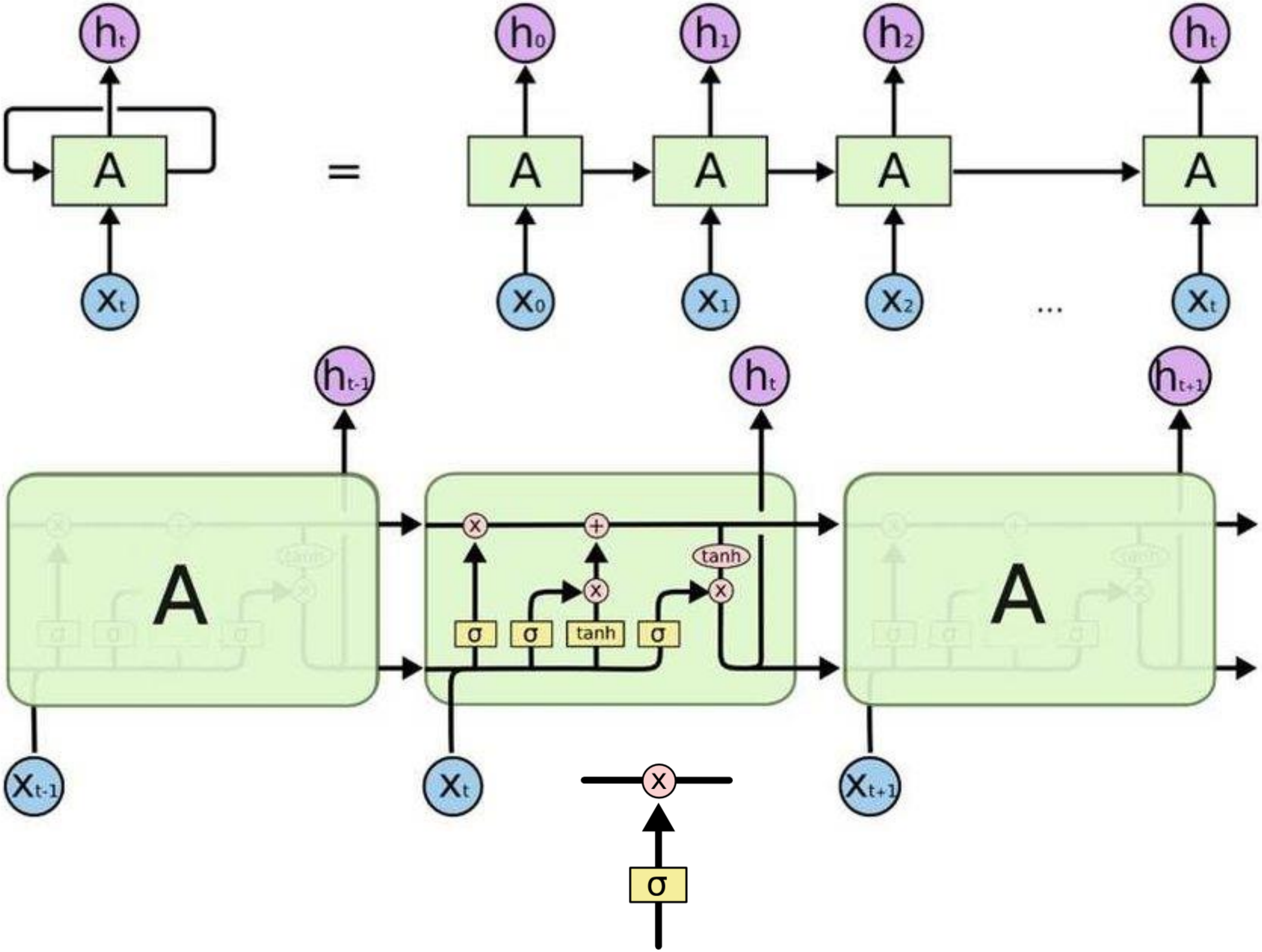
new state

old state

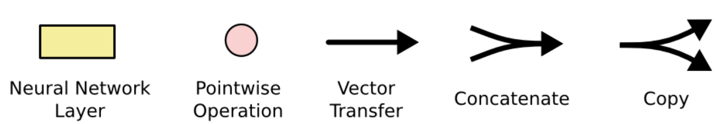
$$\boxed{h_t} = \boxed{f_W}(\boxed{h_{t-1}}, \boxed{x_t})$$

function with parameter W

Input vector at some time step

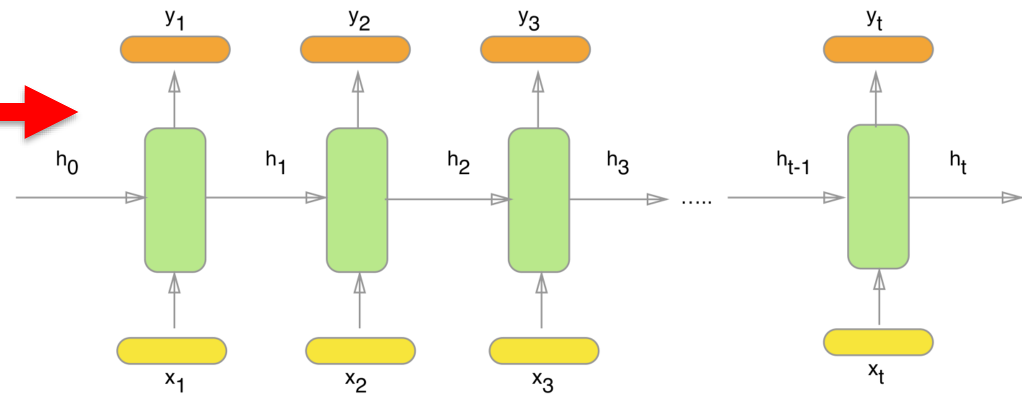


A Sigmoid Gate



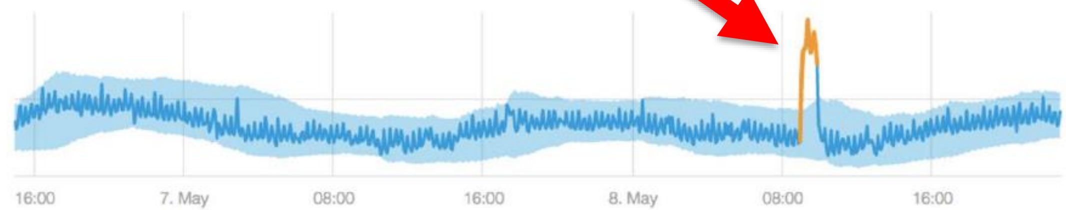
Early Detection and Anomaly Detection

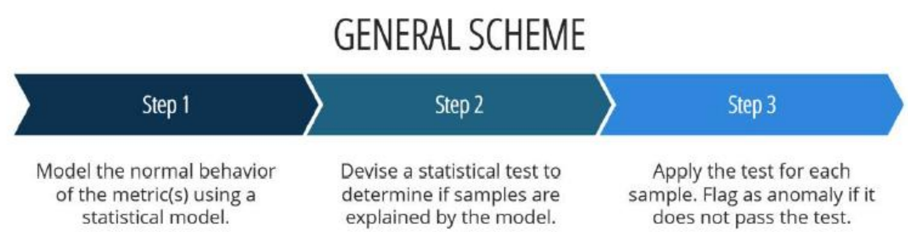
Train the RNN with data that corresponds to periods of absence of crisis. Model is created that describes multi-dimensional data consistent with absence of crisis.



$$f(x_1, \dots x_n, y_1, \dots y_m, t)$$

Anomaly detection





Early detection relies on detection of departure from expected behavior of the data, or anomaly detection. The more accurate the model, the earlier the detection.

Early detection algorithm of new waves used for planning

Early Detection of COVID-19 Waves

Requested by Government and the South African National Defense Force in order to assess the level of risk of new waves

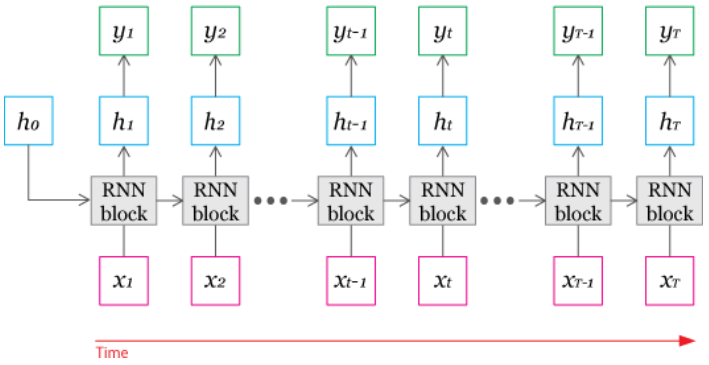
Use of classical analytics and AI.

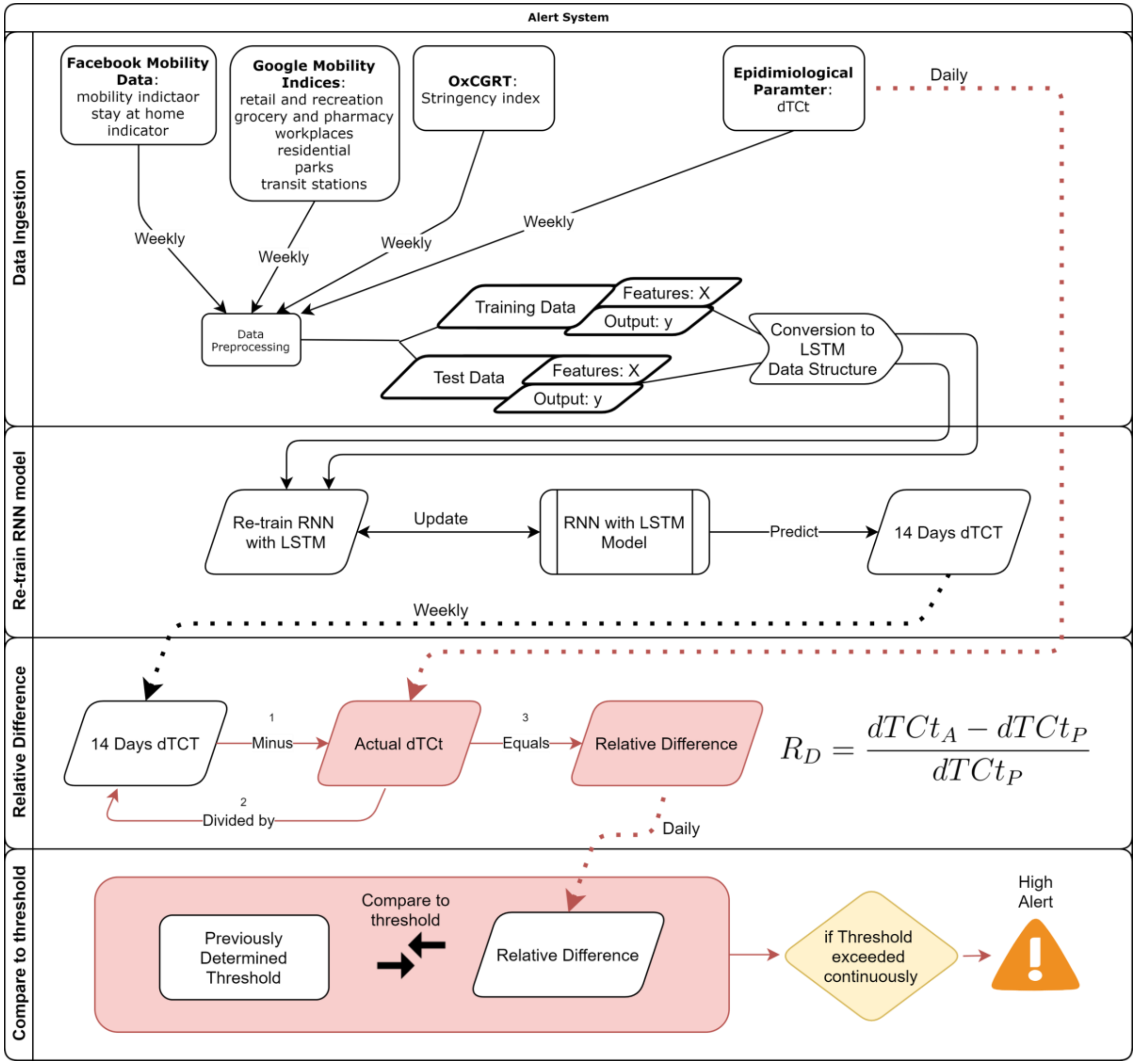
Int. J Environ Res Public Health 2021 Jul 9;18(14):7376 doi: 10.3390/ijerph18147376.

https://www.covid19sa.org/riskindex-ai

Released a new early detection algorithm based on Artificial Intelligence

Created a risk index that helps understand whether the data is consistent with low or high risk of a new wave

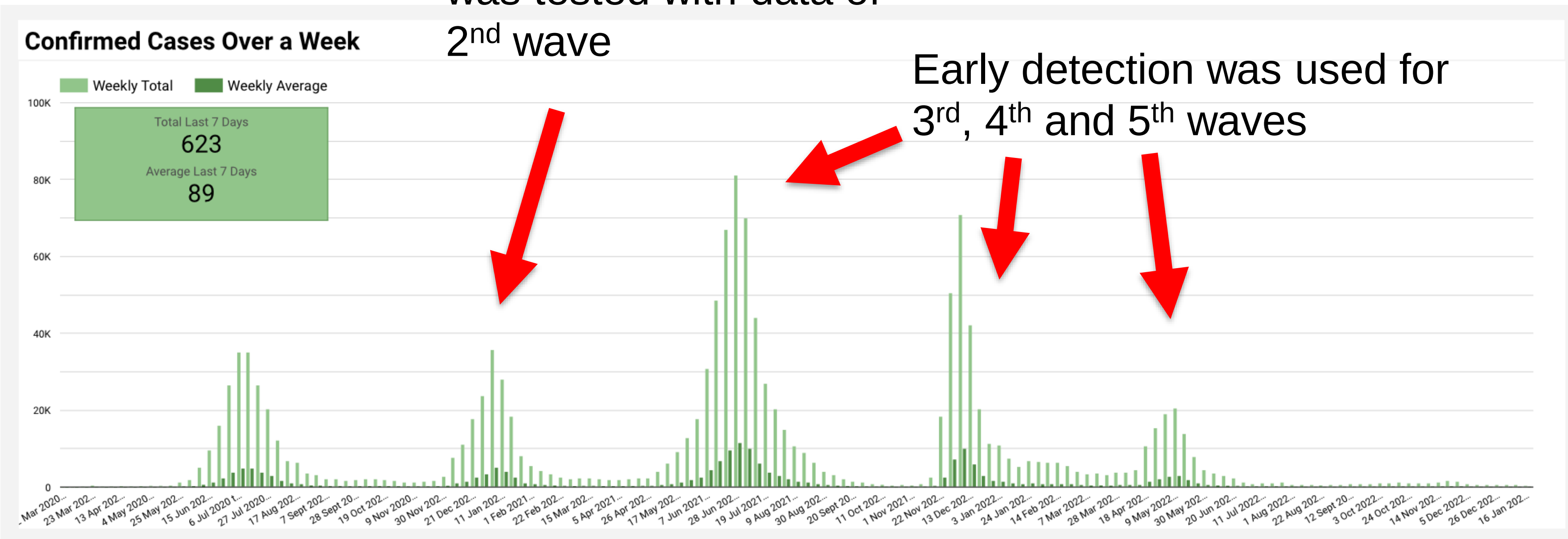




Int. J Environ Res Public Health 2021 Jul 9;18(14):7376 doi: 10.3390/ijerph18147376.

Performance of early detection algorithm was tested with data of 2nd wave

Early detection was used for 3rd, 4th and 5th waves



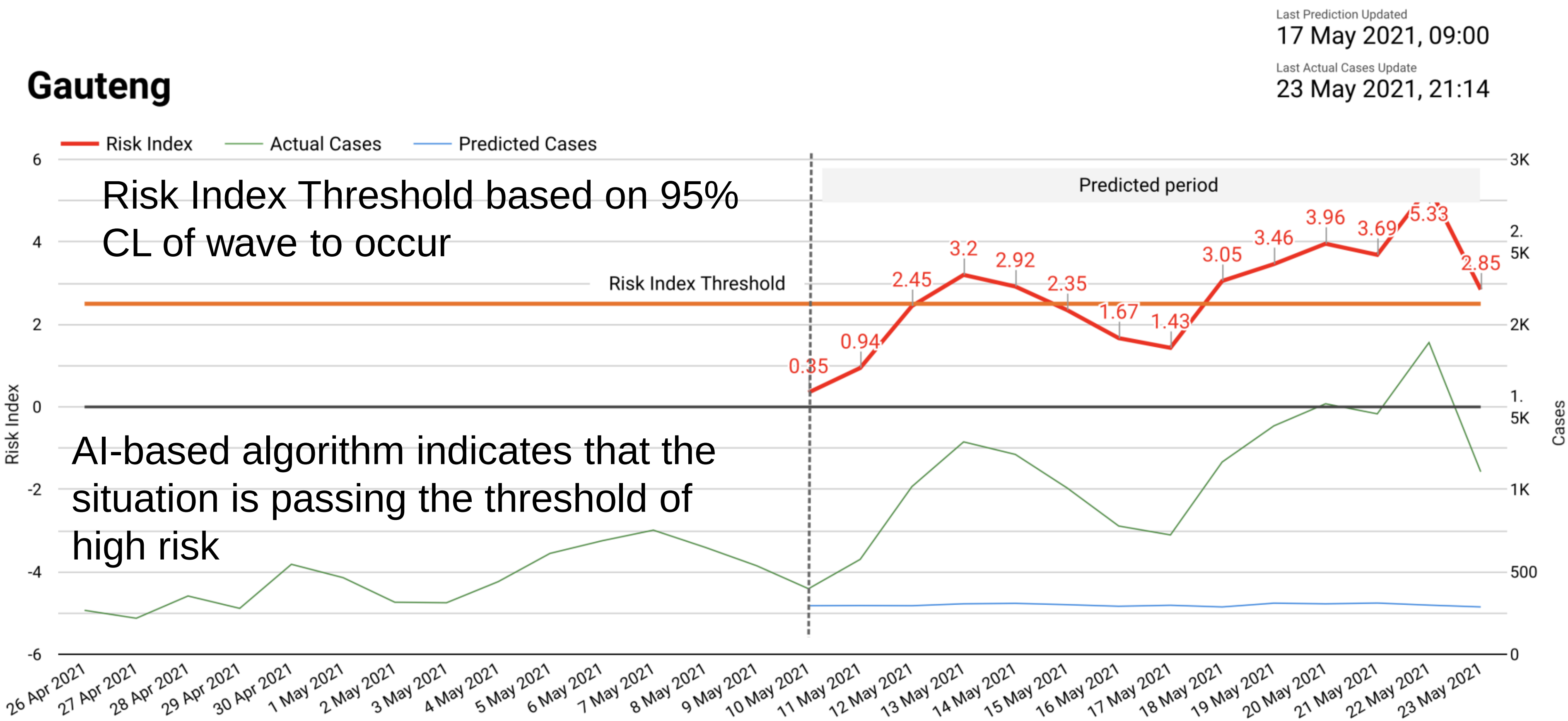
LSTM algorithm trained with multi-dimensional data of periods in between the 1st-2nd and 2nd-3rd waves.

Early detection does not rely on the specifics of the anomaly. As such, the model was able to detect the presence of new variants even before the genomics team.

