Air pollution prediction using machine learning techniques – A concept to replace existing monitoring stations with virtual monitoring stations

M.Sc. Abdul Samad

ICUC11 Sydney, 28.08.2023

Introduction

Air pollution and its harmful effects

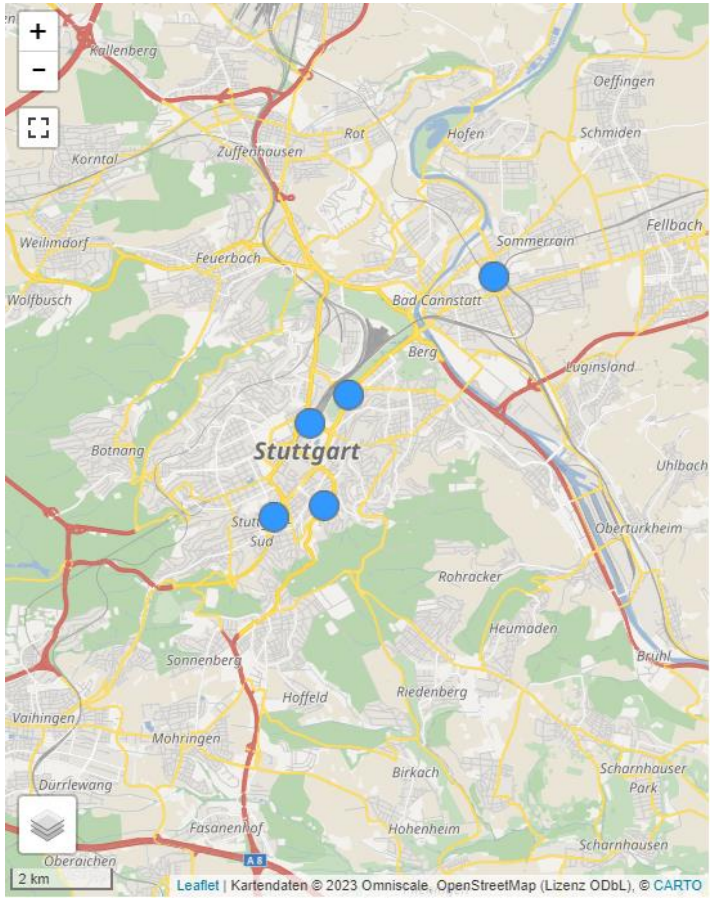
- 7 million premature deaths estimated annually worldwide due to air pollution
- Air quality is an important aspect of the urban climate
- Air quality in the cities is often poor
- More than half of human population live in cities who are exposed to these emissions
- Parameters affecting air quality
  - Meteorological parameters e.g. air temperature, wind speed, solar radiation, precipitation, etc.
  - Thermal wind systems e.g. cold air flows, slope winds, temperature inversions, etc.
  - Local emission sources e.g. traffic, combustion, etc.

Source: World Health Organization (Nov. 2021): Air Pollution silent killer (Retrieved from https://www.who.int/multi-media/details/air-pollution-silent-killer)

Introduction

Air quality monitoring

- Low number of air quality monitoring stations
- Reason: Measurement devices are very expensive and regular maintenance required
- Few air quality monitoring stations do not provide adequate information on the spatial distribution of air pollutants
- Mobile measurements and low-cost sensor networks can be used to study spatial and temporal variability of air pollutants
- Urban climate models can be used to simulate pollutant concentrations on a larger scale
- Machine learning techniques can be applied to estimate pollutant concentrations



Source: https://www.lubw.baden-wuerttemberg.de/luft/messwerte-immissionswerte#karte

Problem Identification

- Limited data with low spatial resolution is available on weather and climate as well as aerosols and air pollutants
- There is hardly any air pollutant concentration prediction with machine learning techniques done in order to complement the existing monitoring stations

Problem Solution

- Collecting extensive dataset (training, testing and validation) with the aim to solve above mentioned problems
- Developing a technique (machine learning model) by using a part of this measurement data as input for modelling that can serve as a simulated monitoring station in order to complement existing monitoring stations

UNSW SYDNEY

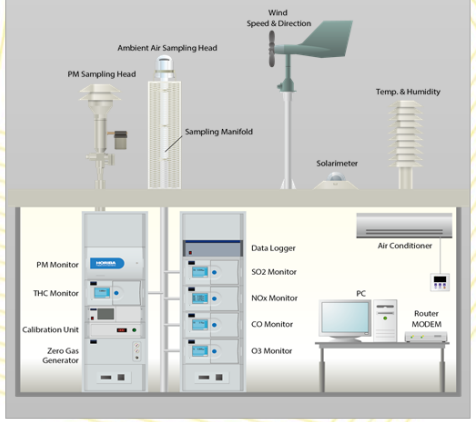
IAUC

Machine Learning Technique – Methodology

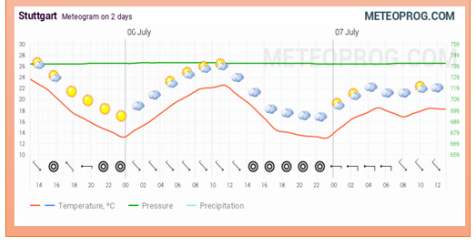
Air pollution prediction using machine learning techniques – An approach to replace existing monitoring stations with virtual monitoring stations

Deterministic Local Inputs

Air quality monitoring networks



Meteorological measurements



Machine Learning Model Inputs

Calendar data

Hour of the day, Day of the week, Week and Month of the year

Meteorological data

Air temperature, Relative humidity, Air pressure, Precipitation, Wind speed

Traffic

Number of vehicles per hour

Nearby air quality networks

Mean, minimum and maximum of NO, NO_2, PM10 and O_3 concentrations from nearby air quality monitoring networks

