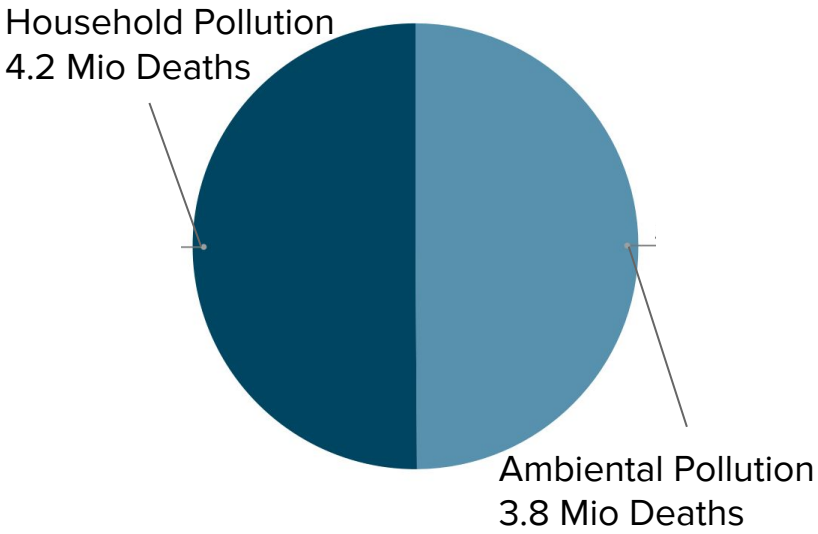
Forecasting Ugandan Air Quality using Weather Recordings

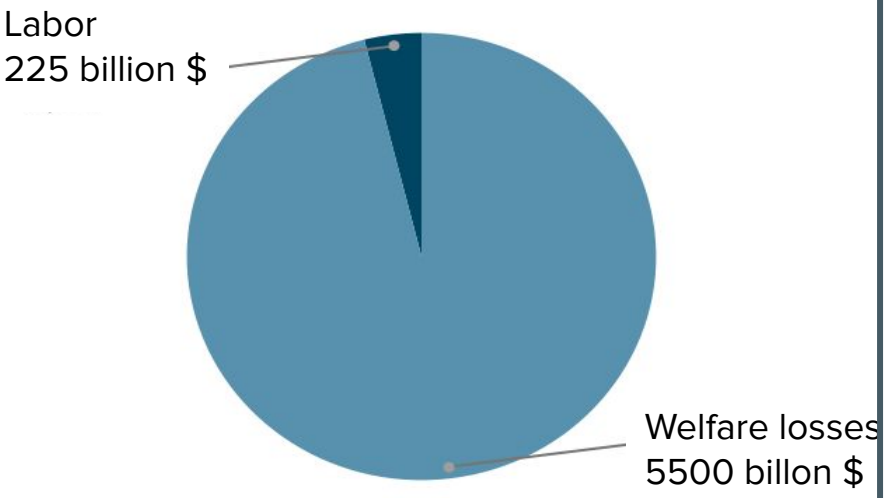
Alexander Engelhardt, Alwin Klick, Beatriz Amanda Watts, Fabian Klauke

Consequences of Air Pollution

8 Million Deaths Worldwide (WHO 2016)



Economic Losses (WHO, 2013)



Avis, W. (2020) Urban Mobility, Air Pollution and Covid-19. ASAP-East Africa, University of Birmingham