Gauteng – 3rd wave risk index powered by AI

https://www.covid19sa.org/riskindex-ai

Risk index of 3rd wave powered by AI.
Index updated daily. Each province has specific model with specific thresholds

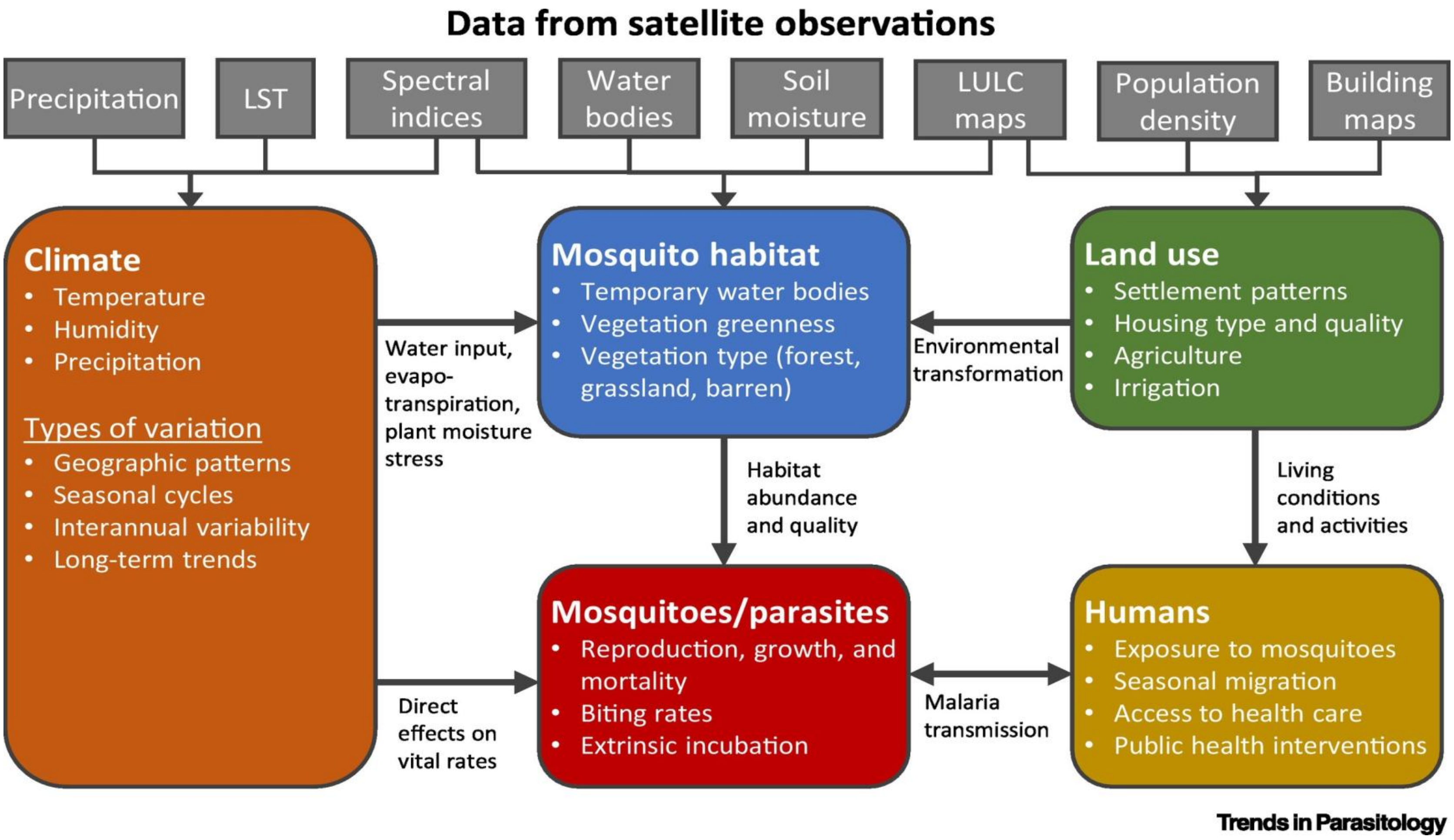


Satellite data and early pandemic detection with AI: Malaria, a Showcase

Background

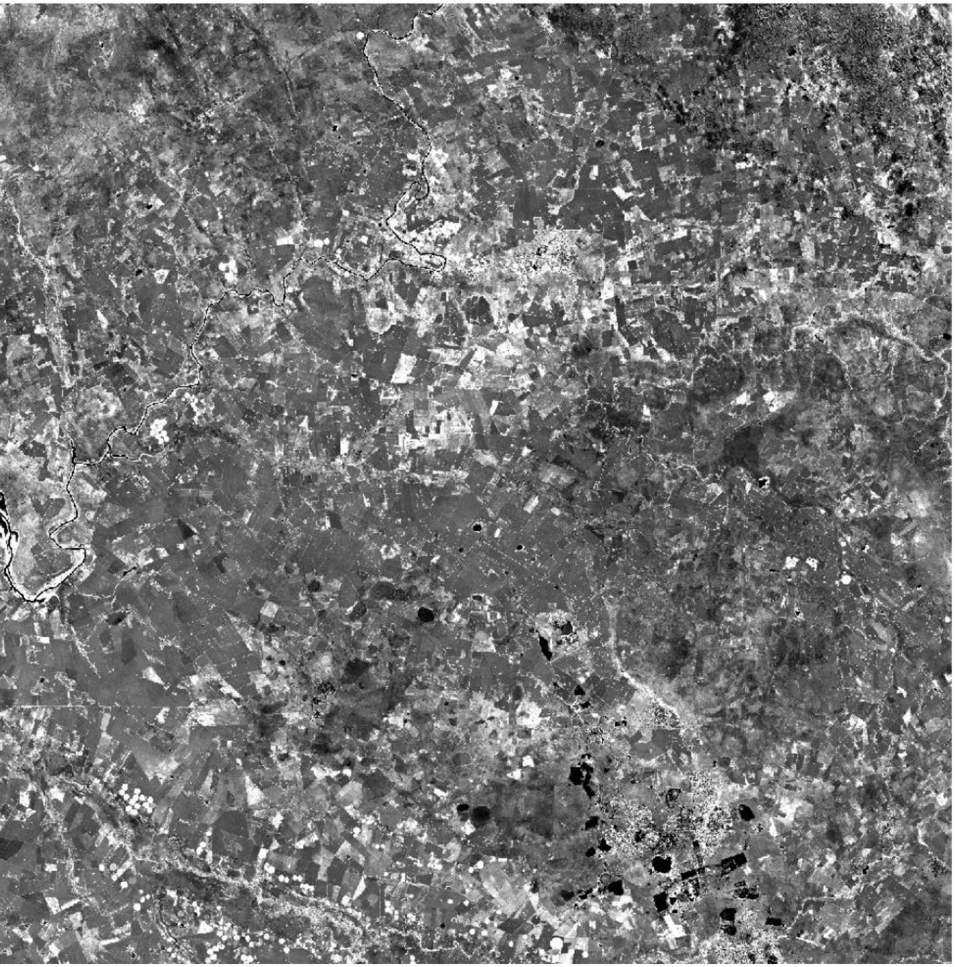
- Malaria is a disease caused by a parasitic protozoan called Plasmodium in mosquitoes. This Plasmodium can then spread to humans when female Anopheles mosquitoes feed on human blood. Once the Plasmodium is in the human body it causes malaria. Female Anopheles are the primary breeders and spreaders of the Plasmodium parasite (Miller, 2002).
- This means that the more conditions are favourable for carrier mosquitos to reproduce the better it is for the Plasmodium's survival. Malaria can spread faster when conditions are favourable for the female Anopheles to breed.
- Motivation for the study: While malaria is curable and treatable it is still one of the leading causes of death in sub-Saharan Africa. Africa accounts for 95% of the total cases globally and 96% of global deaths due to malaria. Cases of malaria have increased from 213 million to 228 million from 2019 to 2020 (WHO, 2021).

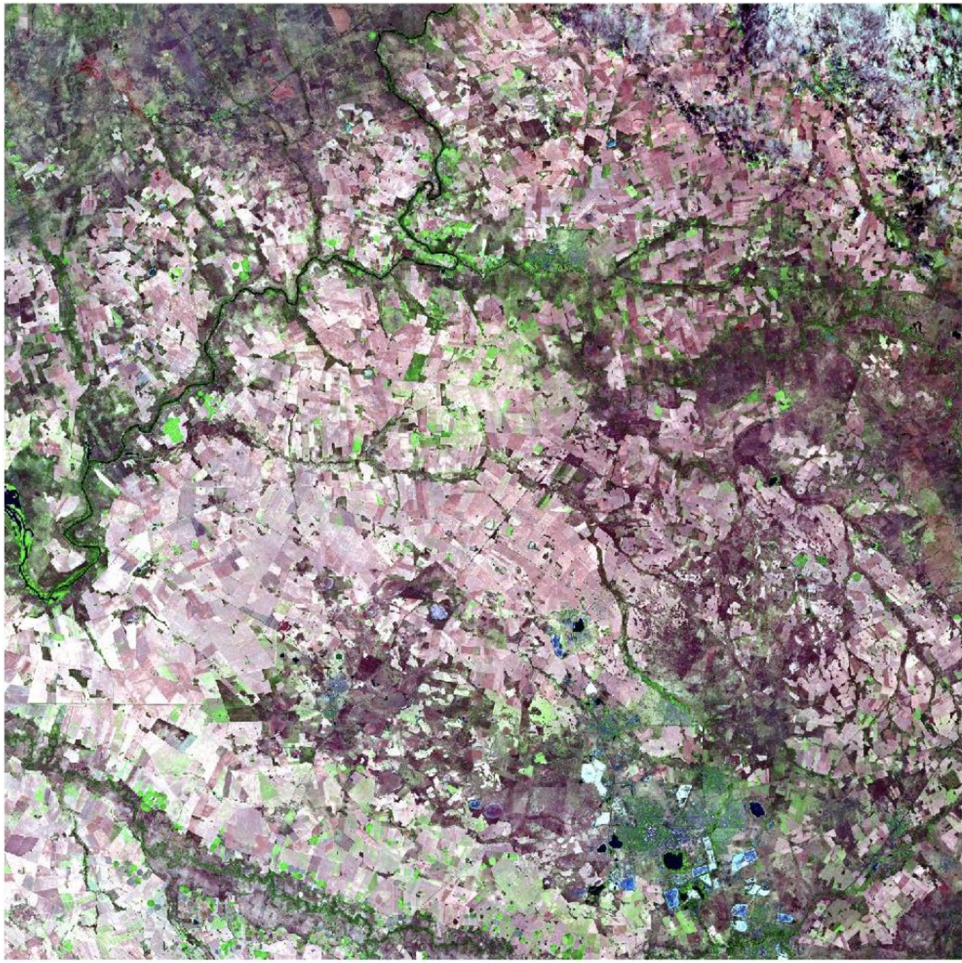
Satellites from the European Space Agency continuously scan the African continent. This data is known to give sufficient information as to the habitat of mosquitos, which is essential to monitor and predict malaria outbreaks at the Pan-African level using Big Data and Artificial Intelligence.



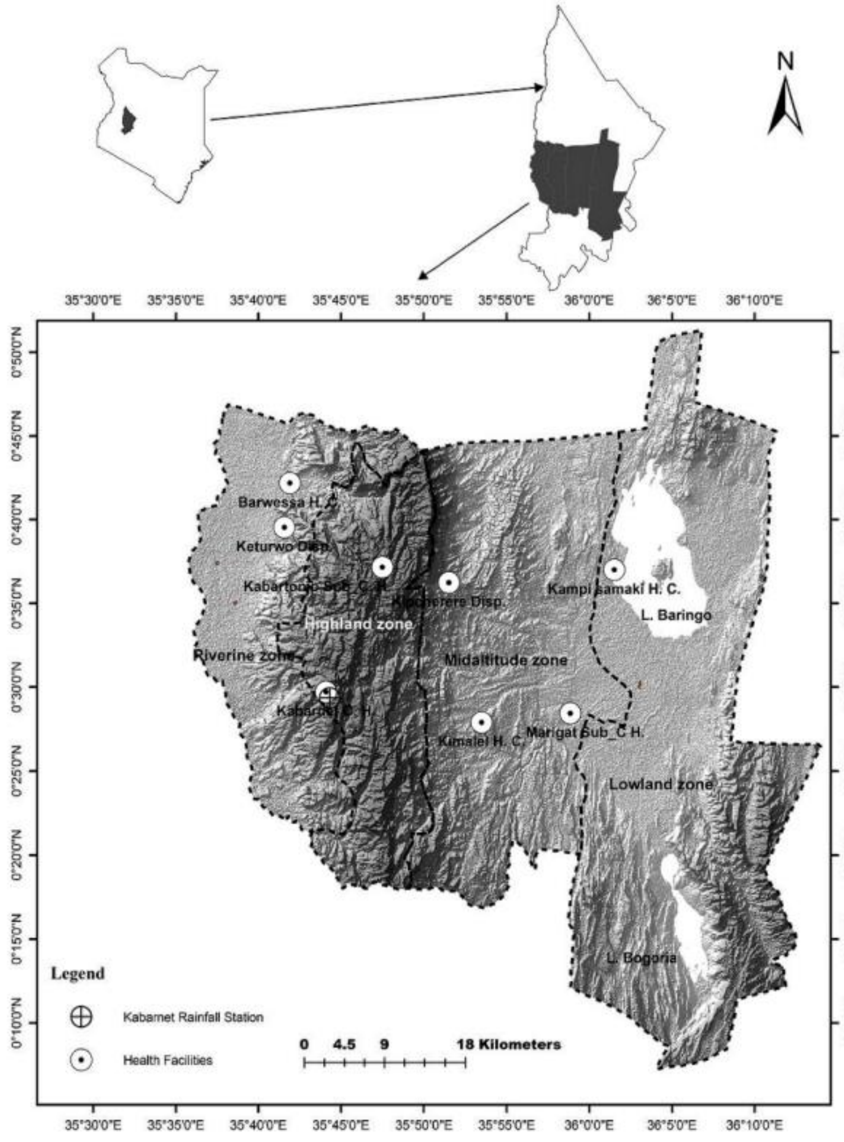
Massive amounts of data from the Sentinel satellites are available and remained untapped for Malaria monitoring and predicting Malaria outbreaks.

Below are examples of Satellite data processed for the extraction of vegetation indexes.





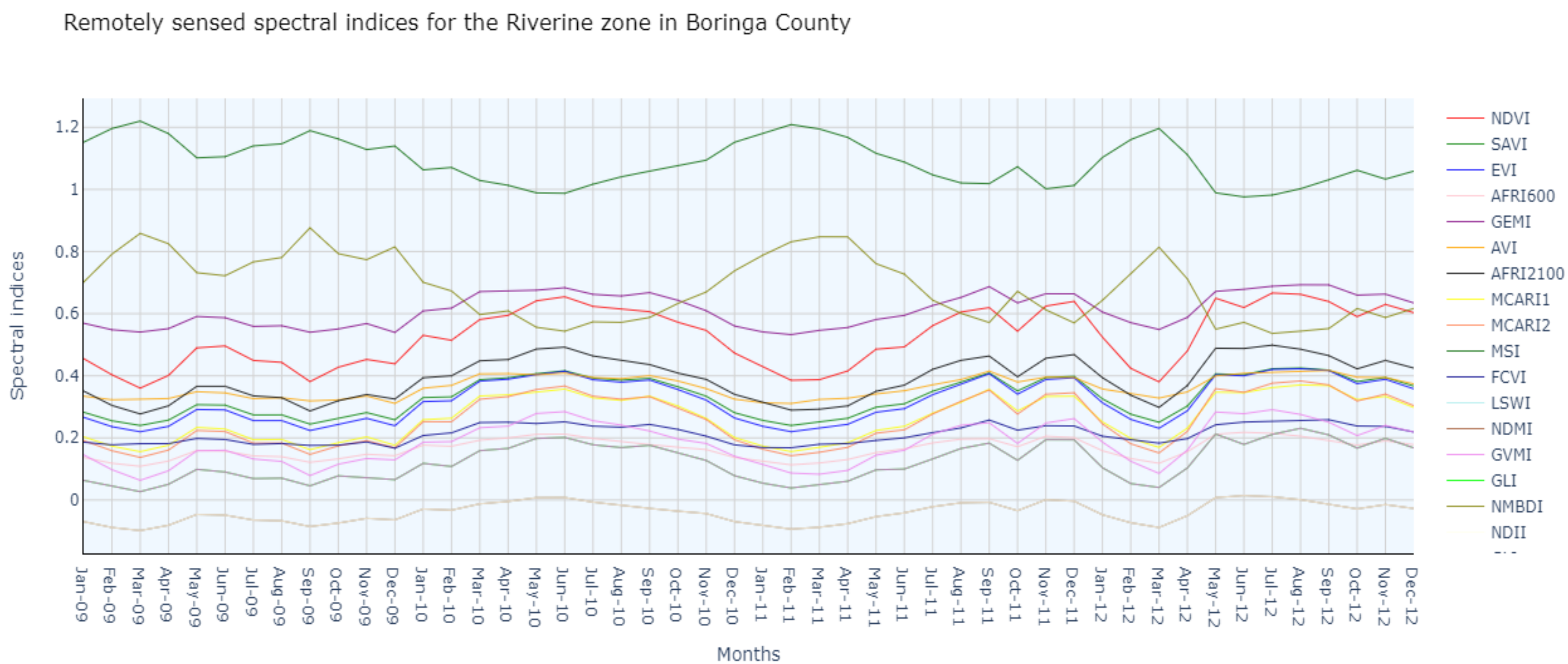
Showcase: Kenya



Satellite Data

Bands	Standard	Sentinel-1	Sentinel-2	Landsat-89	Landsat-457	MODIS
Aersols	A		B1	B1		
Blue	B		B2	B2	B1	B3
Green	G		B3	B3	B2	B4
Red	R		B4	B4	B3	B1
Red Edge 1	RE1		B5			
Red Edge 2	RE2		B6			
Red Edge 3	RE3		B7			
NIR	N		B8	B5	B4	B2
NIR 2	N2		B8A			
SWIR 1	S1		B11	B6	B5	B6
SWIR 2	S2		B12	B7	B7	B7
Thermal 1	T1			B10	B6	
Thermal 2	T2			B11		
Polarization	HV	HV				
Polarization	VH	VH				
Polarization	HH	HH				
Polarization	VV	VV				

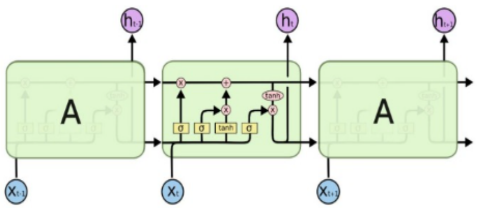
Satellite-Derived spectral indices for Riverine ecological zone



The custom LSTM model

LSTM model

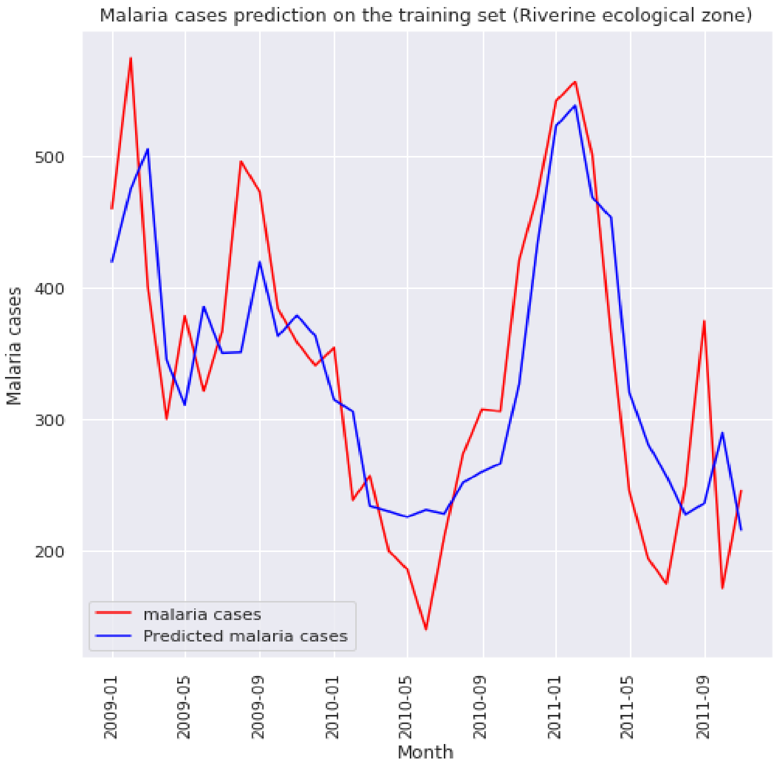
Layers	Activation function	Layer Size
LSTM	Relu	64
LSTM	Relu	32
LSTM	Relu	16
Dropout	-	Custom
Dense	-	1