Pollutant concentration estimation using Machine Learning models

- Ridge
- Support Vector
- Random Forest
- Extra Trees
- XG Boost

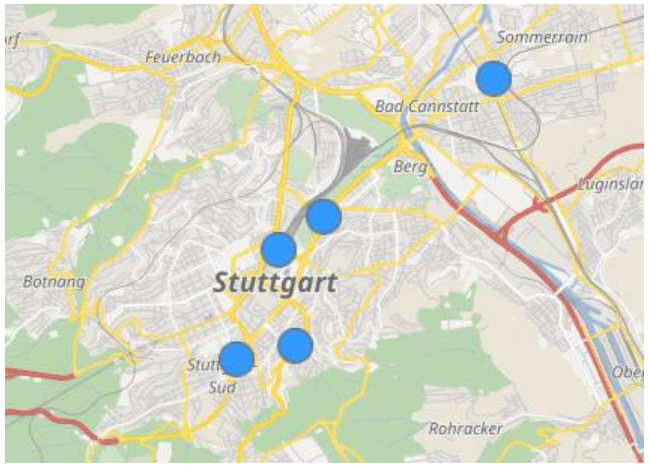
Evaluation

Comparision between actual and modeled data

Performance measures:

R^2 value
MAE
MAPE
MSE

Machine Learning Technique – Methodology

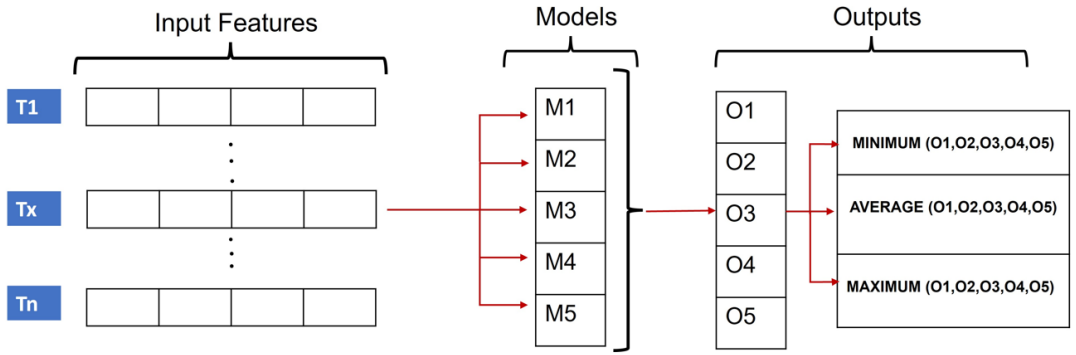


Modelling Scenarios

- Scenario 1: Meteorological Data
- Scenario 2: Meteorological Data + Traffic
- Scenario 3: Meteorological Data + Traffic + Same pollutant from one station
- Scenario 4: Meteorological Data + Traffic + Pollutants from nearby stations

Pollutants to be modelled

- $PM_{2.5}$, PM_{10} and NO_2



Locations

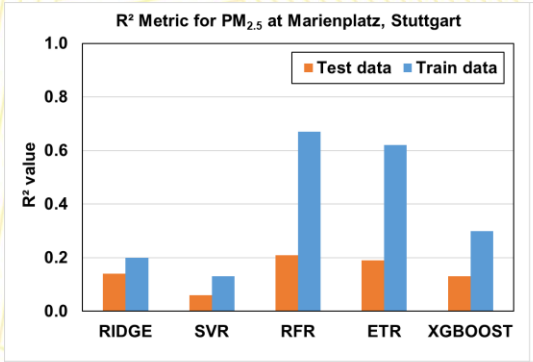
- Marienplatz and Am Neckartor

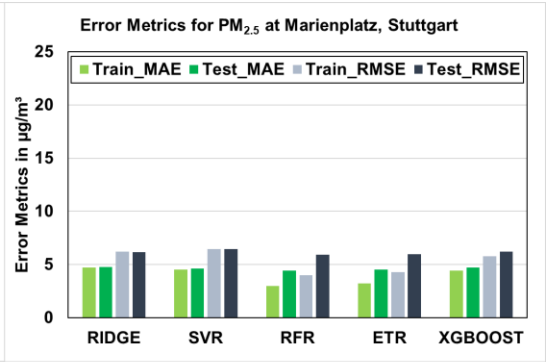
Training and Test data set

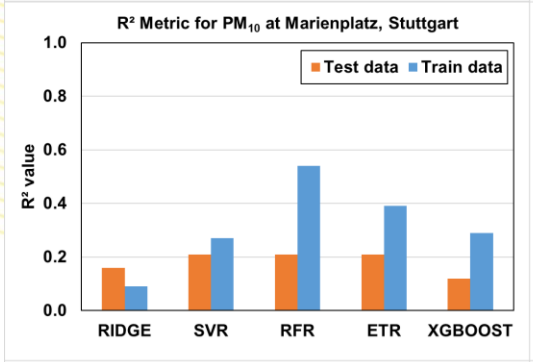
- Training data set : 1st Jan 2018 – 31st March 2021
- Test data set : 1st April 2021 – 31st March 2022