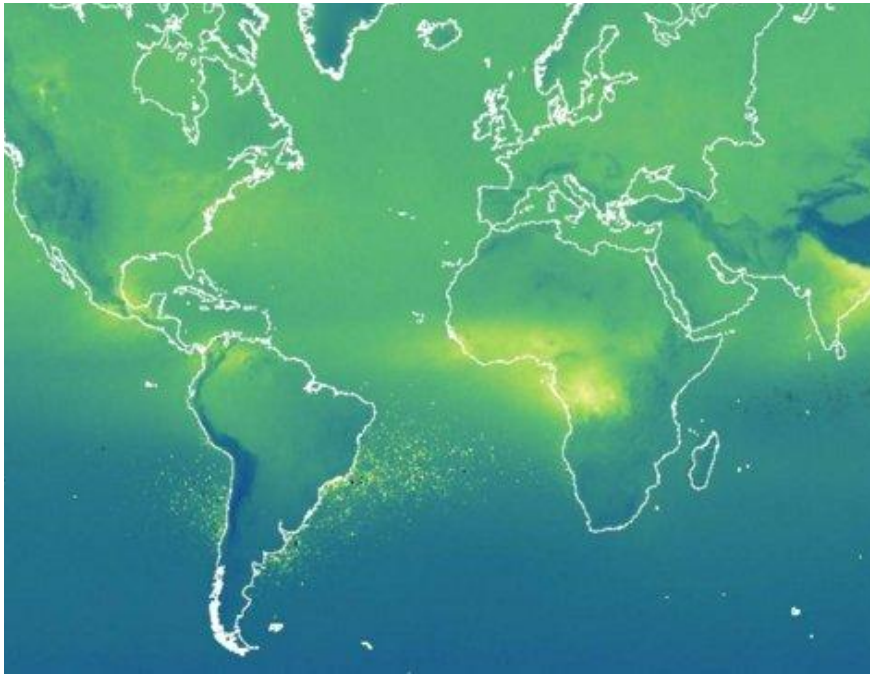
Stakeholder Case

Why not measure air pollution directly?

> It is expensive and too late to issue warnings

The Ugandan Government approached us to generate a predictive model:

> Forecast air pollution based on easily accessible weather data



https://developers.google.com/earth-engine/datasets/catalog/sentinel-5p

Strathmore University · Cardiff University / Prifysgol Caerdydd · KCCA Kampala Capital City Authority · Nord University · Uganda National Roads Authority · Ethiopian Public Health Institute · University of Birmingham · ACTS · AAiT · University of Nairobi

Air Quality: $PM_{2.5}$

- $PM_{2.5}$ can cause
  - Cancer
  - Cardiovascular diseases
  - Respiratory diseases
- More particles per m^3 = higher health risk
- Some people are more sensitive

Air Quality: $PM_{2.5}$

- $PM_{2.5}$ can cause
  - Cancer
  - Cardiovascular diseases
  - Respiratory diseases
- More particles per m^3 = higher health risk
- Some people are more sensitive

Health Concern	$PM_{2.5}$ (μgm^{-3})	Precautions
Good	0 - 12	None
Moderate	13 - 35	Unusually sensitive people should consider reducing prolonged or heavy exertion
Unhealthy for Sensitive Groups	36 - 55	Sensitive groups should reduce prolonged or heavy exertion
Unhealthy	56 - 150	Everyone should reduce prolonged or heavy exertion, take more breaks during outdoor activities
Very Unhealthy	151 - 250	Everyone should avoid prolonged or heavy exertion, move activities indoors or reschedule
Hazardous	250 +	Everyone should avoid all physical activities outdoors.

Air Quality: $PM_{2.5}$

- $PM_{2.5}$ can cause
  - Cancer
  - Cardiovascular diseases
  - Respiratory diseases
- More particles per m^3 = higher health risk
- Some people are more sensitive

Health Concern	$PM_{2.5}$ (μgm^{-3})	Precautions
Good	0 - 12	None
Moderate	13 - 35	Unusually sensitive people should consider reducing prolonged or heavy exertion
Unhealthy for Sensitive Groups	36 - 55	Sensitive groups should reduce prolonged or heavy exertion
Unhealthy	56 - 150	Everyone should reduce prolonged or heavy exertion, take more breaks during outdoor activities
Very Unhealthy	151 - 250	Everyone should avoid prolonged or heavy exertion, move activities indoors or reschedule
Hazardous	250 +	Everyone should avoid all physical activities outdoors.

Air Quality: $PM_{2.5}$

- $PM_{2.5}$ can cause
  - Cancer
  - Cardiovascular diseases
  - Respiratory diseases
- More particles per m^3 = higher health risk
- Some people are more sensitive

Health Concern	$PM_{2.5}$ (μgm^{-3})	Precautions
Good	0 - 12	None
Moderate	13 - 35	Unusually sensitive people should consider reducing prolonged or heavy exertion
Unhealthy for Sensitive Groups	36 - 55	Sensitive groups should reduce prolonged or heavy exertion
Unhealthy	56 - 150	Everyone should reduce prolonged or heavy exertion, take more breaks during outdoor activities
Very Unhealthy	151 - 250	Everyone should avoid prolonged or heavy exertion, move activities indoors or reschedule
Hazardous	250 +	Everyone should avoid all physical activities outdoors.