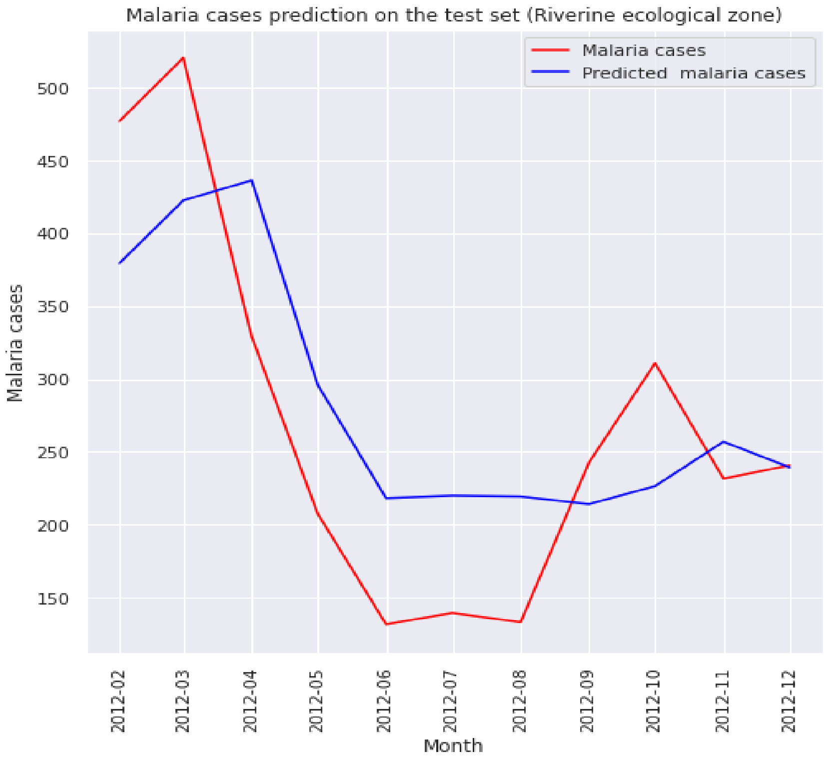


Parameters	
Epochs	2000
Batch Size	4
Learning Rate	0.0001
Drop out rate	0.2

Training and testing parameters for LSTM model

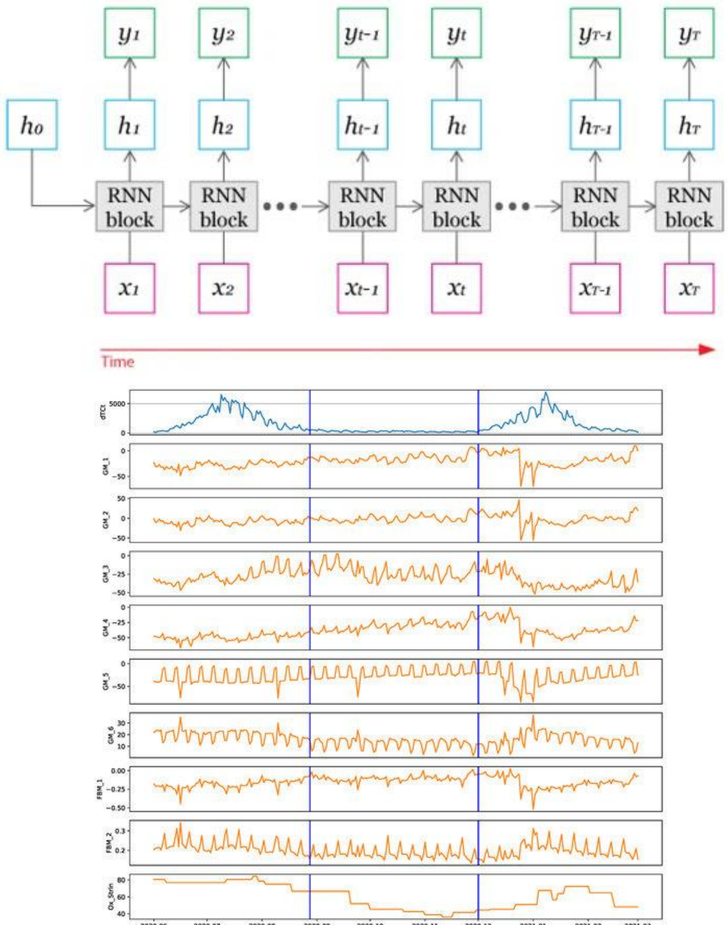
Riverine ecological zone training and testing results from the LSTM model



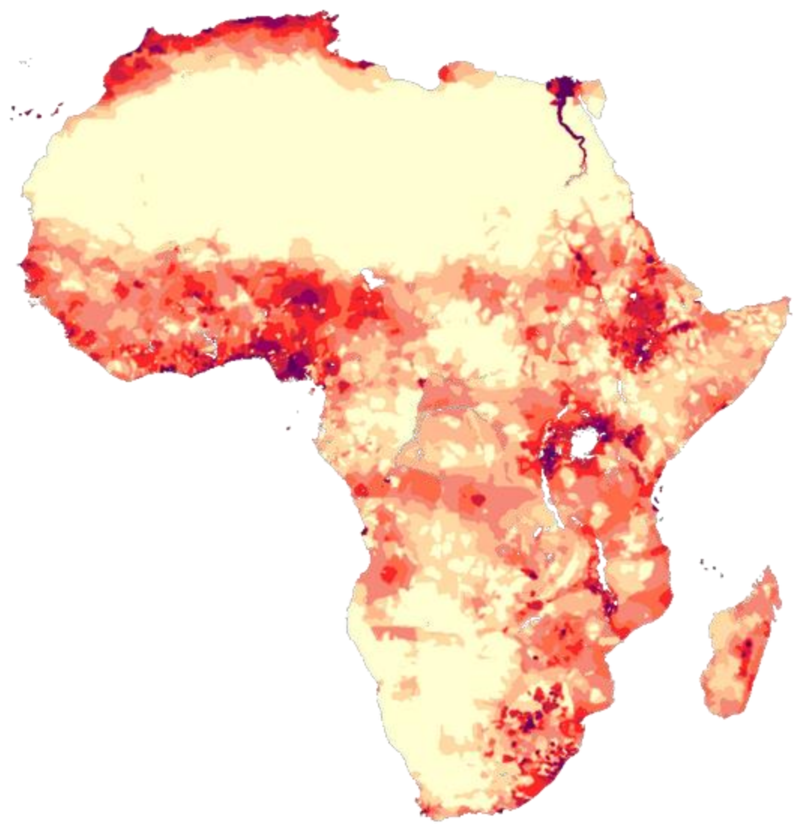


Our team has years of experience processing Satellite Data from the European Space Agency for different projects that range from urban population studies to vegetation indexes studies. Artificial Intelligence are used to process the data automatically in order to monitor vast amounts of territory.

<u>Our team will adapt Artificial Intelligence tools that were developed for COVID-19 for prediction of outbreaks with which to create an African map of Malaria, including predictions for outbreaks</u>



Proposed Malaria map of Africa




Air Quality Monitoring for the Global South
Solving the affordability problem with AI and IoT

Why Air Quality Matters

-It is an effective probe into a broad range of public health issues.

-It is important to manage public health responses and governance.

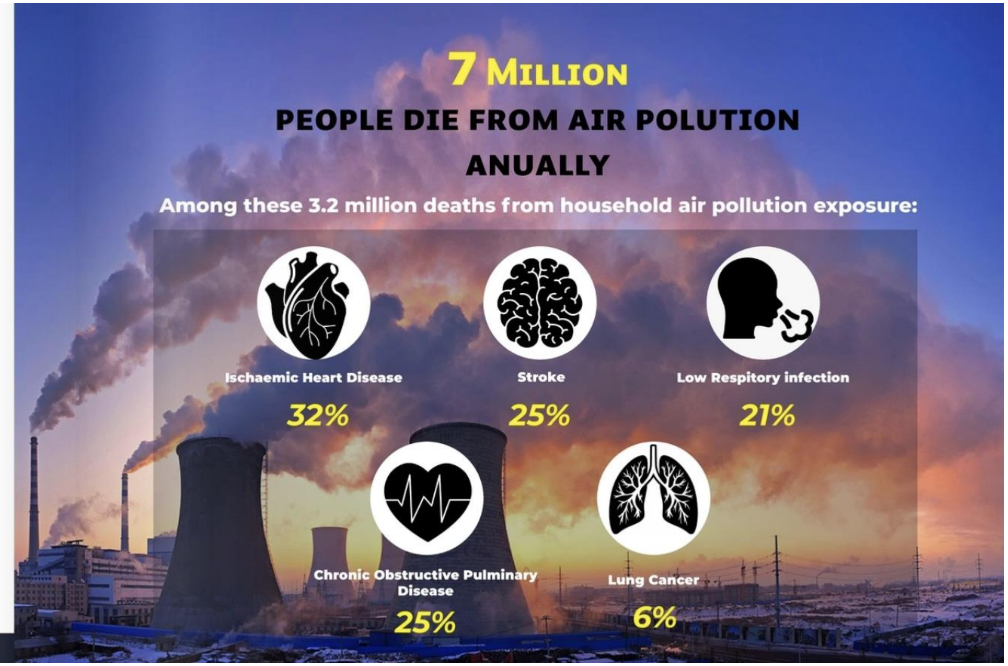
-Poor air quality is known to be linked to chronic conditions and other health-related issues.

-Reducing air pollution levels, countries can reduce the burden of disease from stroke, heart disease, lung cancer, and both chronic and acute respiratory diseases, including asthma.

"In 2019, 99% of people were found to breath air that exceeds World Health Organisation (WHO) air quality guidelines.

Each year, 7 million people die prematurely from illnesses attributable to the household air pollution caused by the incomplete combustion of solid fuels and kerosene used for cooking

This includes more than 1.7 million child deaths a year worldwide. "- says WHO



Social vulnerability

Poverty fosters increased air pollution

Household air pollution is generated by the use of inefficient and polluting fuels and technologies in and around the home that contains a range of health-damaging pollutants, including small particles that penetrate deep into the lungs and enter the bloodstream. In poorly ventilated dwellings, indoor smoke can have levels of fine particles 100 times higher than acceptable. Exposure is particularly high among women and children, who spend the most time near the domestic hearth. Reliance on polluting fuels and technologies also require significant time for cooking on an inefficient device, and gathering and preparing fuel.

Most sources of outdoor air pollution are well beyond the control of individuals and demands concerted action by local, national and regional level policy-makers working in sectors like transport, energy, waste management, urban planning, and agriculture. Poor Air transcends the direct impact on health.

Air quality is a measure of social vulnerability which is linked with poor service delivery and vunerability towards diseases including communicable diseases.

Consortium of Air Quality Monitoring

The International consortium was founded with the goal of bringing together government institutions, research institutions, and private enterprises into a mutually beneficial ecosystem in order to deliver a low-cost intelligent IoT air quality monitoring system.

The consortium consists of a footprint of global partners from the US, Switzerland and Norway and including researchers from South Africa. Supported by the Canadian, South African and Swiss governments.

OUR MISSION

SACAQM

The consortium is developing a novel low-power wireless air quality monitoring and analysis system that will improve public health. It will also support the global effort to achieve carbon neutrality goals. The consortium was founded with the goal of bringing together government institutions, research institutions, and private enterprises in a mutually beneficial ecosystem in order to deliver a low-cost intelligent IoT air quality monitoring system.

Global Partners

- Africa | Canada Artificial Intelligence and Data Innovation Consortium
- WITS University
- Gauteng Provincial Government, Republic of South Africa
- Gauteng Province, Office of the Premier, Republic of South Africa
- CERN
- SAAQIS – South African Air Quality Information System
- science & innovation – Department: Science and Innovation, Republic of South Africa
- forestry, fisheries & the environment – Department: Forestry, Fisheries and the Environment, Republic of South Africa
- Vniversitat de València
- Dahdaleh Institute for Global Health Research
- Global South AI4PEP Network / Sud Global Réseau AI4PEP
- Perovskia
- Nordic Semiconductor
- University of Zululand
- York U
- Schweizerische Eidgenossenschaft / Confédération suisse / Confederazione Svizzera / Confederaziun svizra
- healthmetryx
- NRF National Research Foundation – iThemba LABS, Laboratory for Accelerator Based Sciences
- Canada
- Sida
- IDRC | CRDI – International Development Research Centre / Centre de recherches pour le développement international

healthmetryx

New Data. Better Decisions.

Analysis of exhaled breath carries very important information for early detection systems.

Low-cost breathing device coupled with cloud-driven AI at the core of the system.

Respiratory Cloud Machine

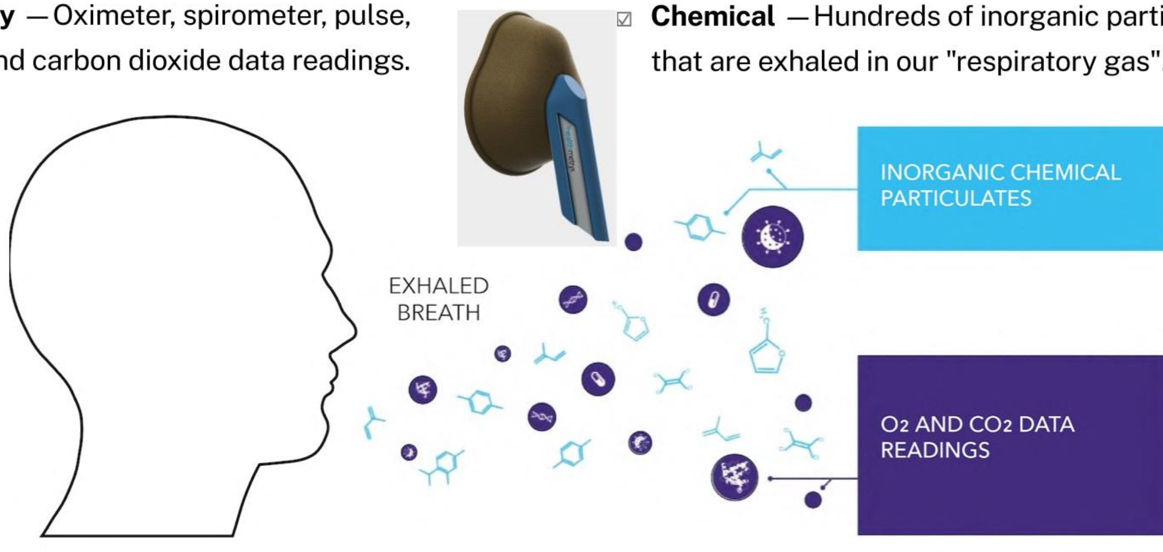
Differentiation

Two Proprietary RCM Data Sets

The company has data collection programs and partners. Leveraging a non-invasive, advanced sensor, breath collection device, the company originates two types of data:

☑ **Pulmonary** — Oximeter, spirometer, pulse, oxygen, and carbon dioxide data readings.

☑ **Chemical** — Hundreds of inorganic particulates that are exhaled in our "respiratory gas".



https://www.sacaqm.org

Need to turn this into a Pan-African consortium

Home About AI_r Projects Partners News Contact Search...

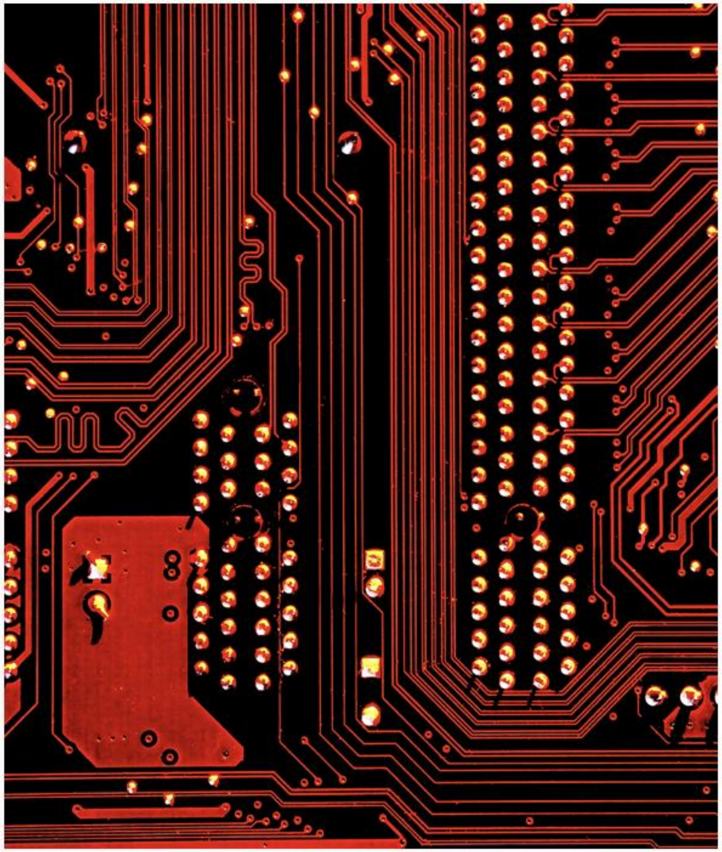
The South African Consortium of Air Quality Monitoring

Solving the affordability problem with AI-powered IoT

Why does air quality matter?

- It is an effective probe into a broad range of public health issues.
- It is important to manage public health responses and governance.
- By reducing air pollution levels, countries can reduce the burden of disease on their

Our Solution: AI-powered IoT



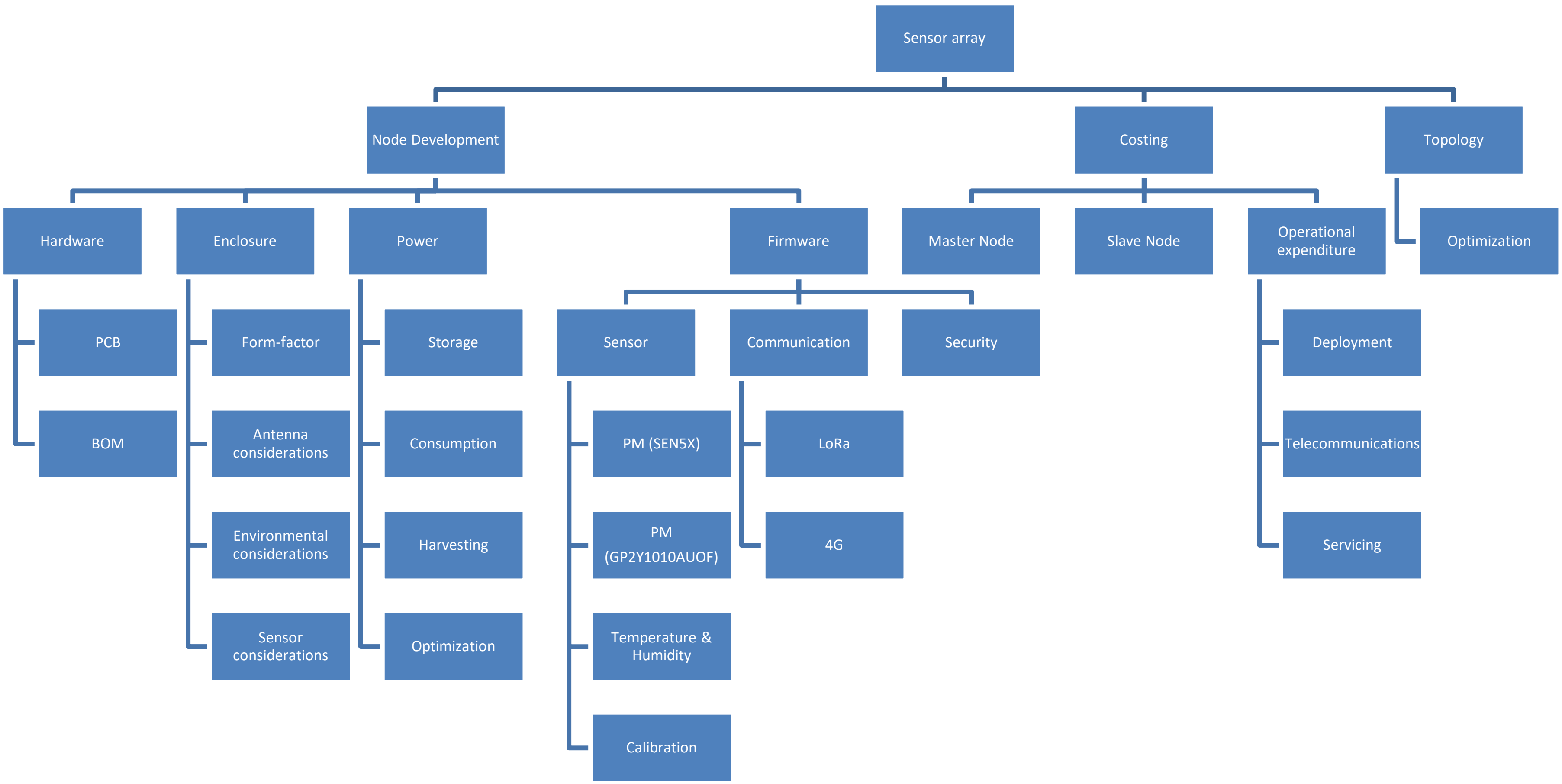
The AI_r system

An industry disrupting air quality monitoring, analysis and prediction system. It combines state-of-the-art air quality sensors with a low-cost Internet-of-Things (IoT) network architecture powered by Artificial Intelligence (AI).