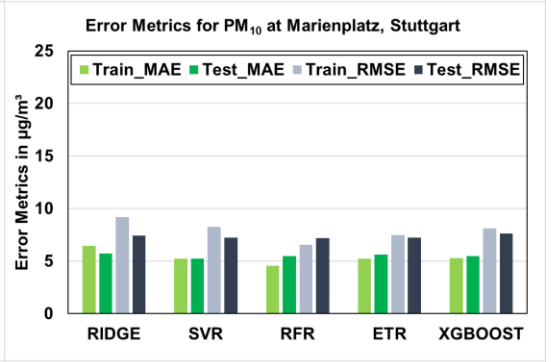
Machine Learning Model Comparison

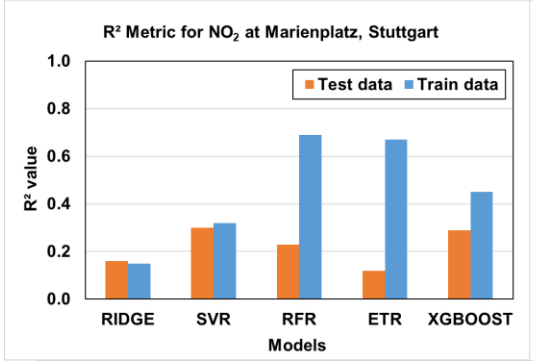
Scenario 1

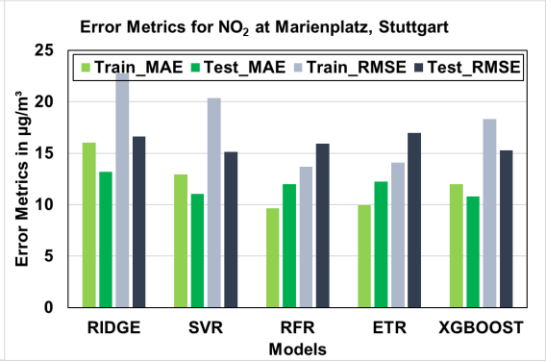




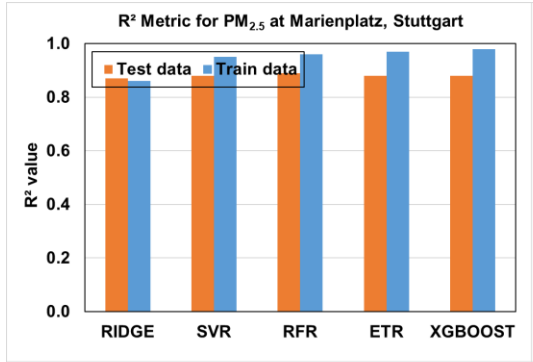


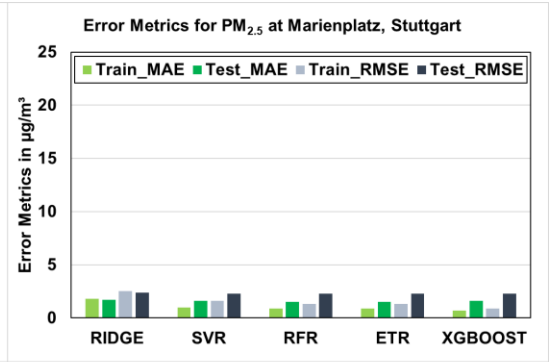


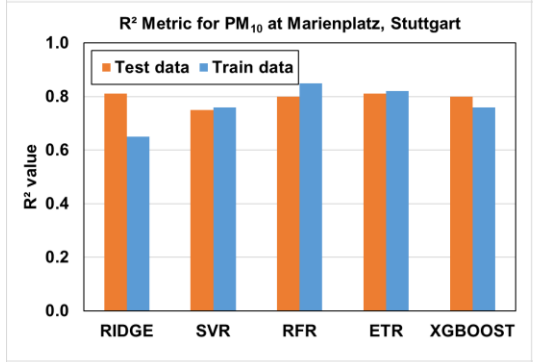


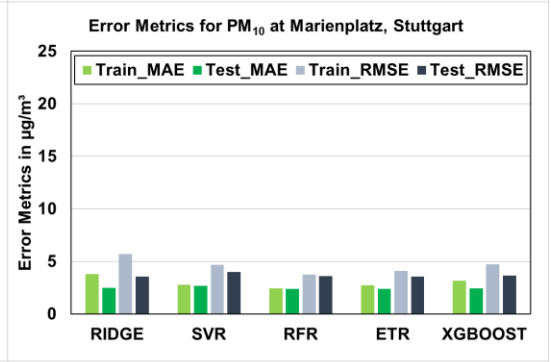


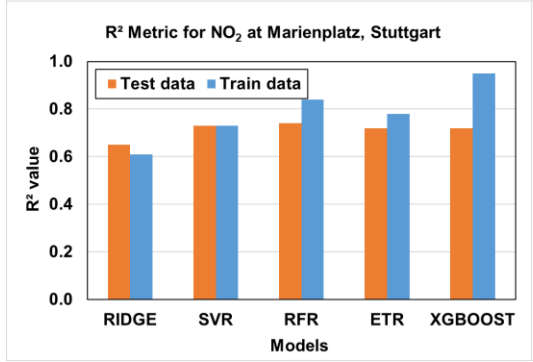
Scenario 4

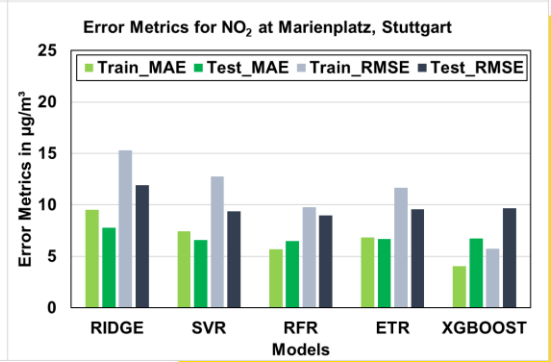






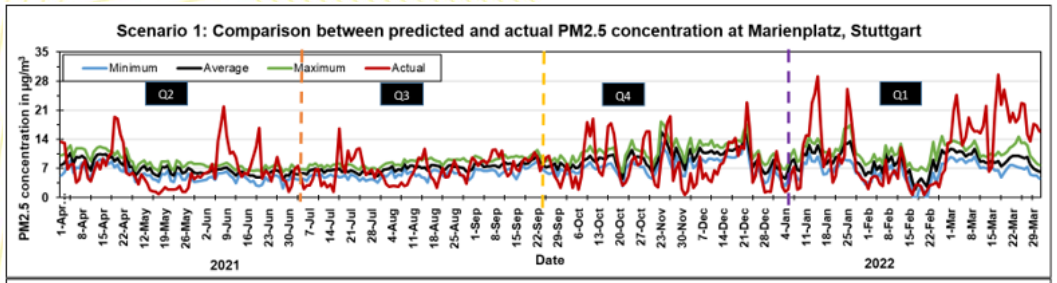


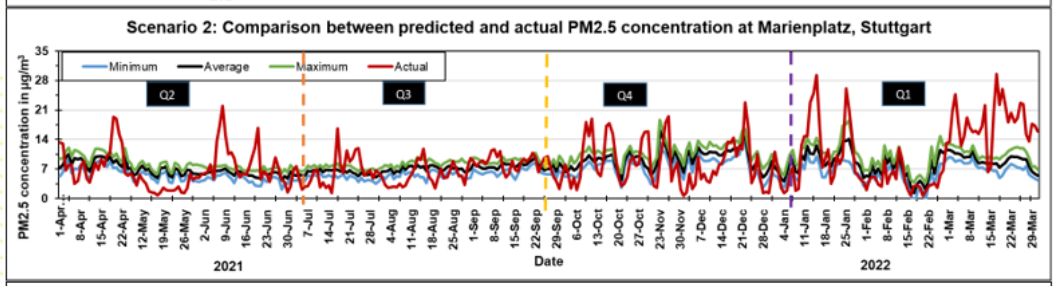


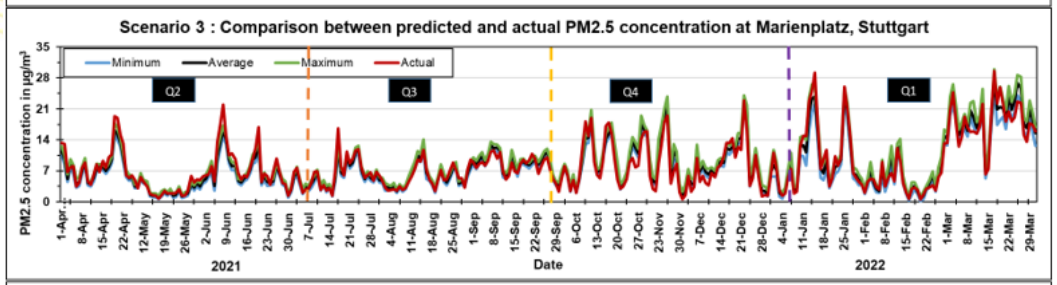


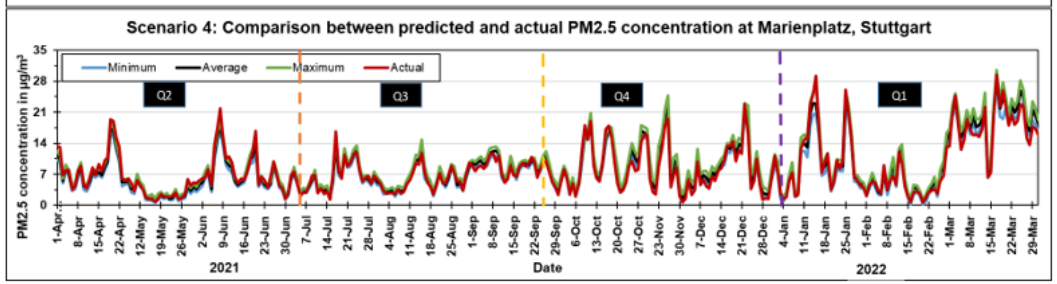