Our Central Question

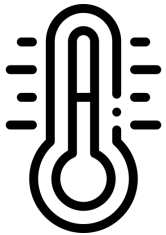
Can we forecast air pollution in Uganda based on easily accessible and cheap-to-use weather recordings?

We assume

- Weather data is sufficient to predict $PM_{2.5}$ concentration 24 hours in advance
- Forecasts are precise enough to not span more than one category

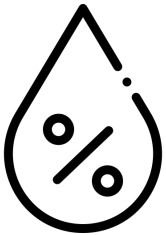
The Dataset

Features

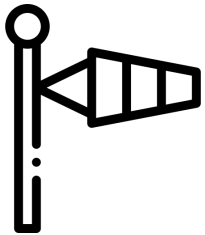


Temperature

Precipitation



Humidity



Wind

Pressure

Location

Data Cleaning and Preparation

- Features extracted from each recording period:
  - Mean
  - Standard deviation
  - Last measurement
  - Time of the day
  - Sensor location
- Observations with missings >30% removed

Modelling Strategy

- Use weather features to predict $PM_{2.5}$ concentration
- Optimisation: Minimise average prediction error (RMSE)

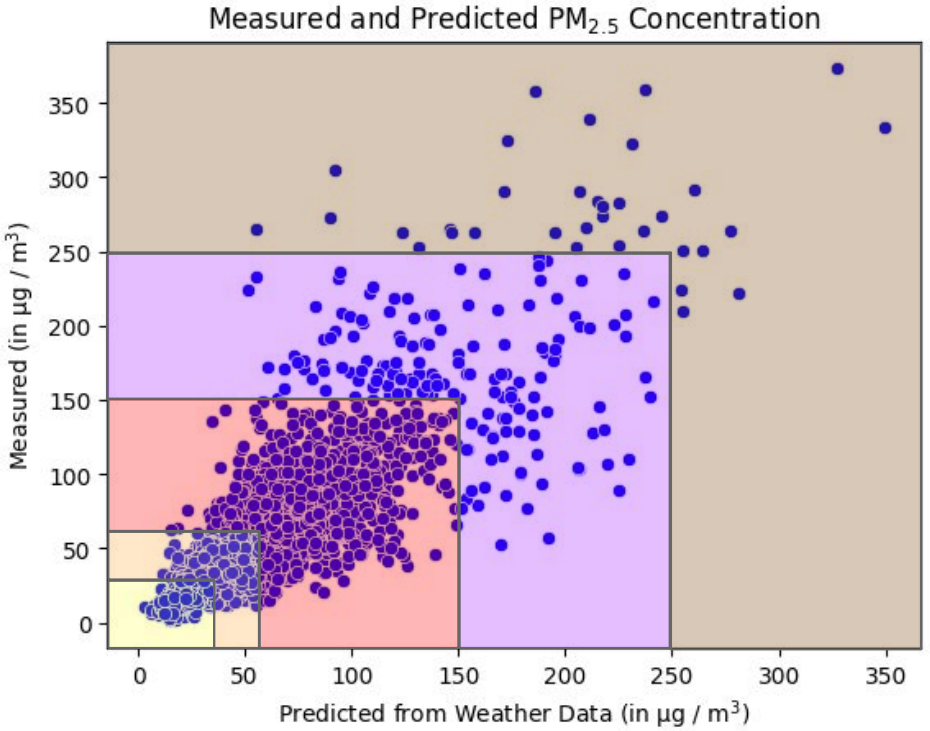
- **First baseline model**
  - Multiple linear regression using the mean feature values
  - **RMSE**: 41.11

Findings

Final Model:

XGBoost Regressor

RMSE: 24.47



Health Concern	$PM_{2.5}$ (μgm^{-3})
Good	0 - 12
Moderate	13 - 35
Unhealthy for Sensitive Groups	36 - 55
Unhealthy	56 - 150
Very Unhealthy	151 - 250
Hazardous	250 +

Implications

1. Forecasts air pollution based on cheap weather data
2. Higher predictive power than competing algorithms
3. Can be used to issue warnings 24 h in advance

Outlook

1. Identify most important features
2. Train separate model for every sensor
3. Include: traffic data, datetime info, holidays, ...


Thanks for your attention!