Read more

IoT Development

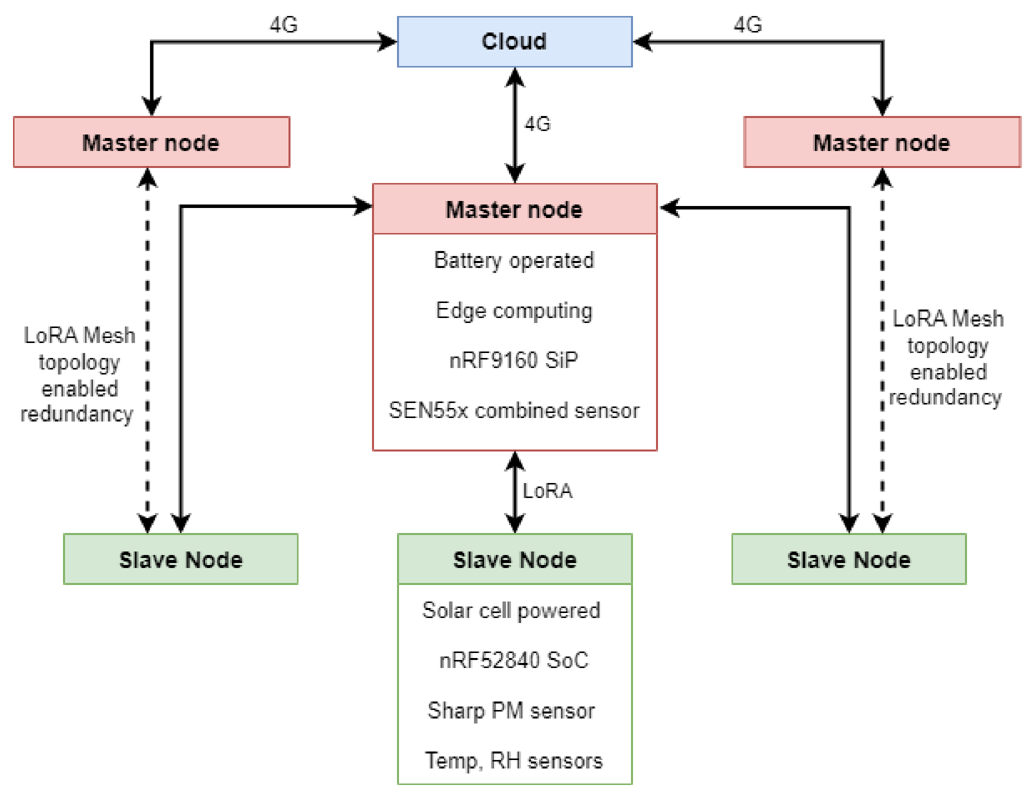
AI_r System Sensor Array Development Overview



Systems building is a problem of systems engineering. This requires interdisciplinary and transdisciplinary collaboration.

AI_r System High Level IoT Topology

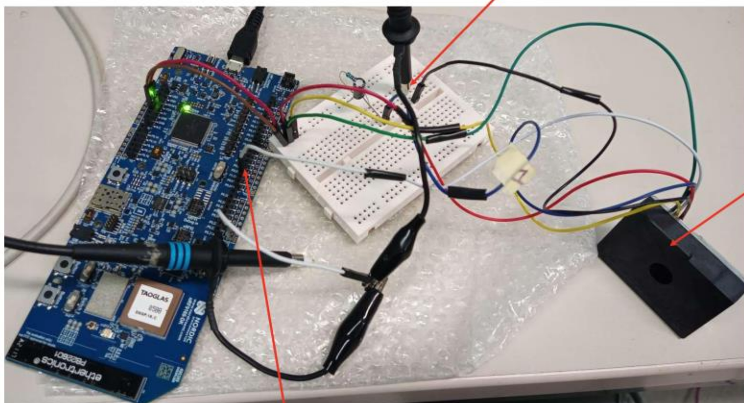
- ❑ **A result of optimization studies to provide maximized price vs coverage and accuracy/precision ratio.**
- ❑ **Master Node – Higher accuracy/precision, 4G communication to cloud via existing network infrastructure.**
- ❑ **Slave Node - Lower cost, increased sensor array coverage.**



GP2Y1010AUOF Testing - Results

GP2Y1010AU0F PM Sensor - nRF9160DK experimental setup

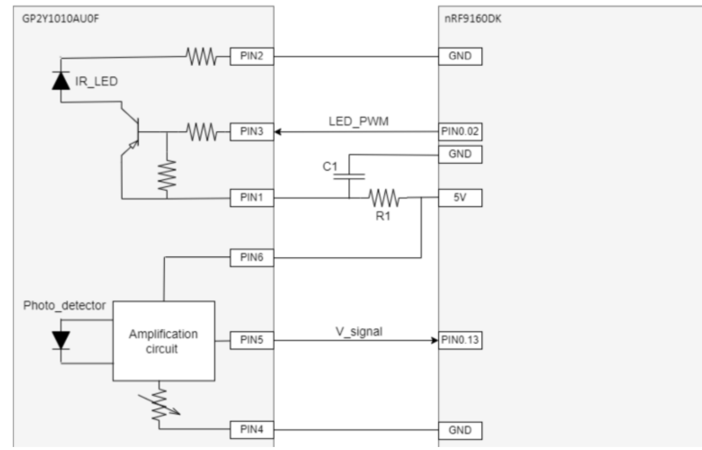
Ch2: V_signal



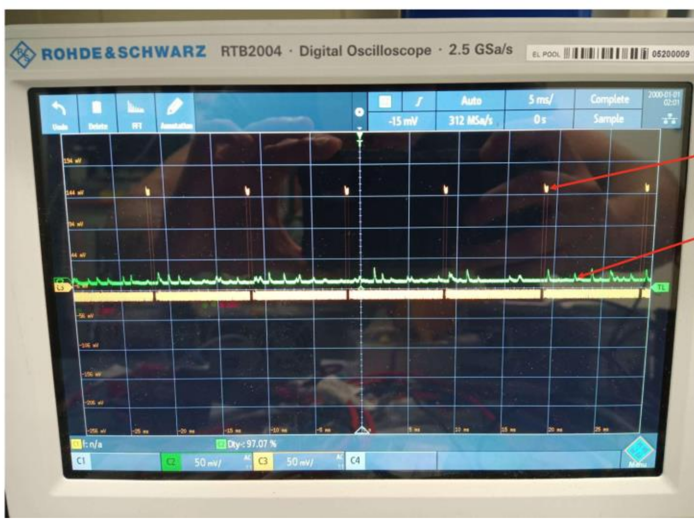
GP2Y1010AU0F

Ch3: PWM input

GP2Y1010AU0F PM Sensor - nRF9160DK hardware connection



4

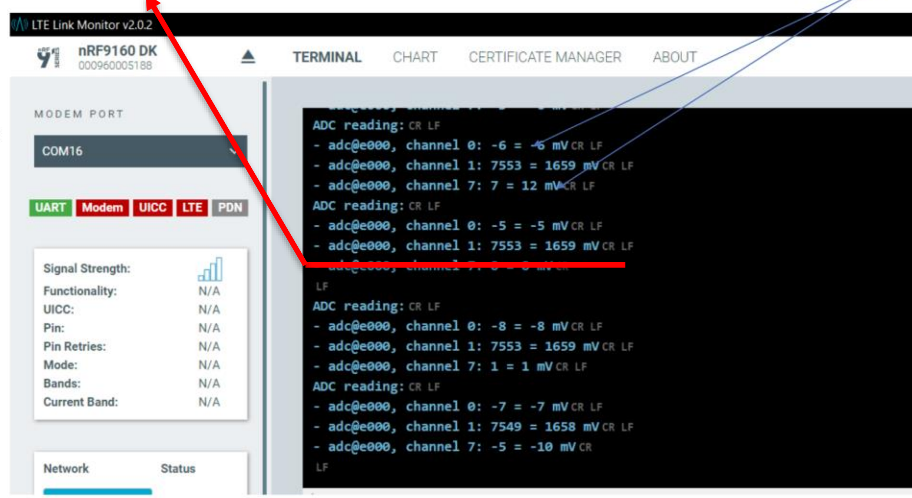


Ch3: PWM input

Ch2: V_signal

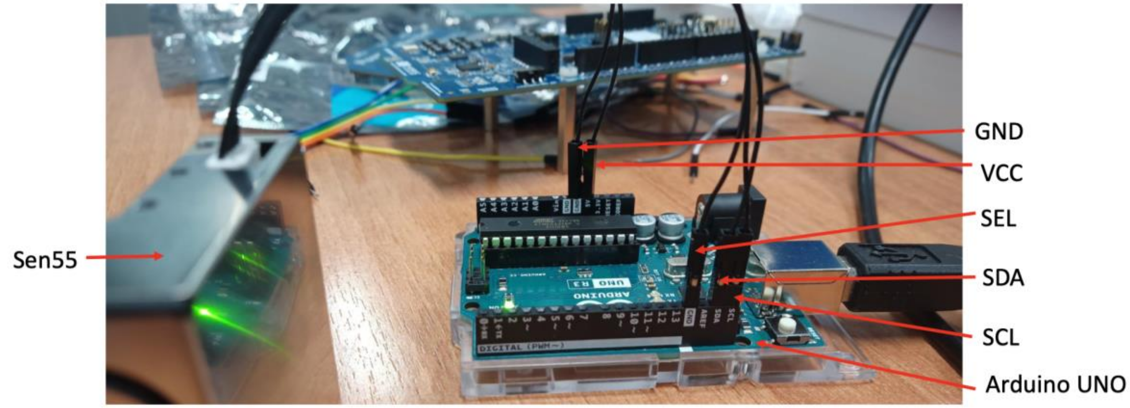
ADC output

Left floating during this measurement



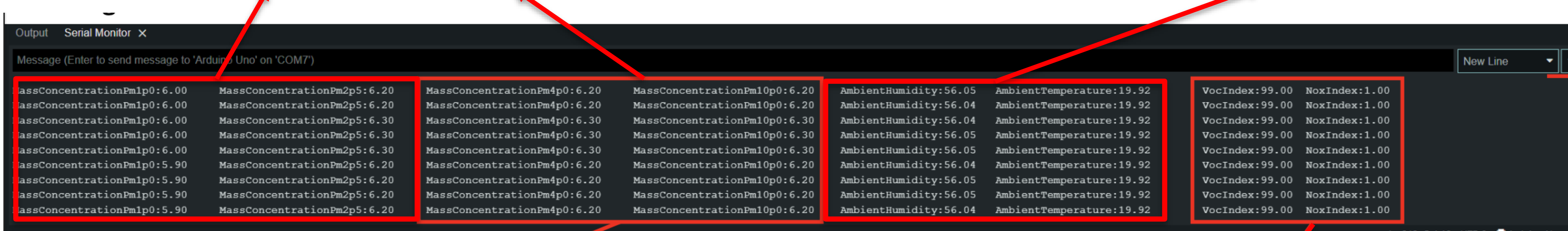
SEN55 combined Evaluation

- ❑ Preliminary test setup – Arduino Uno, SEN55
- ❑ Motivation – rapid assessment of SEN55
- ❑ Firmware in-hand
- ❑ Outcome -> Migration to SEN50 sensor
- ❑



PM4.0 and PM10 outputs of Sensirion's PM sensors are estimated from PM0.5, PM1.0

Can be measured by another sensor (Cost optimization)

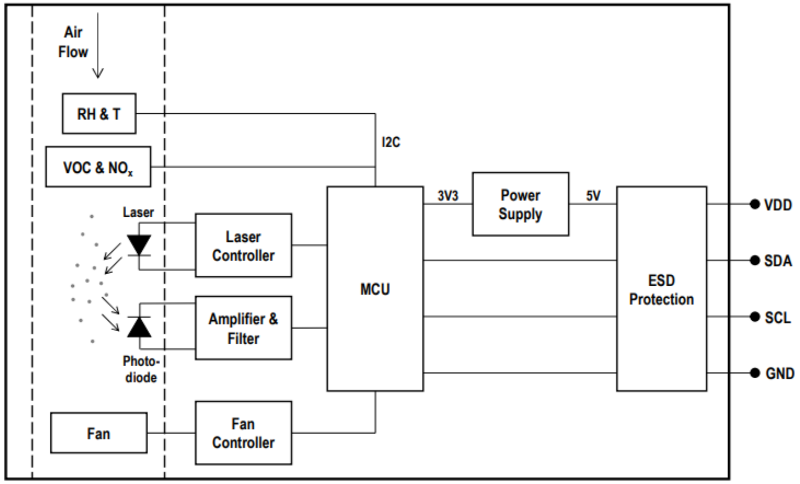


Not useful due to the way in which these values are derived.

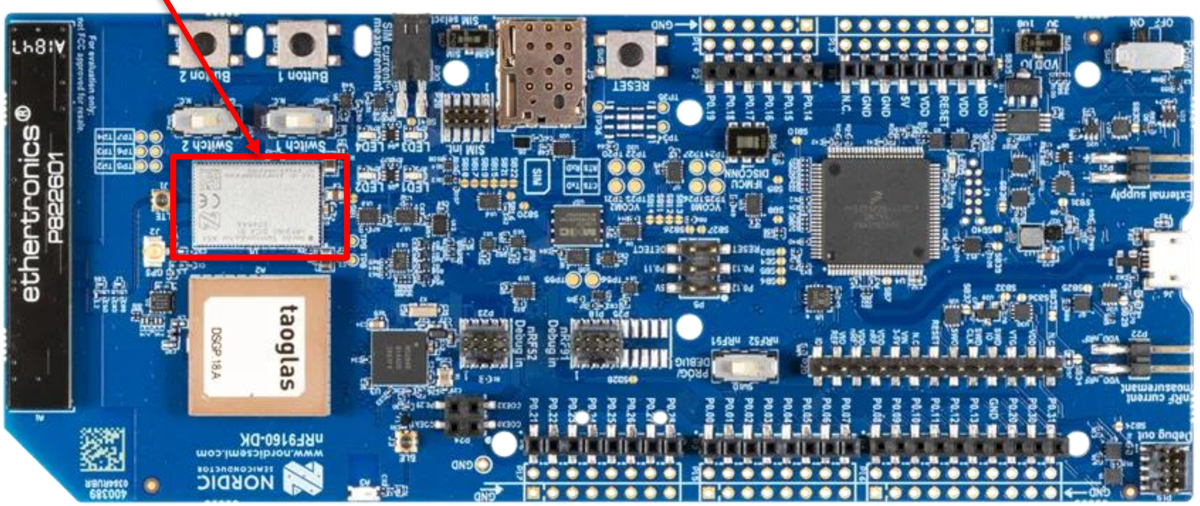
Master Node Development

- The current focus is the integration of the Sensirion sen55 combined sensor with the Zephyr RTOS running on a Nordic Semiconductor nRF9160 SiP.
- Status – I2C driver implementation ongoing.

SEN55 Simplified block diagram

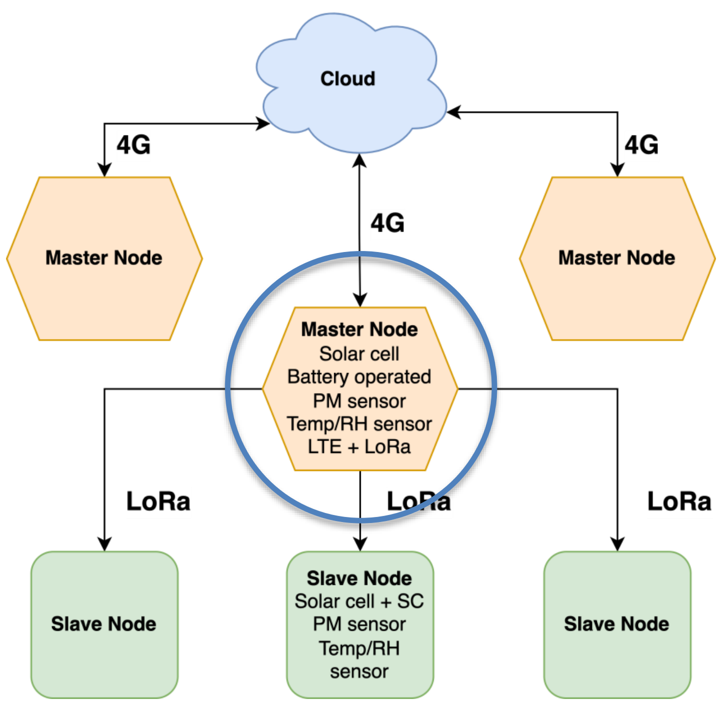


nRF9160 DK



Master Node

- ❑ **Include**
  - ❑ **LoRa module**
  - ❑ **PM sensor**
  - ❑ **Temp + RH sensor**
  - ❑ **SiP with LTE**



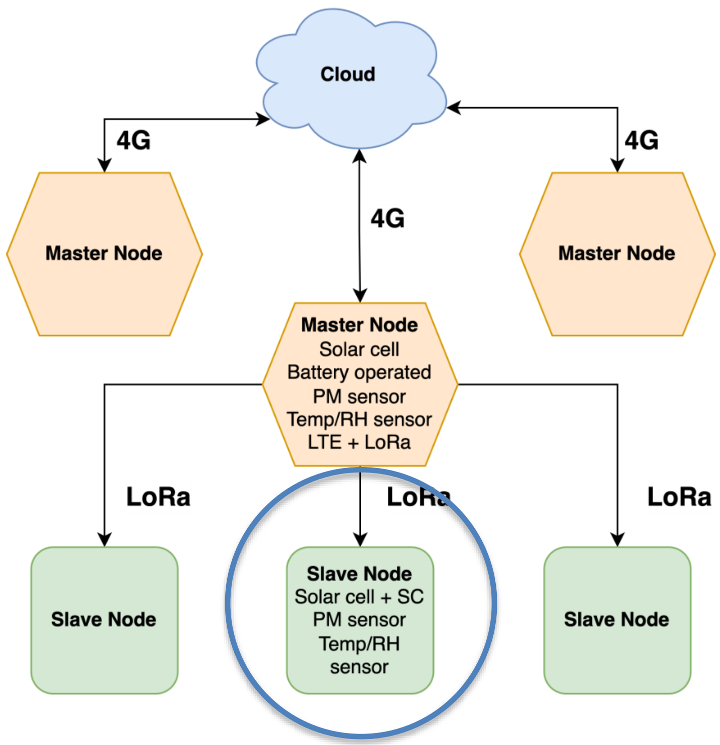
Components	Price 1	Price 100	Info
LR62E	10,11 €	8,40 €	LoRa Module
SEN54	29,92 €	20,40 €	-
nRF9160-SIAA	26,41 €	20,80 €	Only LTE
Charger + passive components (estimated)	20,00 €	10,00 €	Charger to be determined, antenna, caps, resistors…
TOTAL	**86,44 €**	**59,60 €**	**Price per node**

❑ **Replacements**

Components	Price 1	Price 100	Info
nRF9160-SICA	29,94 €	27,62 €	LTE, NB-IoT and GPS
SEN50	23,72 €	16,16 €	Only PM
SX1262	8,26 €	6,18 €	LoRa chip
SHT40-AD1B-R2	2,90 €	1,58 €	RH+T sensor