Machine Learning Model Comparison

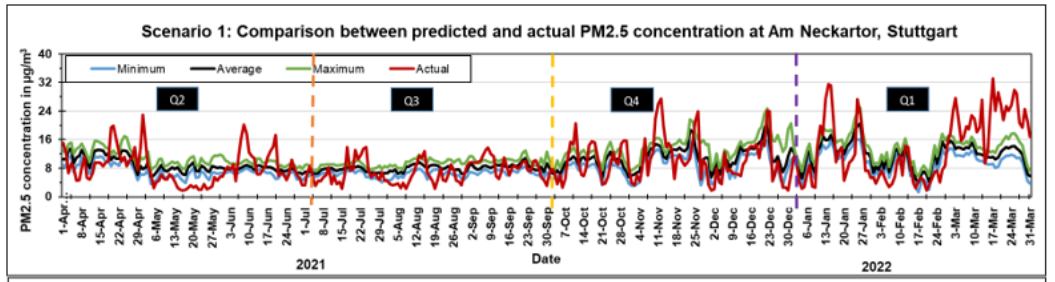
Pollutant comparison – $PM_{2.5}$

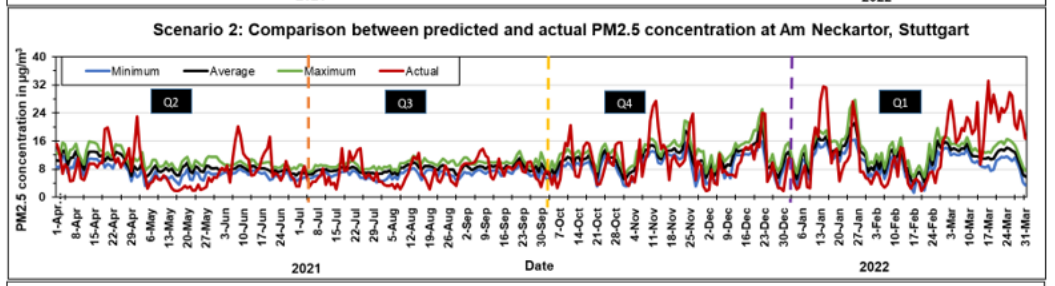


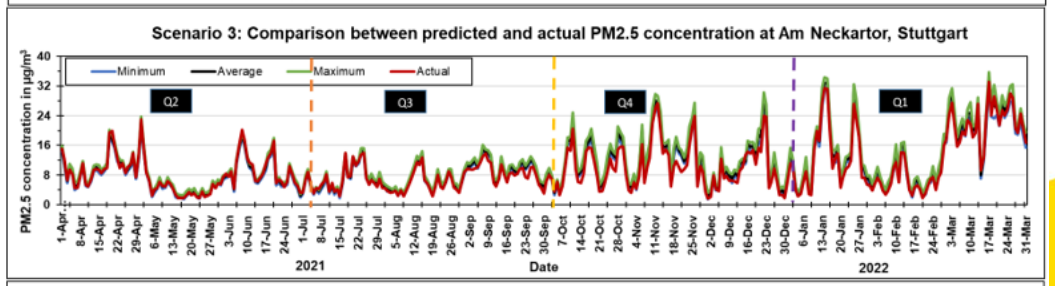


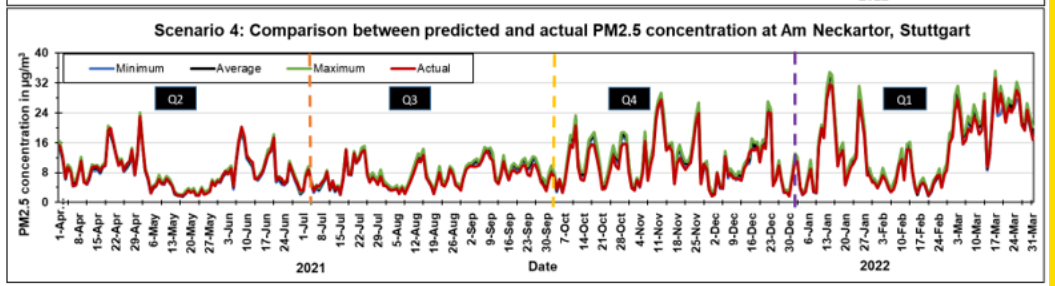






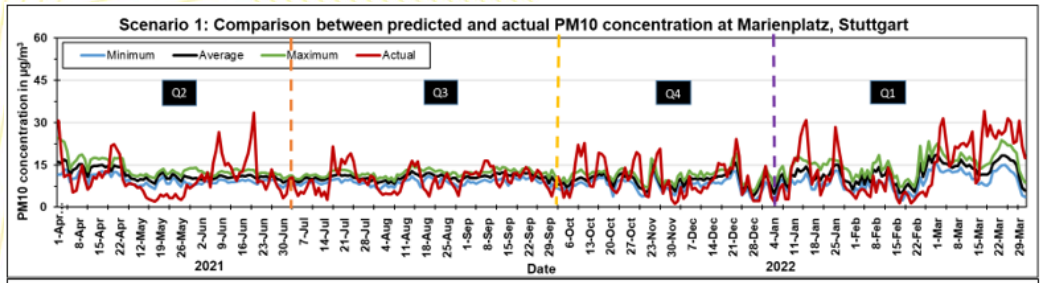


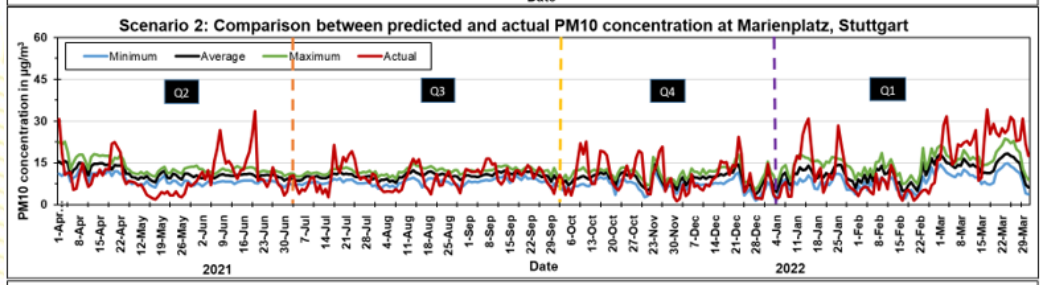


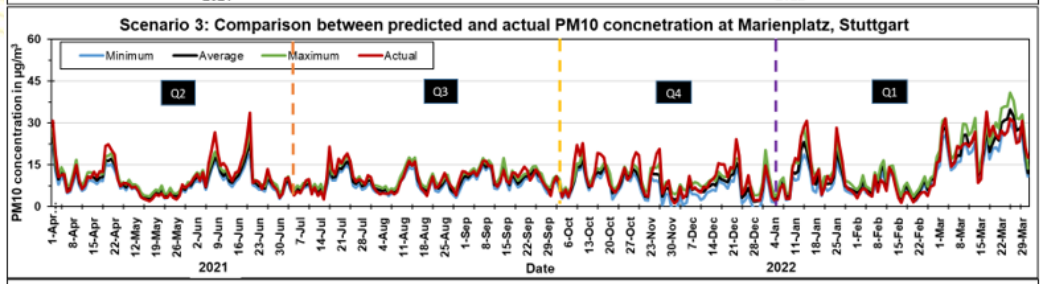


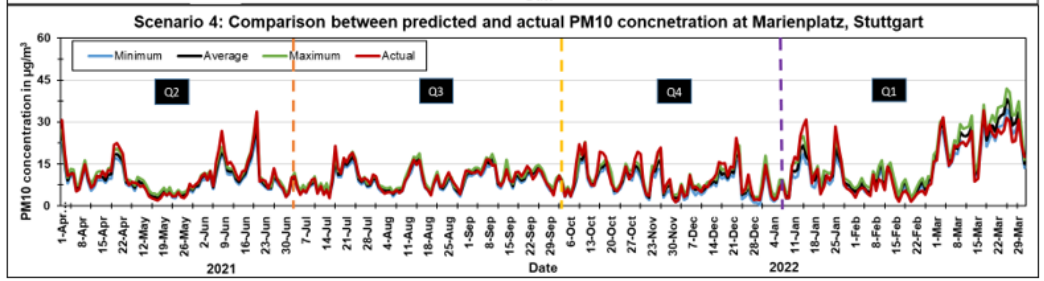
Machine Learning Model Comparison

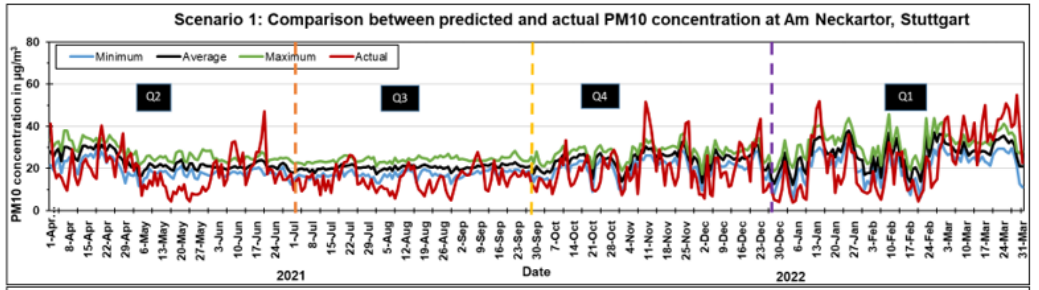
Pollutant comparison – PM_{10}

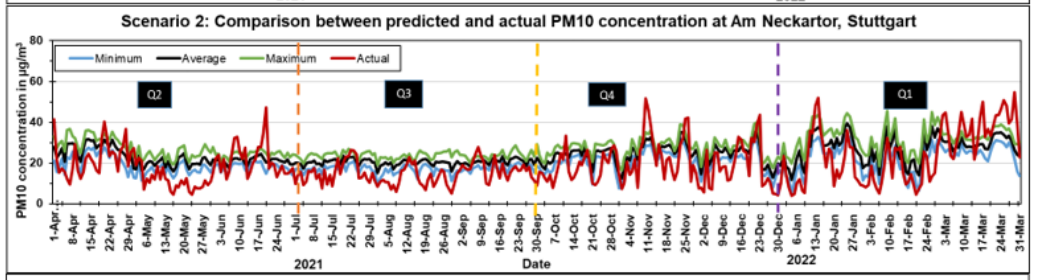


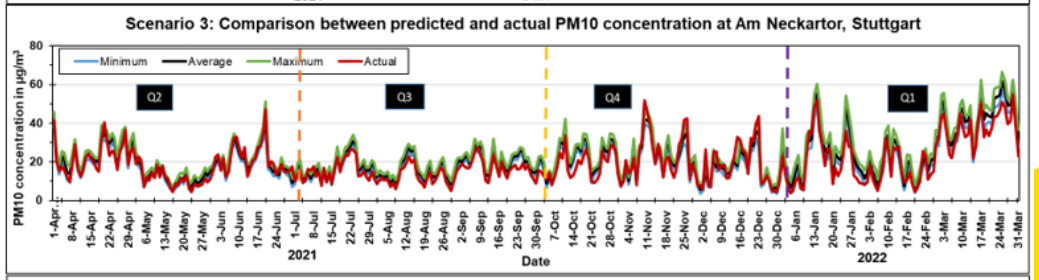