Slave Node

- ❑ **Include**
  - ❑ **LoRa module**
  - ❑ **PM sensor**
  - ❑ **Temp and RH sensor**



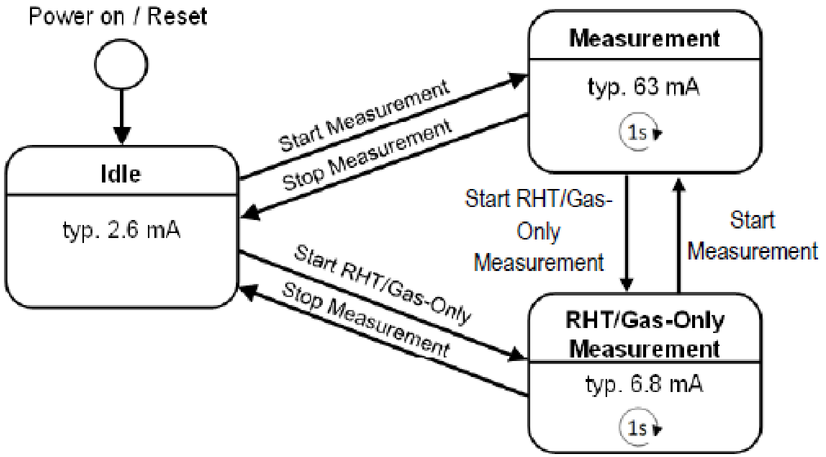
Components	Price 1	Price 100	Info
LR62E	10,11 €	8,40 €	LoRa Module
GP2Y1010AU0F	12,16 €	7,75 €	Sharp PM sensor
nRF52840-QIAA	7,09 €	5,30 €	Microcontroller
SHT40-AD1B-R2	2,90 €	1,58 €	RH+T sensor
Charger + passive components (estimated)	15,00 €	7,00 €	Charger to be determined, antenna, caps, resistors…
TOTAL	**47,26 €**	**30,03 €**	**Price per node**

❑ Replacements

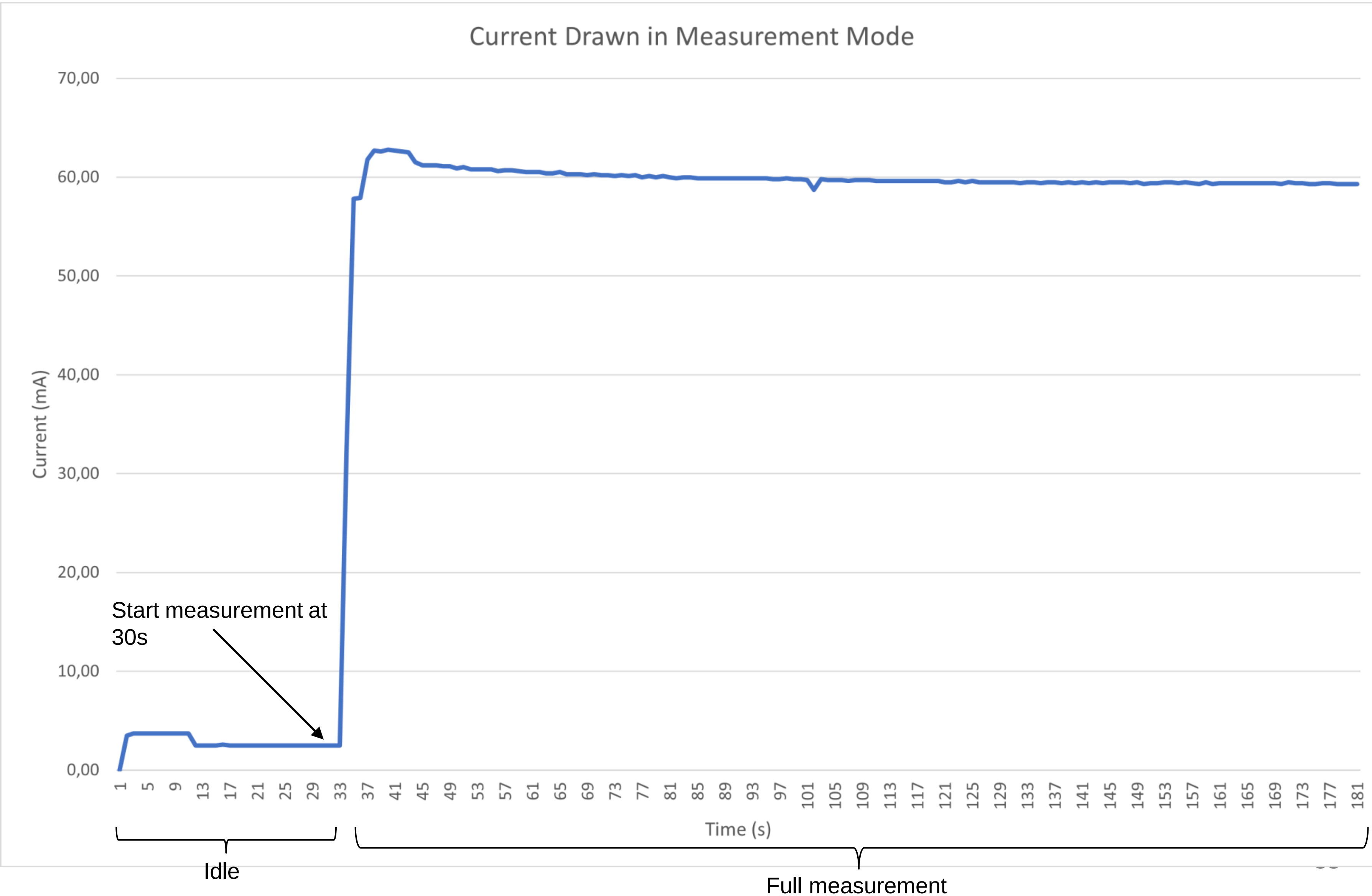
Components	Price 1	Price 100	Info
nRF52840-CKAA	6,69 €	5,00 €	Smaller footprint (worse to route)
SX1262	8,26 €	6,18 €	LoRa chip
SEN54	29,92 €	20,40 €	Include PM + temp + humi sensor
SEN50	23,72 €	16,16 €	Only PM

Sensor Power Consumption

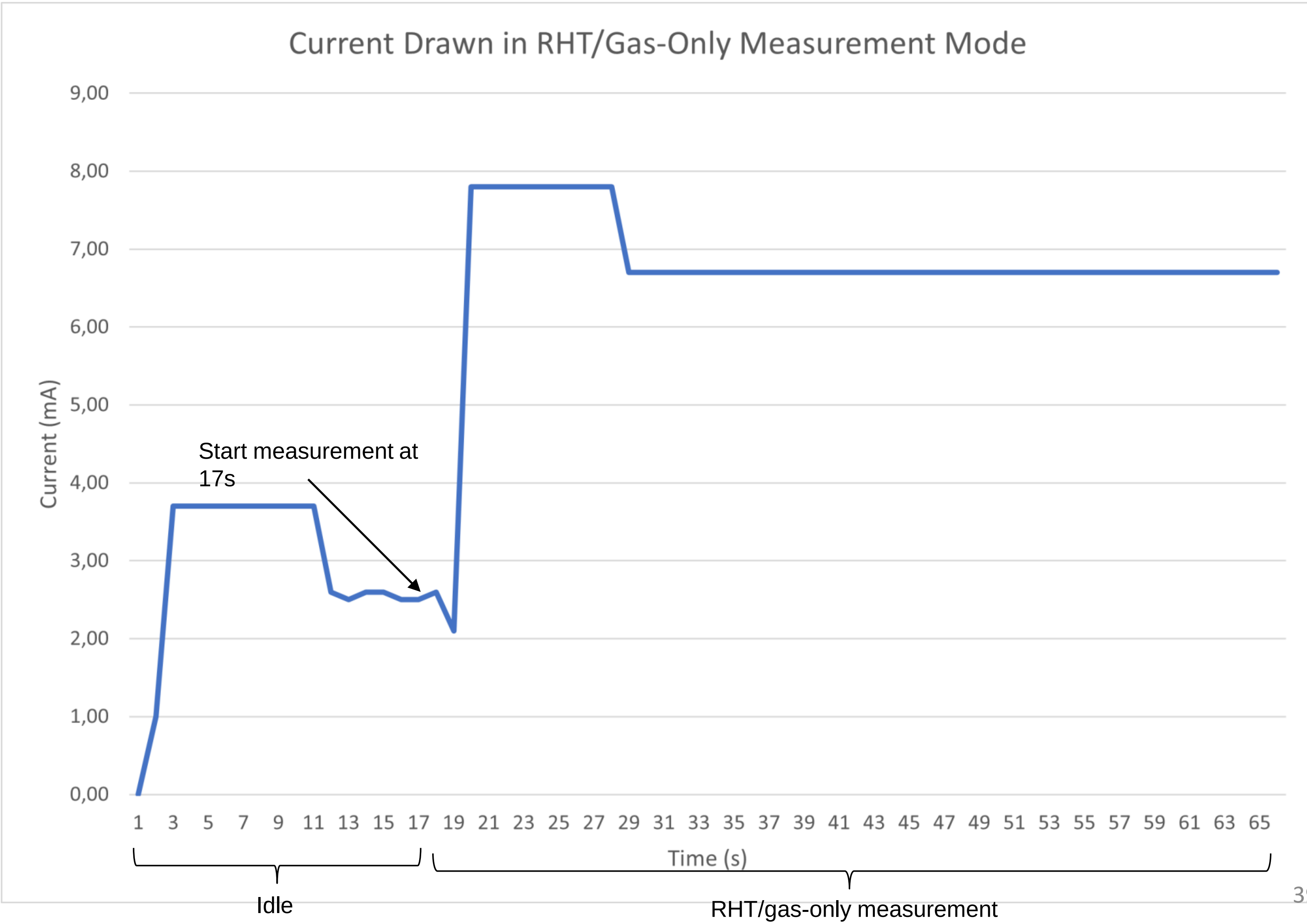
- SEN55 has 3 operating modes
  - Idle: most electronics switched off, ready to receive commands
  - Full measurement mode: all electronics switched on, new readings every second
  - RHT/gas-only measurement mode: RHT and gas sensor on, fan and laser off (no PM measurements), new readings every second



Full Measurement

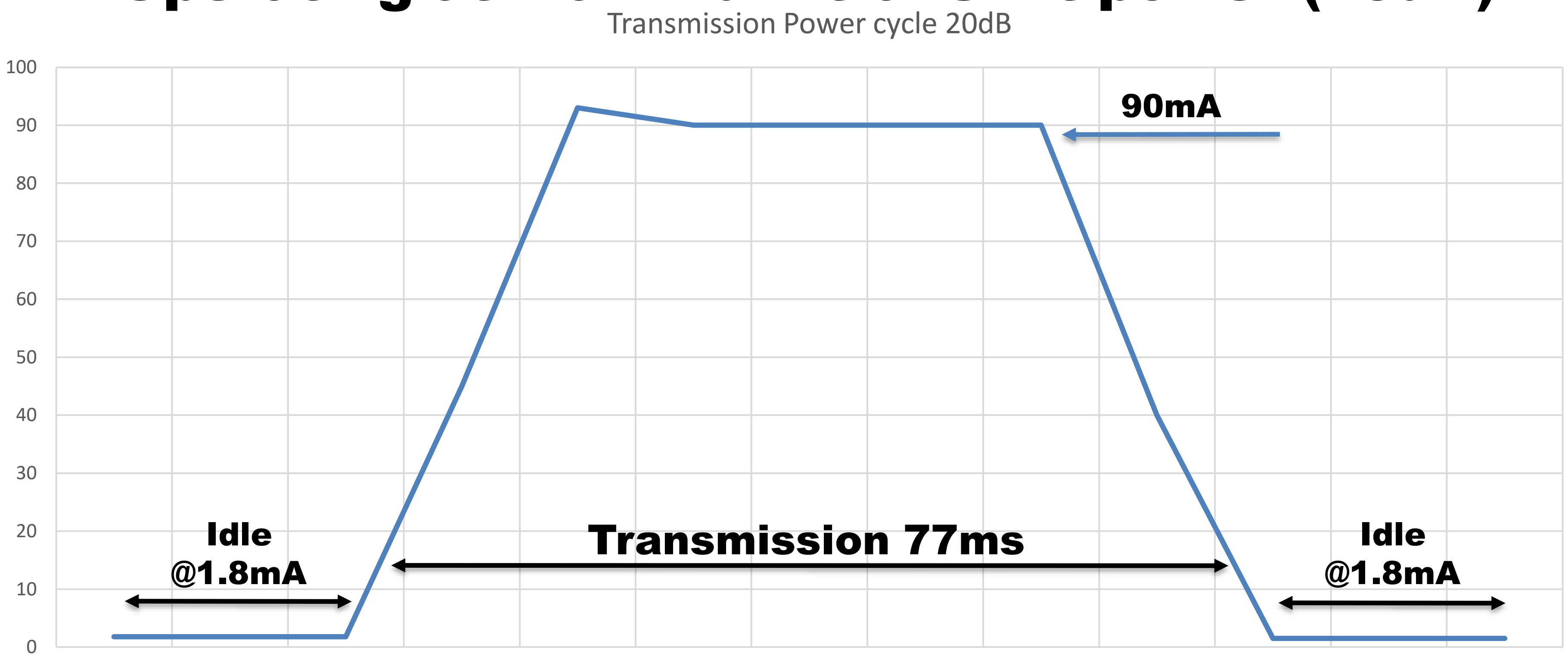


RHT/gas-only Measurement

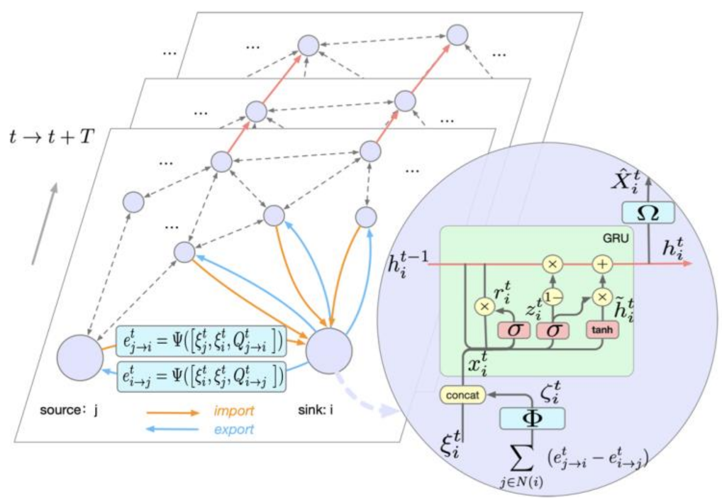


LoRa Transmission Cycle

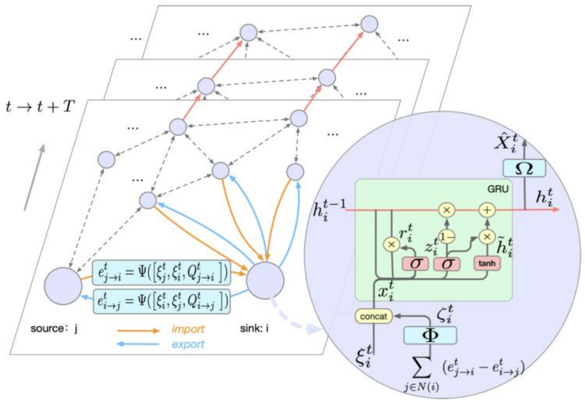
- Operating at maximum transmit power (20dB)

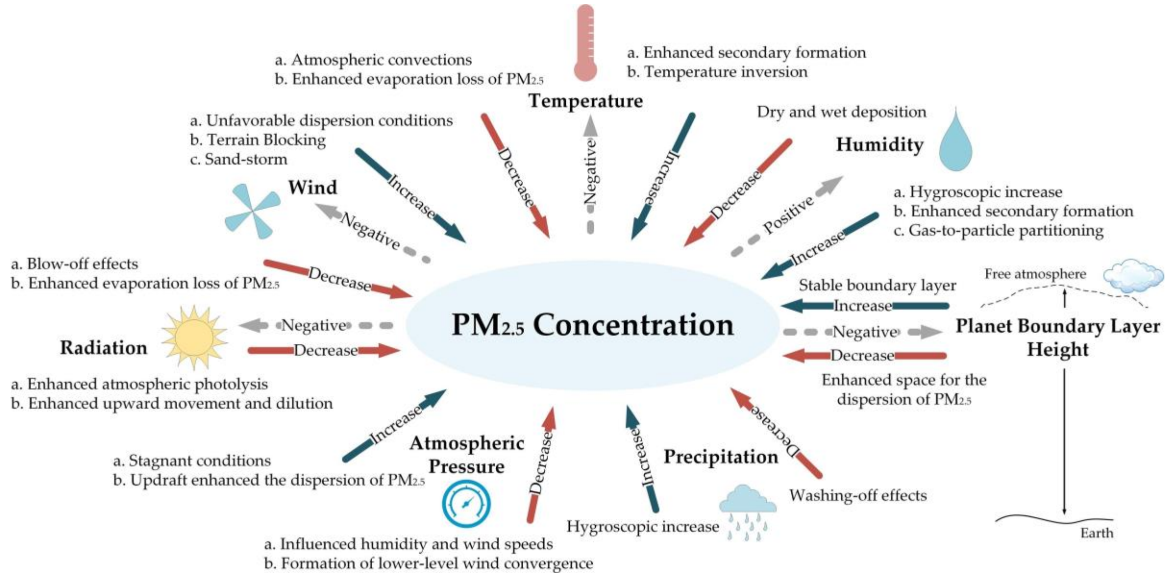


AI-powered Modelling and Early Detection



Methodology





Weather	Pollutants
Temperature	CO
Relative Humidity	O_3
Surface Pressure	PM10
Cloud cover	NOx
Wind speed	NMVOC
Wind Direction	NO_2
Solar Radiation	CPC
Precipitation	EC
	SO_2

Node Features $P^t \in \mathbb{R}^{N \times p}$

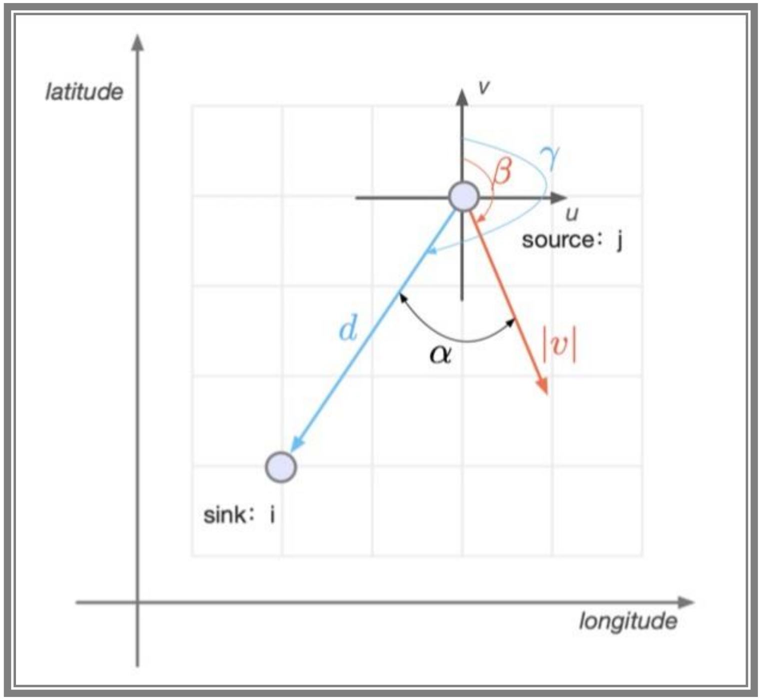
- **Influence the temporal diffusion of the PM2.5 concentration. The meteorological information affects the pollutant concentrations are present in the atmosphere at a given point in time.**
- **Pollutant concentration act as a secondary source of the PM2.5 concentration**