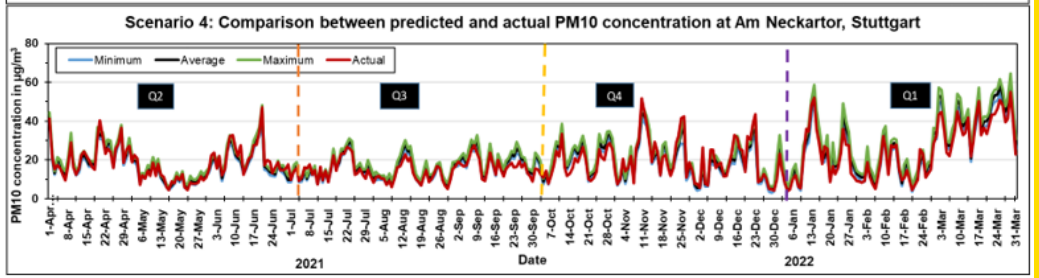






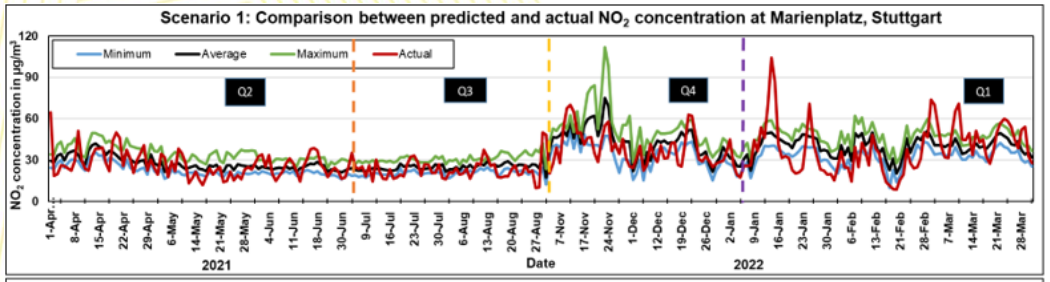


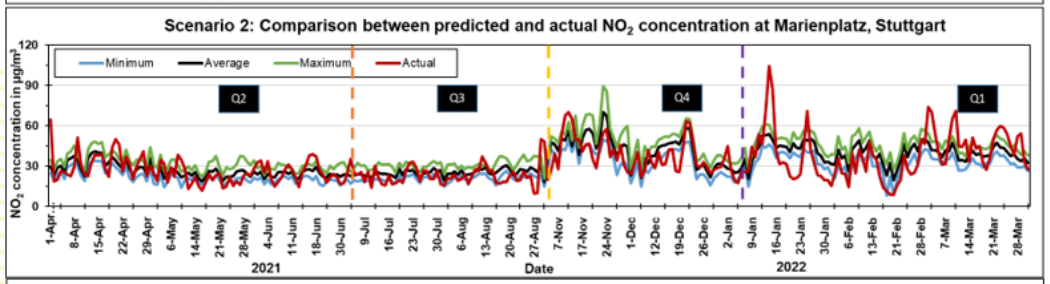


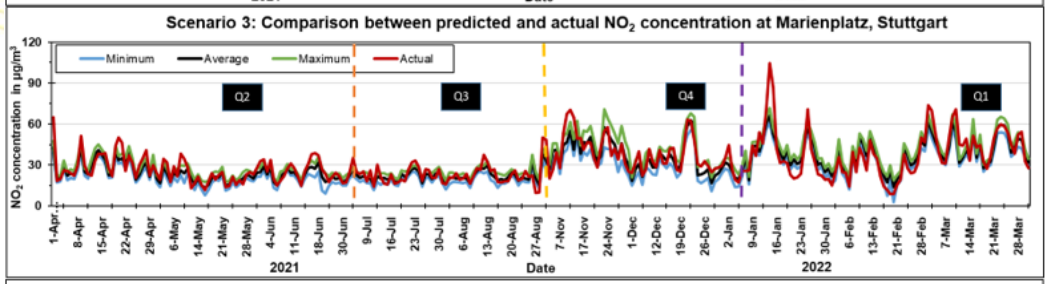


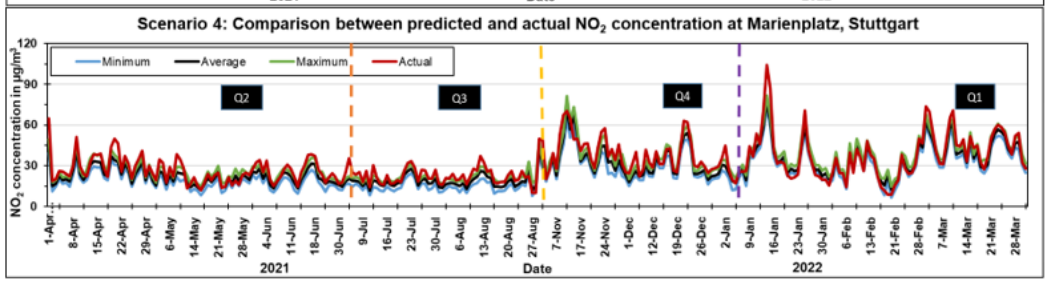
Machine Learning Model Comparison

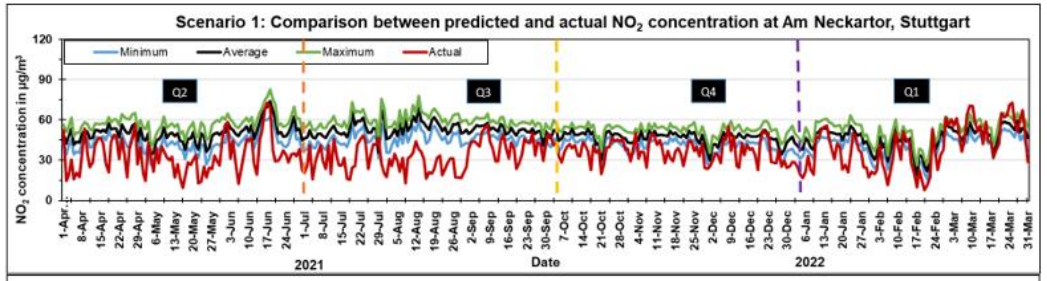
Pollutant comparison – NO_2

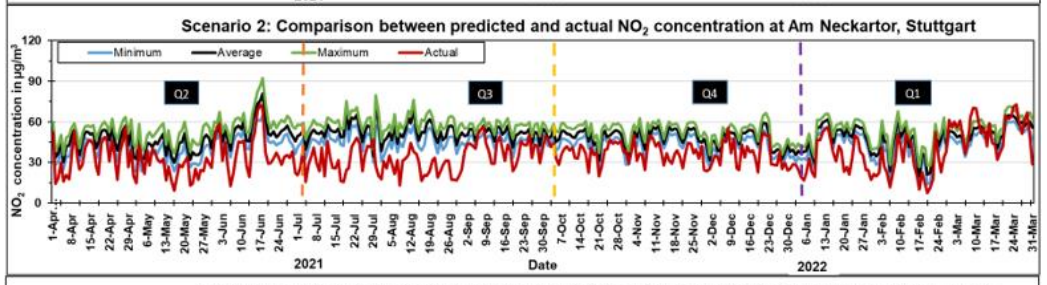


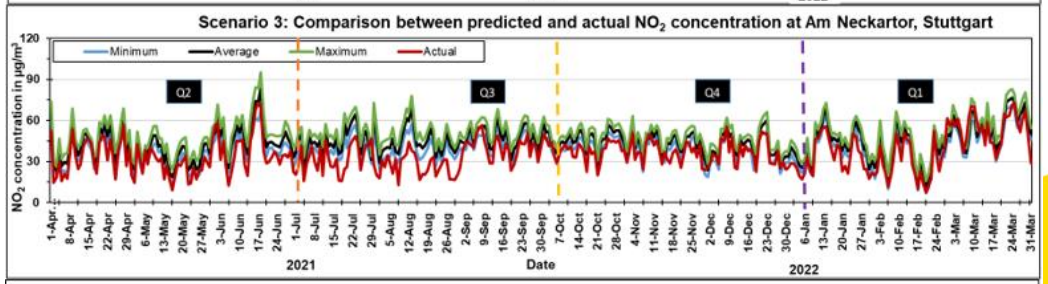


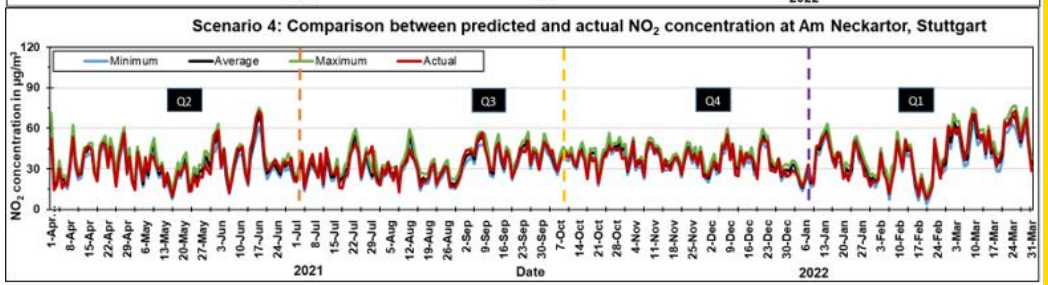






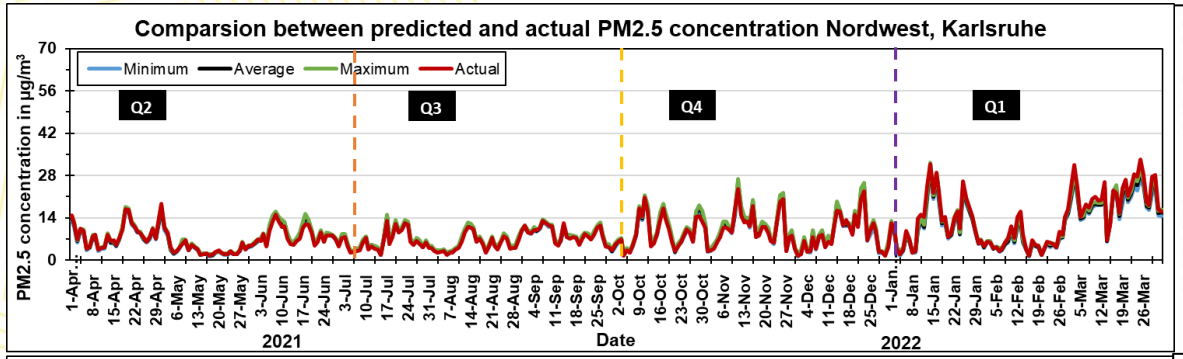


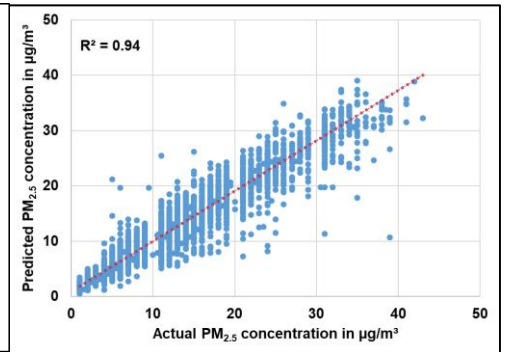


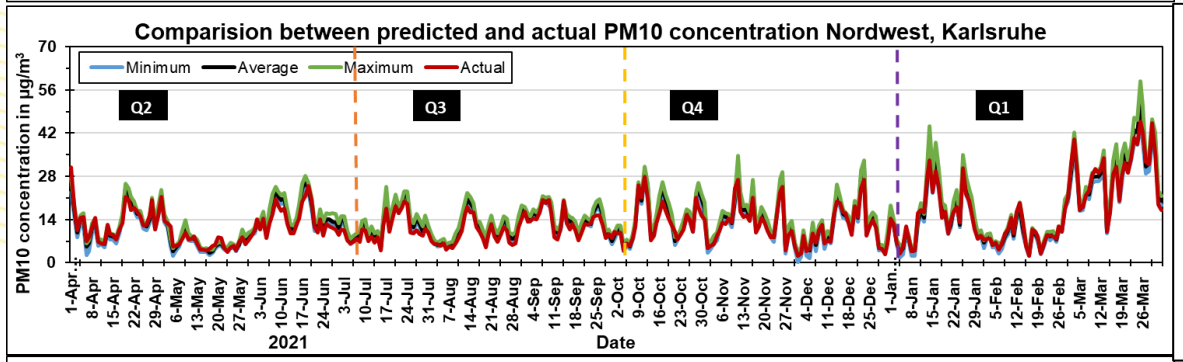


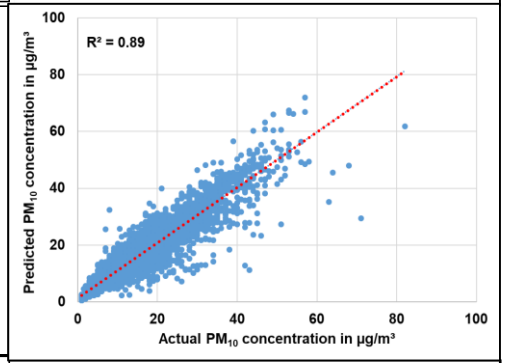
Machine Learning Model Comparison

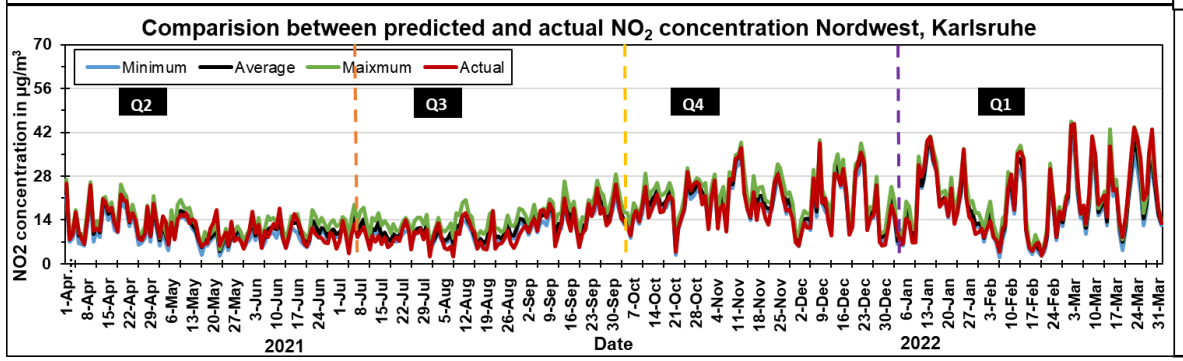
Pollutant comparison – $PM_{2.5}$, PM_{10} and NO_2 at Nordwest, Karlsruhe

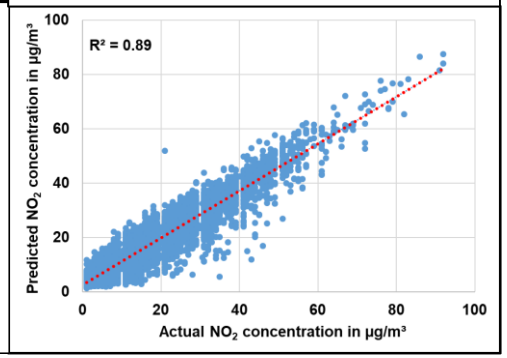












Conclusions

- The field measurements are an important contribution to the data set for individual investigations and model validation
- The developed machine learning model captured most of the trends and achieved a decent generalizing ability in predicting the pollutant concentrations
- The pollutants from other monitoring stations as an input feature (Scenario 4) played a significant role in estimating the pollutant concentration
- The developed technique can be applied to other locations where pollutant prediction is required
- This methodology can be used to produce multiple simulated monitoring stations that can complement the existing air quality monitoring networks and provide air pollutant concentration data with a high spatial resolution

Thank you for your attention!

M.Sc. Abdul Samad

e-mail abdul.samad@ifk.uni-stuttgart.de
phone +49 711 685-63397
fax +49 711 685-63491

University of Stuttgart
Institute of Combustion and Power Plant Technology
Pfaffenwaldring 23 • 70569 Stuttgart • Germany

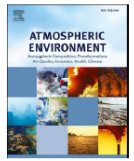
ifk

Atmospheric Environment 310 (2023) 119987

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv



Air pollution prediction using machine learning techniques – An approach to replace existing monitoring stations with virtual monitoring stations

A. Samad [a,*], S. Garuda [b], U. Vogt [a], B. Yang [b]

[a] Institute of Combustion and Power Plant Technology (IFK), Department of Flue Gas Cleaning and Air Quality Control, University of Stuttgart, Germany
[b] Institute of Signal Processing and System Theory (ISS), University of Stuttgart, Germany

UNSW SYDNEY

IAUC