Methodology

Edge Features $Q^t \in \mathbb{R}^{M\times q}$

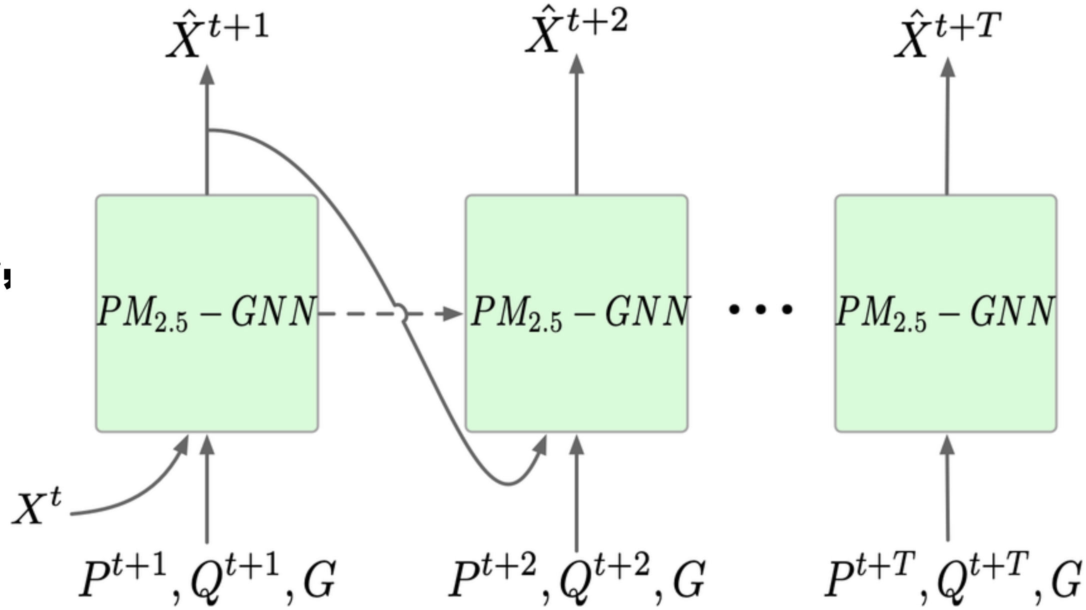
Effect the horizontal transport of the PM2.5 concentration. Edges represent the paths along which the PM2.5 concentration travel from node to node, and are dependent on:

Weather
Wind speed of source node $\lvert v \rvert$
Distance between source and sink d
Wind direction of source node β
Direction from source to sink γ
Advection coefficient S



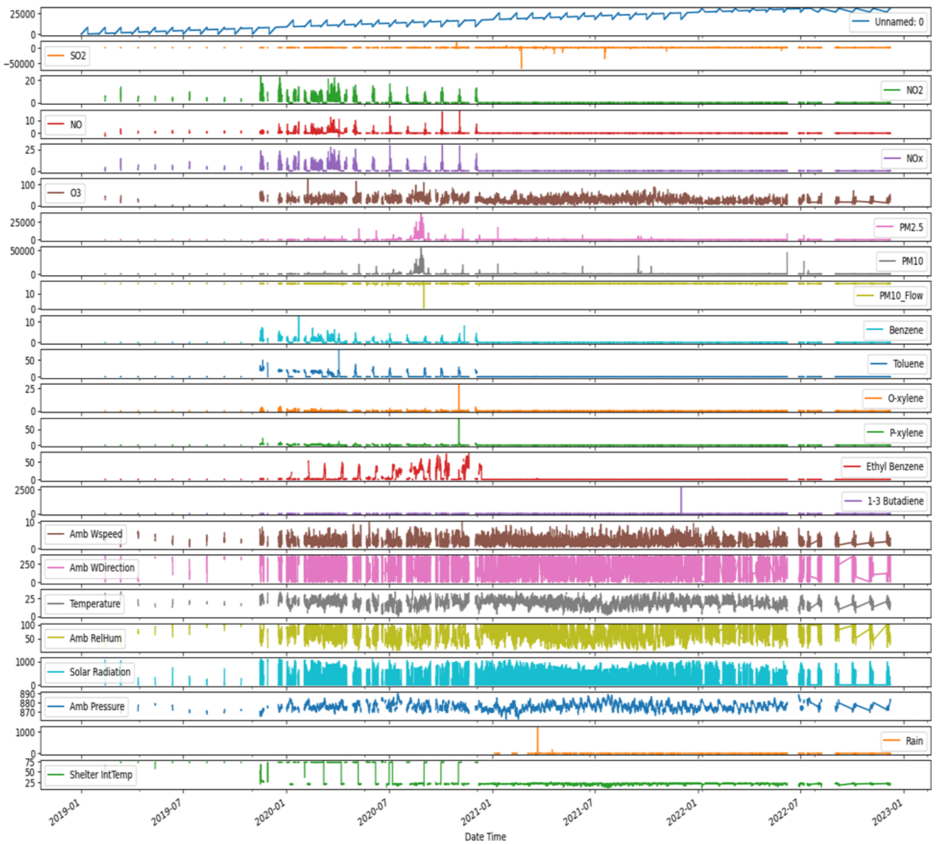
Methodology

- **Problem Definition**
- **X^t - $PM_{2.5}$ concentration at current time step**
- **P^t and Q^t - are node and edge feature matrices respectively for each time step**
- **G directed graph**
- **Input the current X^t,P^{t+1},Q^{t+1}, and G to predict $\hat{X}^{t+1}$**
- **The process is continued for T steps (if you want to predict up to 24, 72, 168 hours)**
- **$\hat{X}^t$ is compared to X^t to test how well the model performs the prediction**

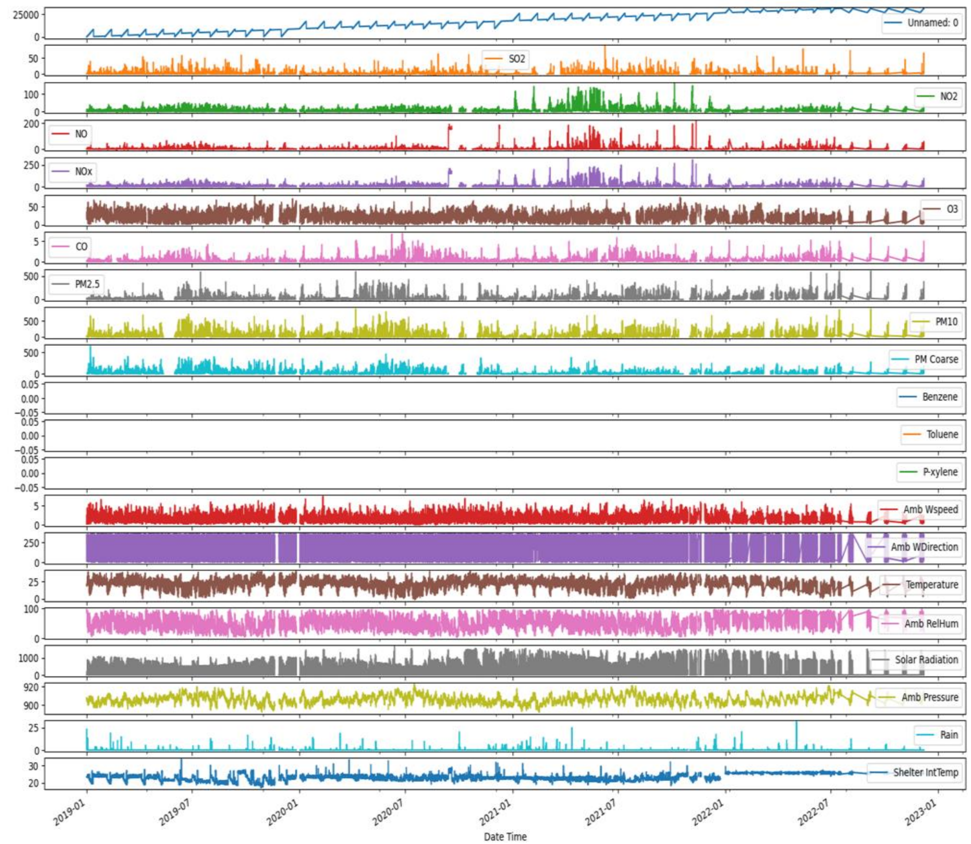


Data collection

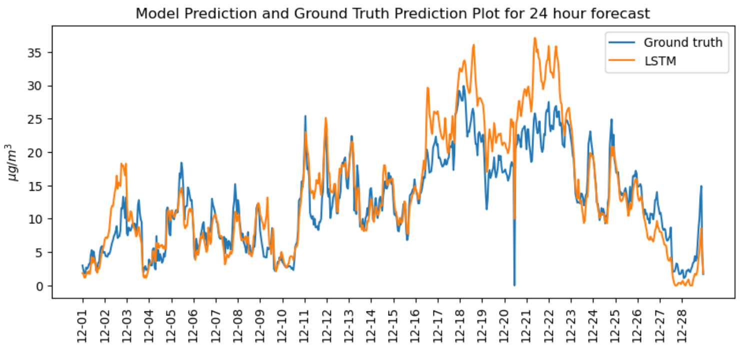
Capricorn

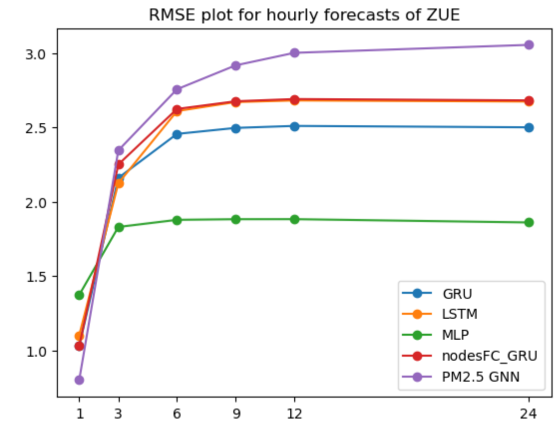


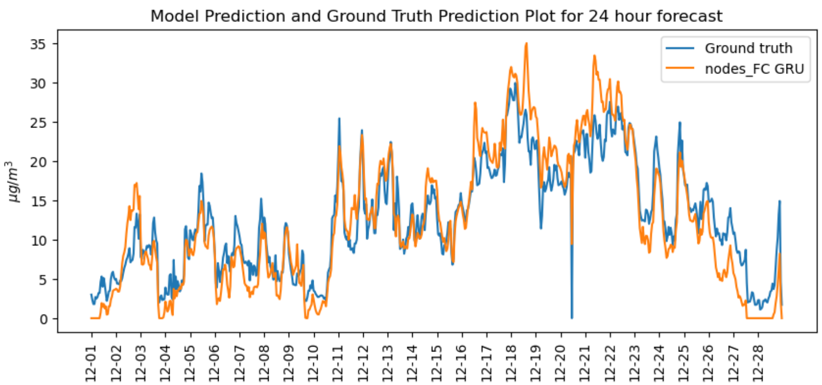
Thabazimbi

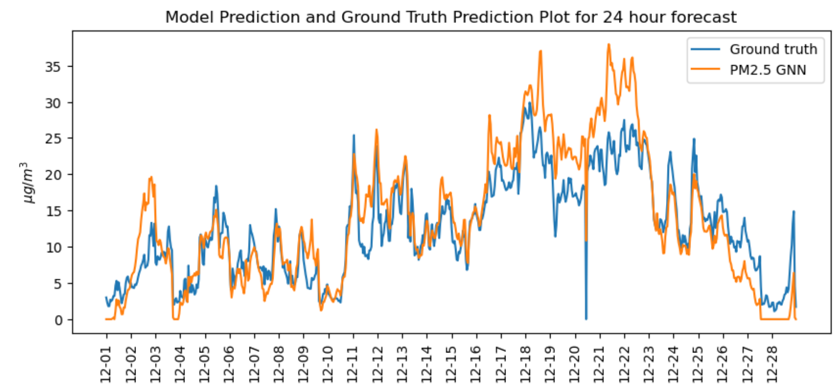


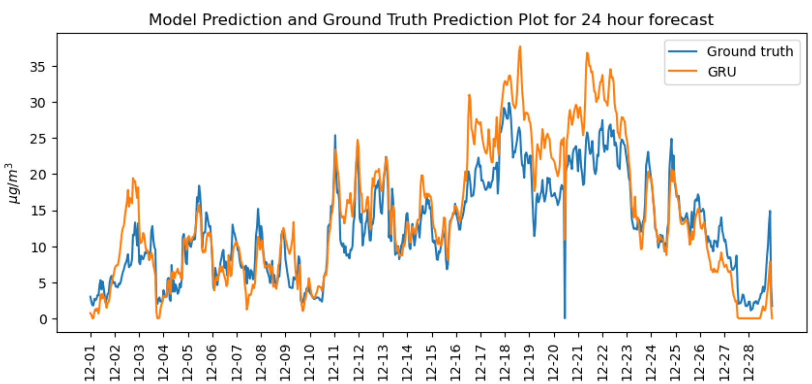
Case Study: Air Monitoring Station in Zurich

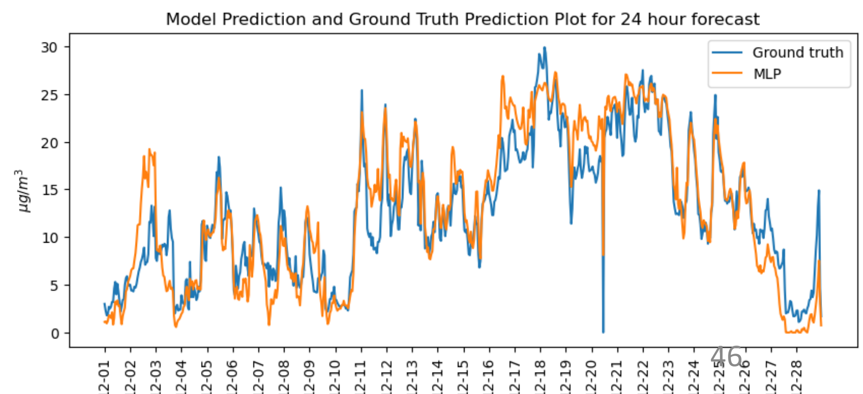




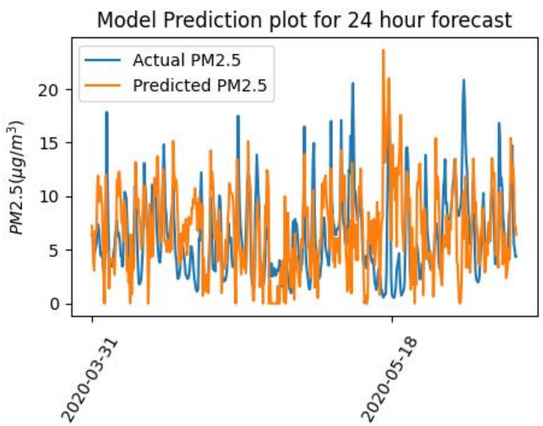


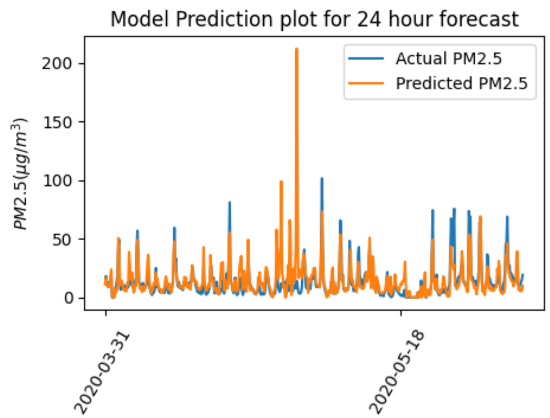


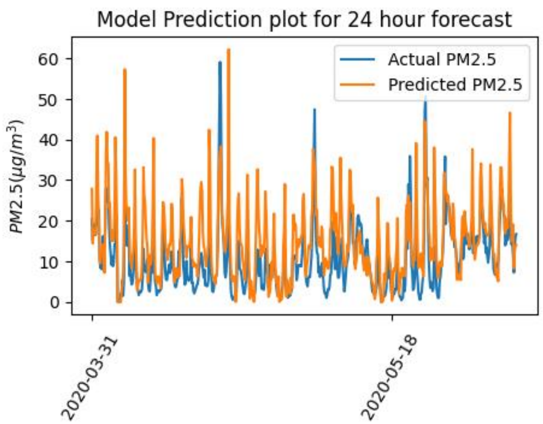


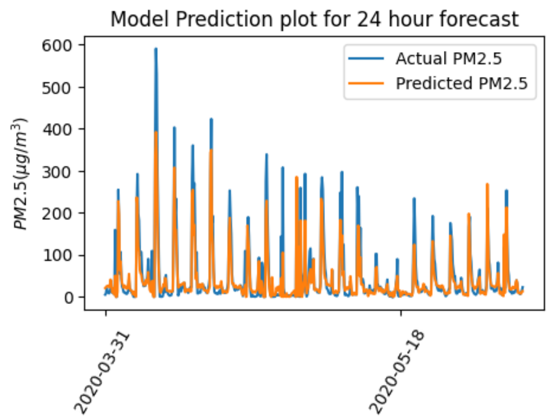


Results with SA Data









Dataset	Train	Validate	Test
1	2019-01-01 to 2019-12-31	2020-01-01 to 2020-03-31	2020-04-01 to 2021-06-30

Deployment of Prototype

An Air Quality monitoring System for the CERN Green Village

- A visionary initiative linking CERN's sustainability roadmap with industrial solutions for sustainability and the talent of young innovators.
- A unique setting, a city within a city, to test and scale up early-stage innovation on site and share technologies and know-how, contributing to Europe's Green Economy.

CERN Green Village | SCE

The CERN campus: A test-bed

- No air quality monitoring system is currently deployed at CERN.
- The high-use Route de Meyrin that connects Saint-Genis Pouilly and Geneva has been identified as a region for deployment,
- The route bisects the CERN Meyrin campus and is a busy arterial road that is frequented by CERN staff, Users, as well as tourists visiting CERN.

Fig. A google maps image illustrating the traffic density of the Route de Meyrin. Credit: Google maps.

Extra Slides

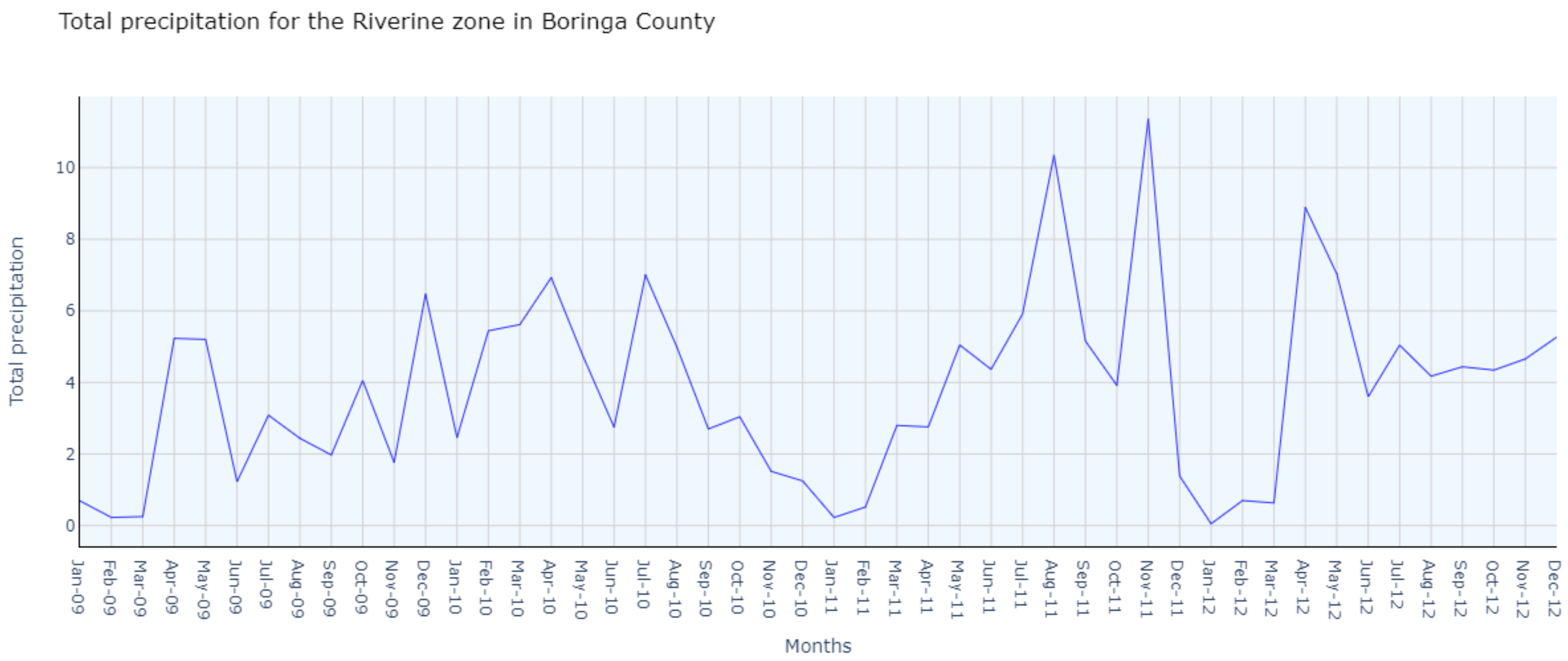
Data processing

- Our study area is the Riverine ecological zone in Baringo county in Kenya.
- The features which make up our dataset are derived from various satellite products between 2009 and 2012 using a cloud-based platform (Google earth engine).
- Features include normalized difference vegetation index (NDVI), average land surface temperature (LST), precipitation, moisture stress index (MSI), normalized difference drought index (NDDI), etc.

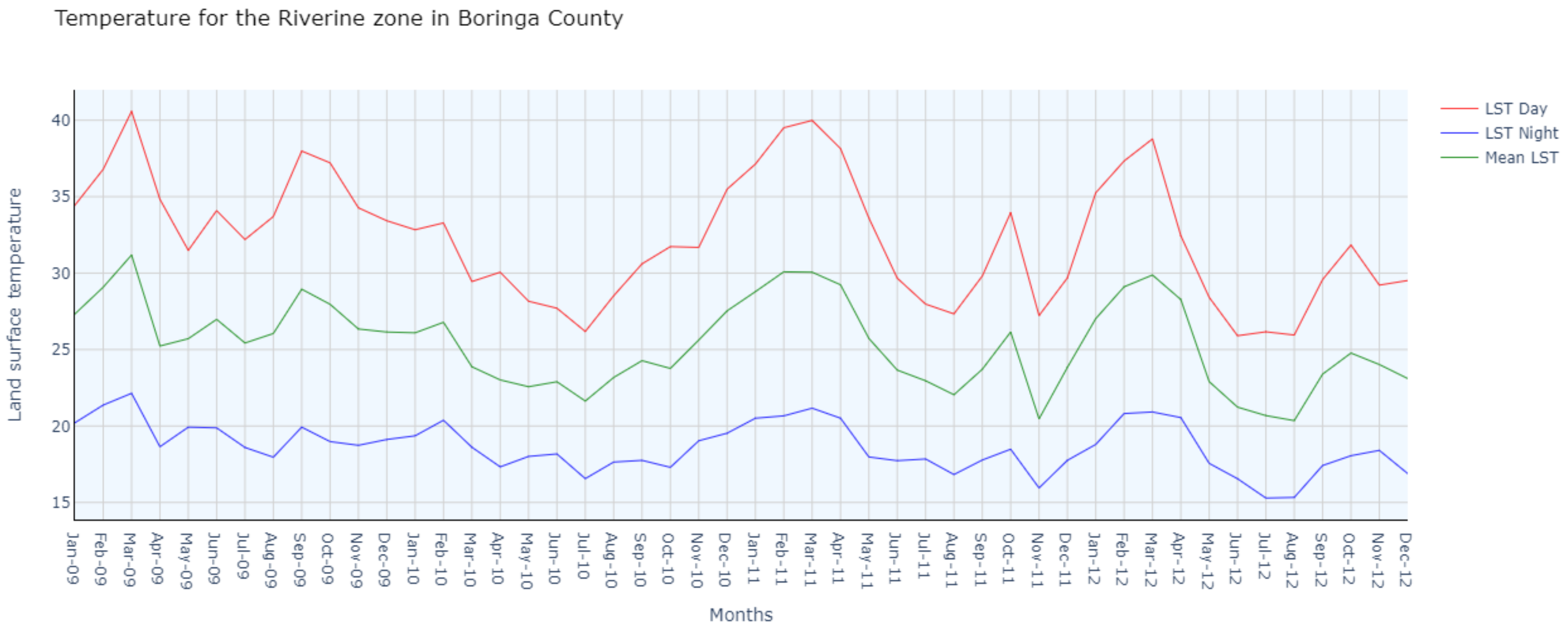
Data processing

- Rainfall: IMERG (Integrated multi-satellite retrievals for GPM) product by NASA provides precipitation estimates at 30-minute intervals for 0.1 arc degree pixels (Wimberly, 2022). This instrument is a better estimate for rainfall compared to the previously proposed methods.
- Land surface temperature: The REACH GEE application uses the MODIS NASA instrument (MOD11A2) which provides 8-day mean surface temperature. For every 8-day cycle temperature values remain the same. LST measures the uppermost surface temperature and estimates ground temperature. Direct ground temperature cannot be retrieved using satellite data.

Average total precipitation for Riverine ecological zone using the Global Precipitation Measurement (GPM)



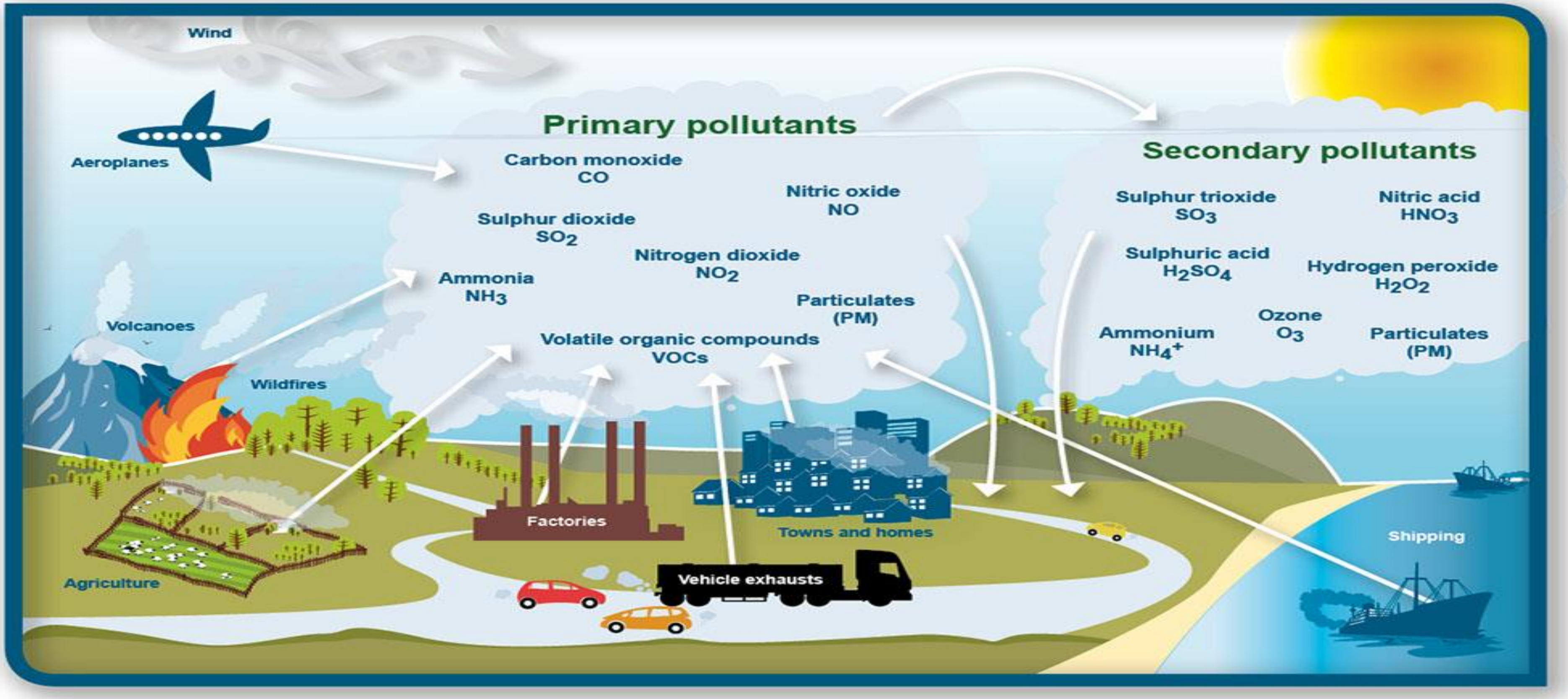
Average land surface temperature for Riverine ecological zone (8 day averages)



Multivariate time series model: LSTM

- We use a multivariate time series model for the prediction of malaria incidents based on the historical data in our study area.
- Our data set has 48-time steps and 27 features. We set up a simple LSTM model with four layers. The first two layers are LSTM layers with 32 and 16 units respectively. The last two layers are comprised of a single dropout layer with a dropout rate and a single dense layer with the final output.
- Brute force hyperparameter tuning was done to produce the most optimal results based on the data we have. We have not done feature selection on the current dataset used for this trial. Future trials will exclude insignificant variables.

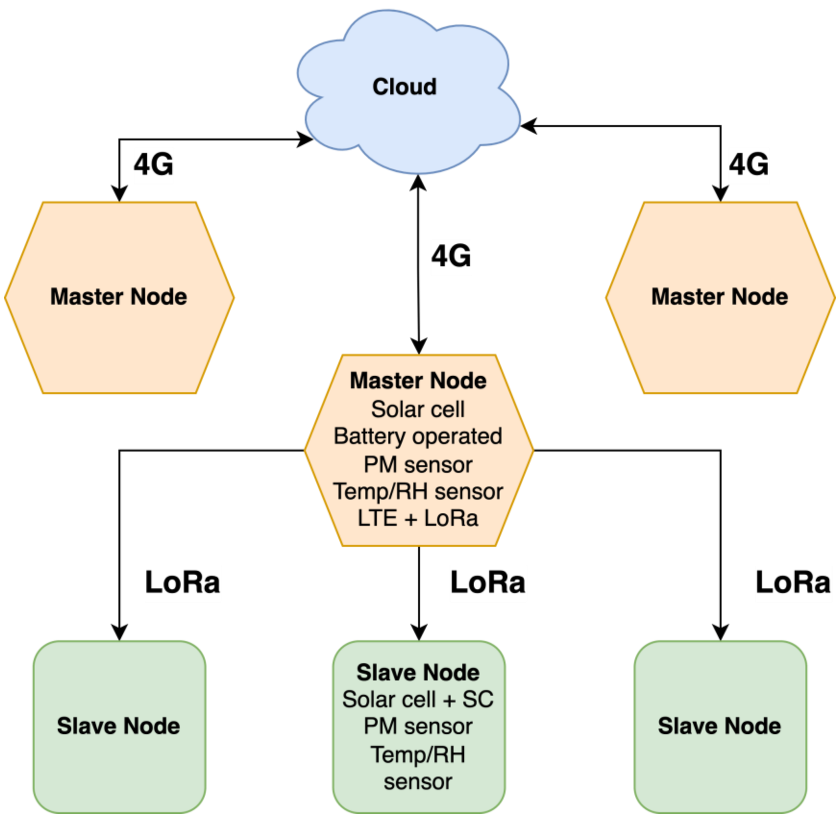
Sources of Air Pollution



- The quality of the air around us is affected by the pollutants released into the atmosphere through human activities, such as transport and industry, as well as from natural sources.
- Multiple sources of pollutions – low level and aloft
- Multiple impacts of pollutions – health, ecosystem (deposition), agriculture (ozone and SO2), water, materials, visibility, climate.

Architecture

- ❑ **Network based in master/slave node topology**
- ❑ **Slaves connect to master via LoRa**
- ❑ **Master connect to Internet via LTE**
- ❑ **Each node is composed by a variety of sensors**
- ❑ **Aiming to reach lowest consumption**
- ❑ **Powered by batteries and super capacitors**
- ❑ **Charge via solar panels**



LoRa Communication

- Hope RFM95W LoRa node/radio
  - Low power, long range
  - Up to 5km in Urban areas
  - Up to 15km in Rural areas
  - 15km+ with direct line of sight
  - Automatic antenna gain calibration
  - Low bit-rate data transfer suitable for this projects needs

LoRa Communication

- Successful prototype of a node to node communication system that:
  - Has low on the air times i.e. transmission is received fairly quickly within range
  - Has good reliability, virtually no erroneous transmissions are received
  - Has good signal strength in urban areas with RSSI > -50dBm in noisy space
  - Consumes very little power (<3mW) when not actively in transmission or receiving mode

LoRa Transmission Power Profile

- During transmission, the node uses a maximum of ±94mA @3.3V ≈ 310mW
  - Typ. ±88mA
    - Rough Measurements awaiting current sensor for more accurate power data
- Sending data from the Sensirion sensor results in a transmission duration of roughly 77ms
- When not actively transmitting the radio goes on standby mode pulling a maximum of 1.8mA @3.3V ≈ 5.9mW

List of components

- Particulate Matter (PM) Sensor
- Temperature and Relative Humidity (RH) Sensor
- LoRa communication
- Controllers

PM Sensors

- SHARP sensor
  - GP2Y1010AU0F
- Sensirion SEN5x
  - SEN50
    - PM sensor
  - SEN54
    - Same as SEN50
    - Temp/RH sensor
    - VOC sensor
  - SEN55
    - Same as SEN54
    - NOx sensor

Temperature and RH sensor

- Lots of modules in the market
- Selection done price/characteristics
- Sensirion SHT40
  - Both sensors included
  - Low power

LoRa Modules

- LoRa Module
  - Fanstel LR62E
  - Includes SX1262
- LoRa chip
  - SEMTECH SX1262

Controllers

- Nordic Semiconductor nRF52840
  - System-on-Chip (SoC)
- Nordic Semiconductor nRF9160
  - System-In-Package (SiP)
  - Includes LTE
  - Different versions
    - nRF9160-SICA
      - LTE-M + NB-IoT + GPS
    - nRF9160-SIAA
      - Only LTE-M
    - nRF9160-SIBA
      - Only NB-IoT

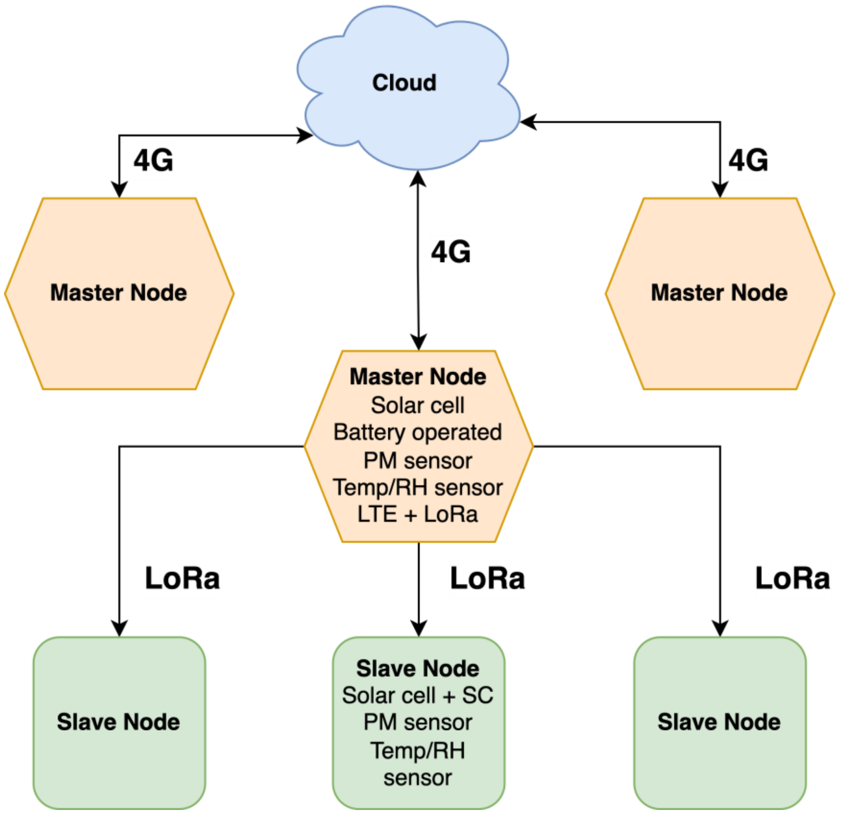
Air Quality Measurement

- **Sensor: Sensirion SEN55**
- **Measurements:**
  - **Particulate matter (PM1, PM2.5, PM4, PM10)**
  - **Temperature**
  - **Relative Humidity**
  - **VOC index**
  - **NOx index**

- PM, NO_x, VOC, RH & T sensor platform
- Fast & easy integration
- One driver for up to 8 data signals
- Superior sensing accuracy and lifetime
- Fully calibrated digital output

Why LoRa?

- Long Range
  - Up to 2km in urban areas
- Low power consumption
- More info about LoRa can be found in here



Sensor Power Consumption

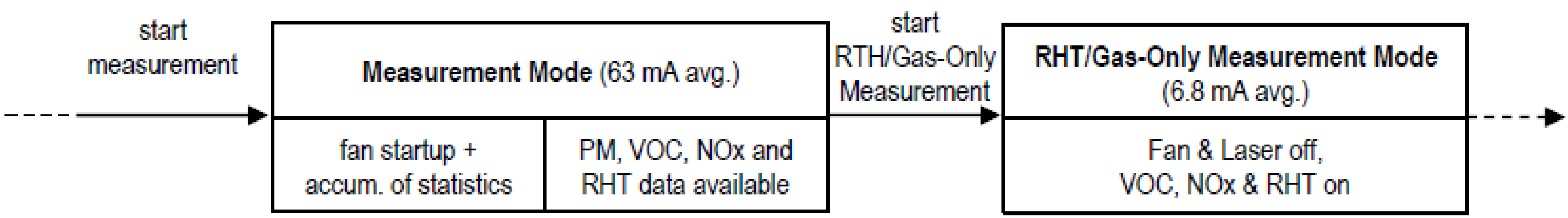
- **Power consumption for each mode of operation was measured experimentally (current drawn at a supply voltage of 5V)**
- **Idle: 3.7mA (18.5mW) for 10s after being switched on, 2.5 – 2.6mA (12.5 – 13mW) after 10s**
- **Full measurement: rises to 62.8mA (314mW) peak after being switched on, falls to 61.1mA (305.5mW) after 15s, falls to 59.3mA (296.5mW) after 180s**
- **RHT/gas-only measurement: rises to 7.8mA (39mW) peak after being switched on, falls to 6.7mA (33.5mW) after 10s**

Reduced Power Operation

- Measurement Interval: 1 hour
- Time in full measurement mode (RHT, gas and PM): 60 seconds
  - 30 seconds for startup
  - 30 seconds to record readings
  - Recommendation: take an average of 30 seconds of readings for each sampling interval to improve measurement stability
- Datasheet recommends operating in RHT/gas-only mode for the rest of the time, but this may not be necessary for the applications of this project

Reduced Power Operation – Datasheet Recommendation

A proper, alternating use of these operation modes as indicated in Figure 2 may reduce power consumption by a factor of 7-9 with only minimal compromises on sensor system performance.



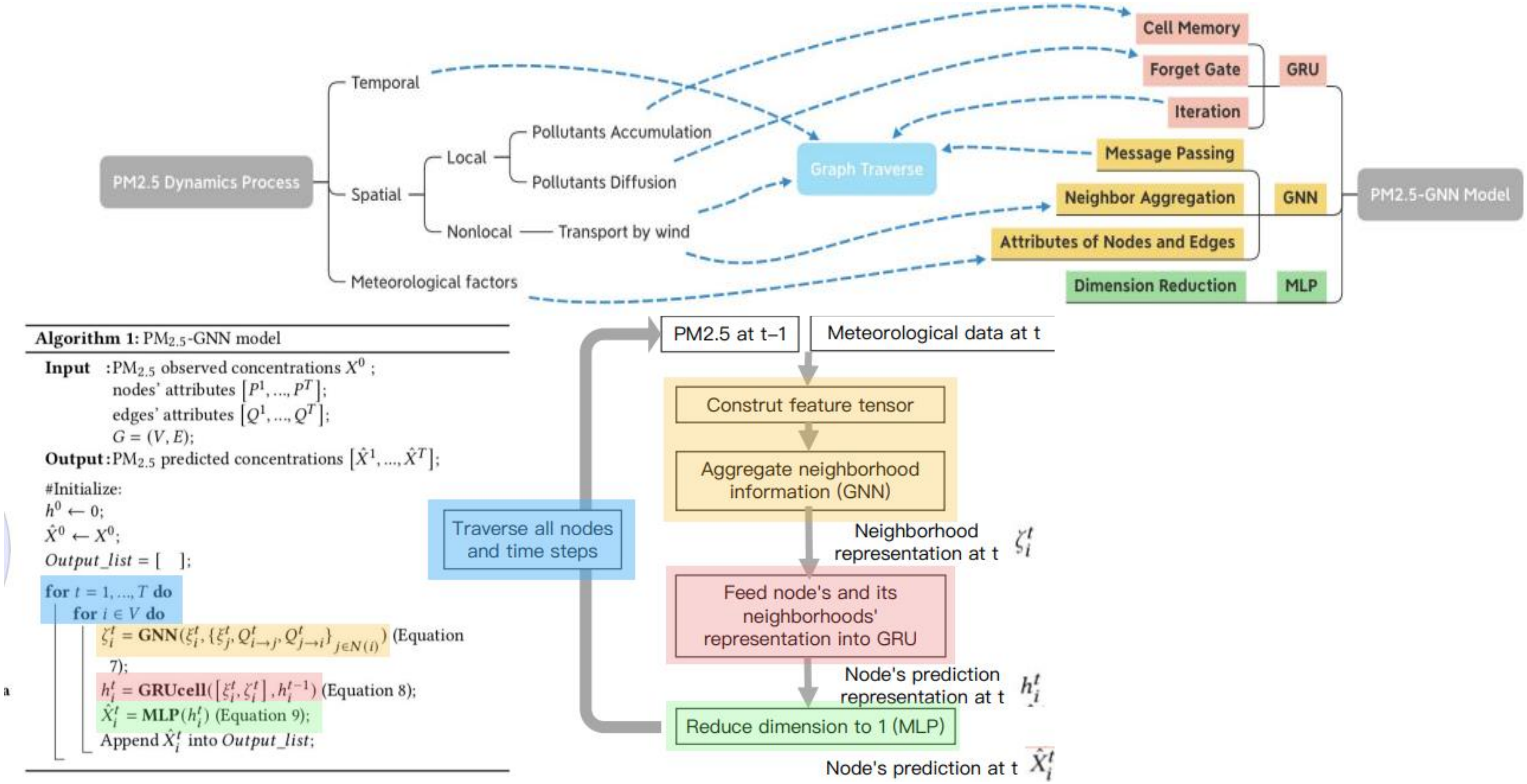
There are two main variables influencing the overall power consumption that need to be traded off with performance of the sensor system: the time spent in measurement mode, as well as the time spent in the RTH/Gas-Only Measurement mode.

Time in Measurement mode ↔ Ability to detect fast pollution events (High power consumption)

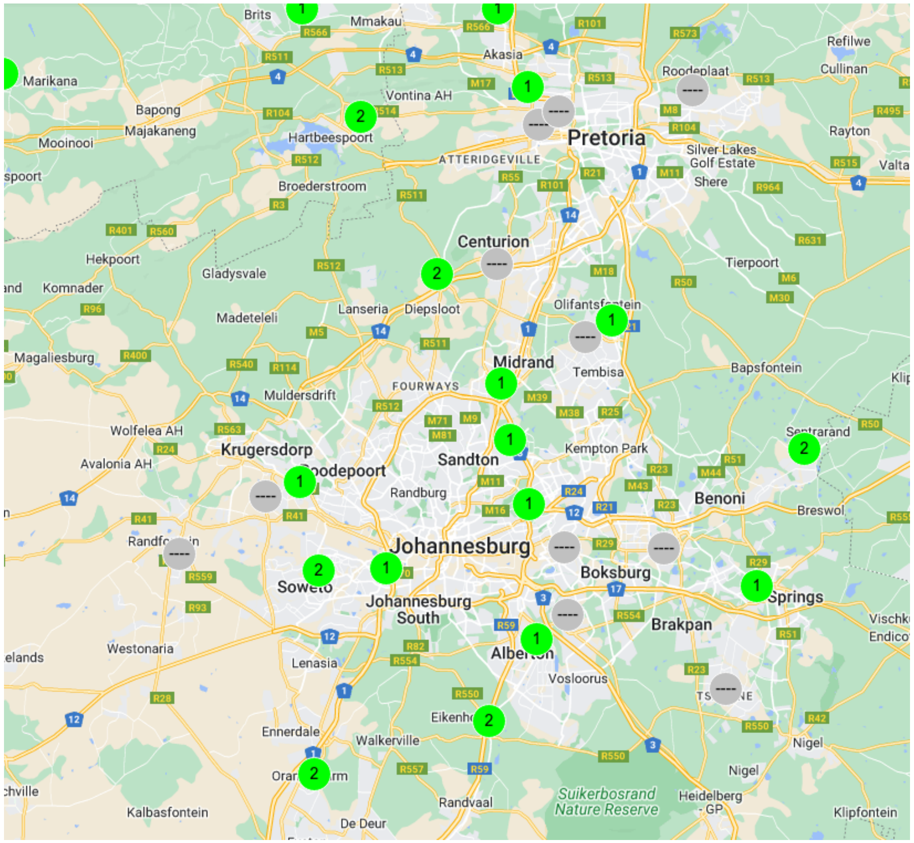
Time in RHT/Gas-Only Measurement mode ↔ Ability to identify trends and slow pollution events (Low power consumption)

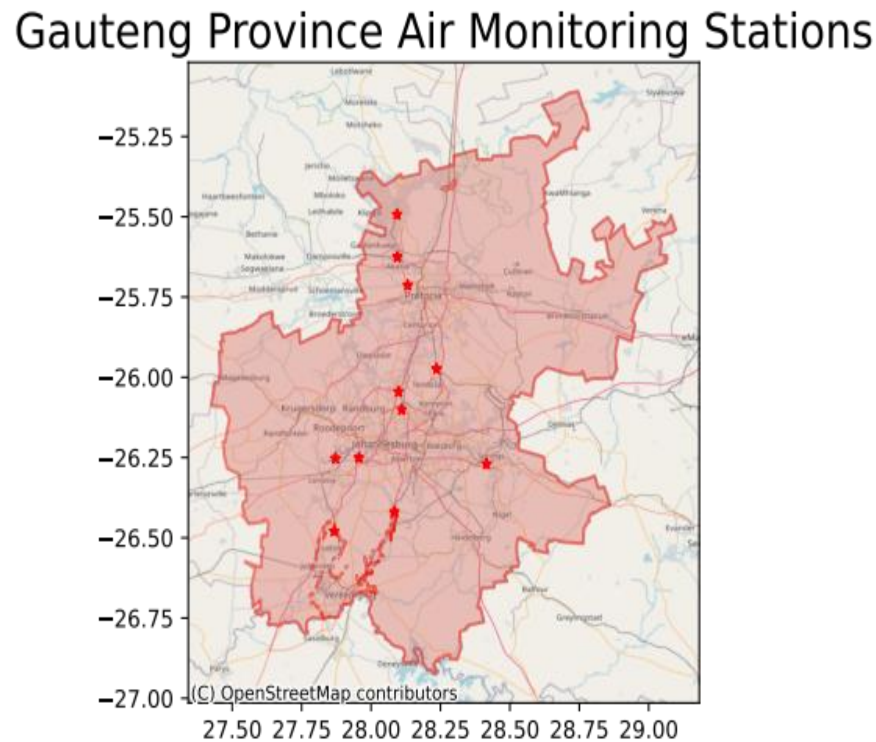
PM2.5 GNN Model

A knowledge-enhanced Graph Neural Network (GNN) is devised to capture pollutants' horizontal transport by leveraging neighbouring information and updating nodes' representations. A spatio-temporal GRU is applied after updates to model pollutants' vertical accumulation and diffusion under the influence of weather



Study Area

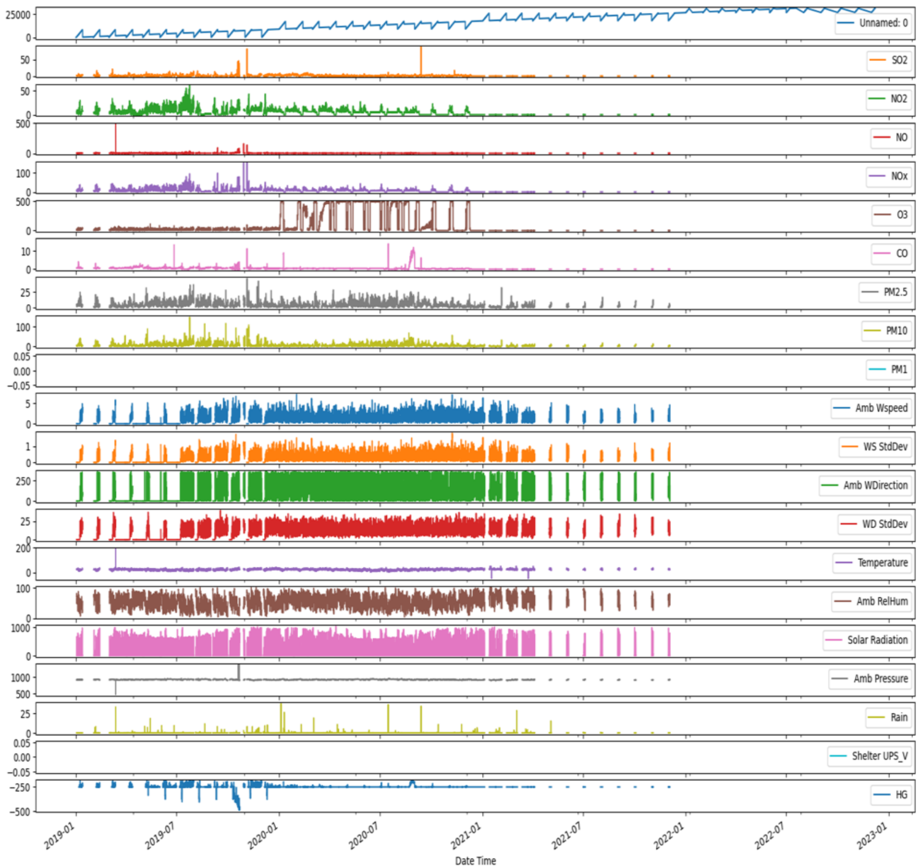




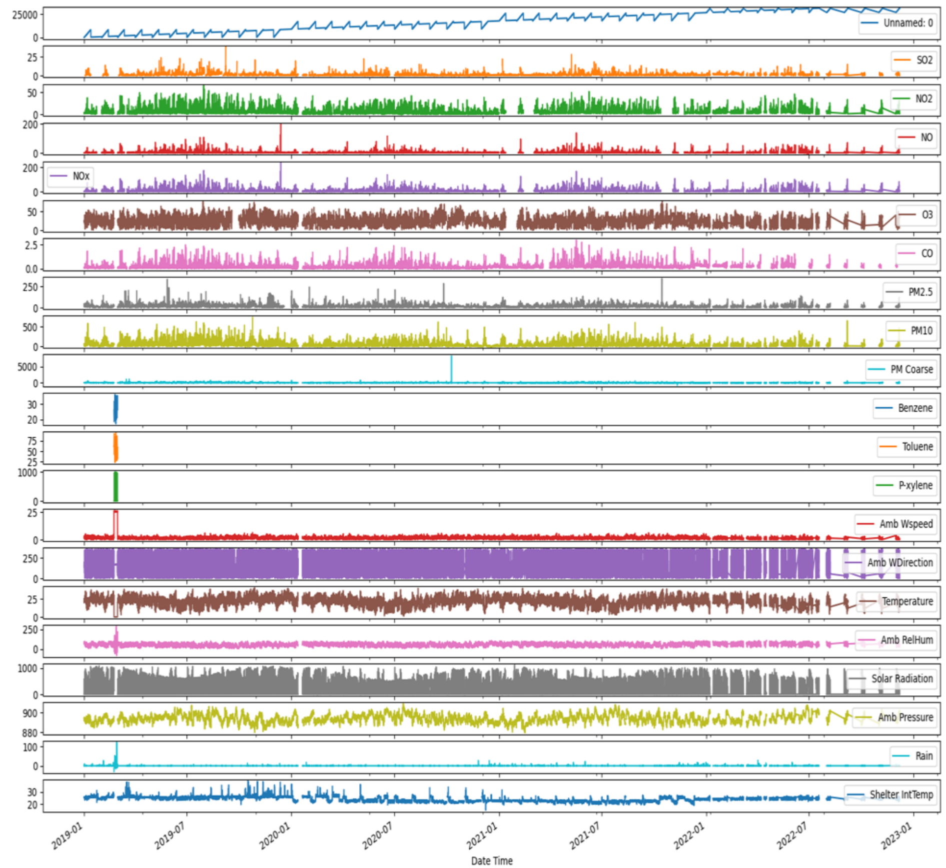
- **Monitoring Networks Included**
- City of Johannesburg (Some stations have low data recovery)
- City of Ekurhuleni Municipality (Most stations have low data recovery)
- City of Tshwane (Some stations have low data recovery)

Done: Data collection

Dilokong

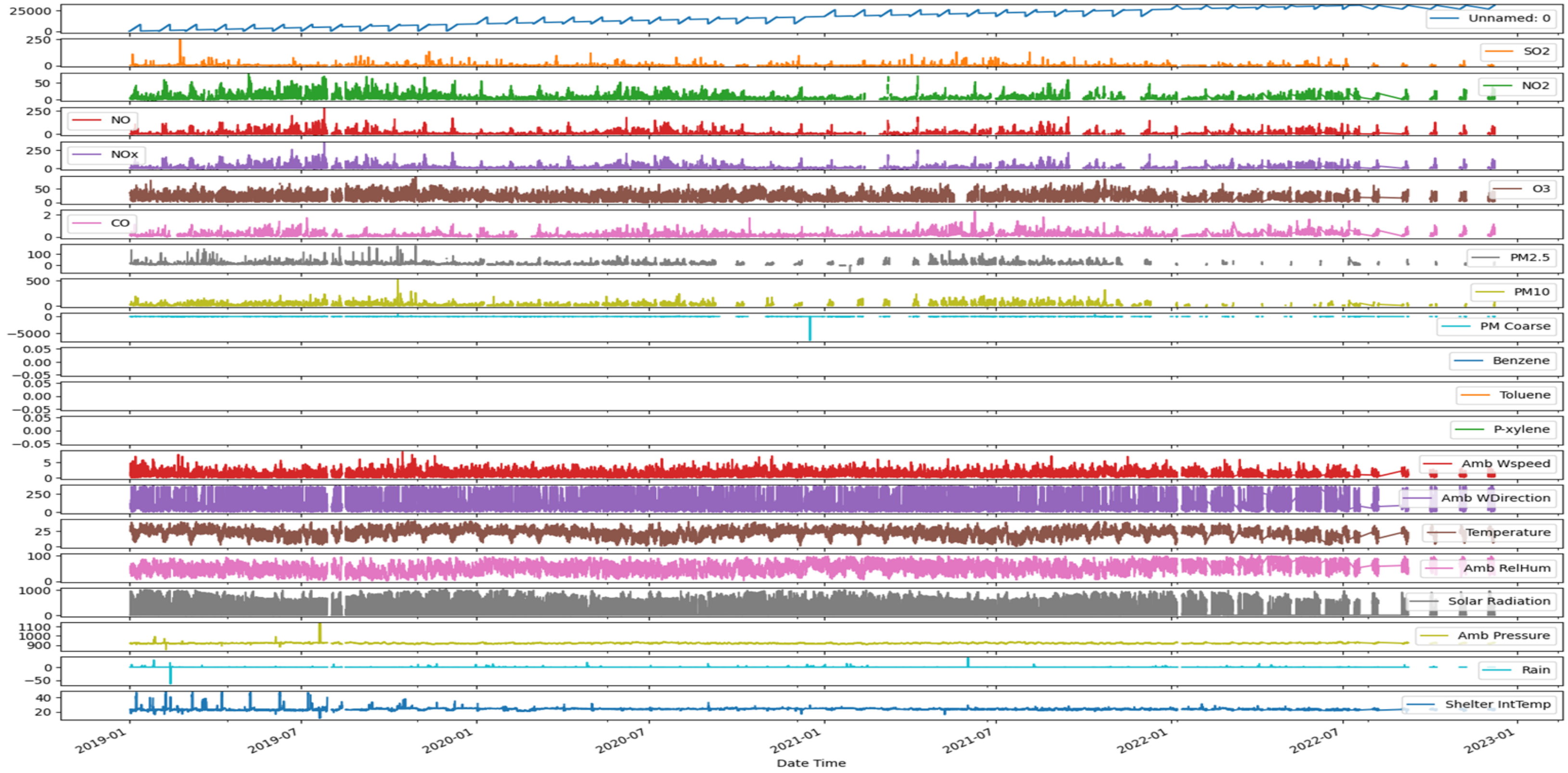


Mokopane

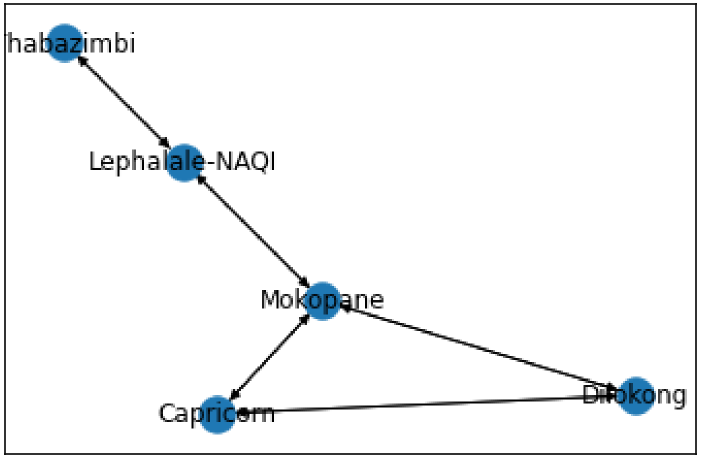


Data collection

Lephalale-NAQI



Graph construction



- **Adjacent matrix shows connected stations.**

```
matrix([[1, 1, 1, 0, 0],
        [1, 1, 1, 1, 0],
        [1, 1, 1, 0, 0],
        [0, 1, 0, 1, 1],
        [0, 0, 0, 1, 1]], dtype=uint8)
```

Adjacent matrix

- **From table 2, the distance between is computed shown in table 1.**

	Station Name	Longitude	Latitude	Altitude	Station ID
0	Dilokong	30.171036	-24.615222	0	0
1	Mokopane	28.983199	-24.155951	1093	1
2	Capricorn	29.405415	-23.884413	1310	2
3	Lephalale-NAQI	27.722012	-23.682068	834	3
4	Thabazimbi	27.391605	-24.591058	977	4

Table 2

Source	Tank	Distance	Direction
0	1	130.798544	-202.641358
0	2	112.238448	-225.994995
1	0	130.798544	336.868195
1	2	52.436616	-325.087815
1	3	138.719772	-201.975368
2	0	112.238448	313.690542
2	1	52.436616	-144.915947
3	1	138.719772	337.513269
3	4	106.133354	-108.379885
4	3	106.133354	-288.514997

Table 1:Distance between stations

Results

Matric	MLP		LSTM		GRU		nodesFC-GRU		PM2.5 GNN	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Train loss	0.7691	0.0044	0.4201	0.0039	0.4660	0.0052	0.4559	0.0069	**0.2893**	**0.0026**
Val loss	0.7824	0.0000	0.3191	0.0000	0.2707	0.0000	0.3571	0.0000	**0.2042**	**0.0000**
Test loss	0.6161	0.0078	0.2218	0.0049	0.1994	0.0013	0.2205	0.0087	**0.1618**	**0.0059**
RMSE	26.7292	0.2876	15.4056	0.3437	15.7377	0.3455	16.1433	0.1190	**13.9910**	**0.0744**
MAE	17.1112	0.0194	8.7349	0.3825	9.2490	0.3378	9.7907	0.0377	**7.8570**	**0.0583**