Active Learning for Air Quality Station Deployment

S. Deepak Narayanan, Apoorv Agnihotri, Nipun Batra

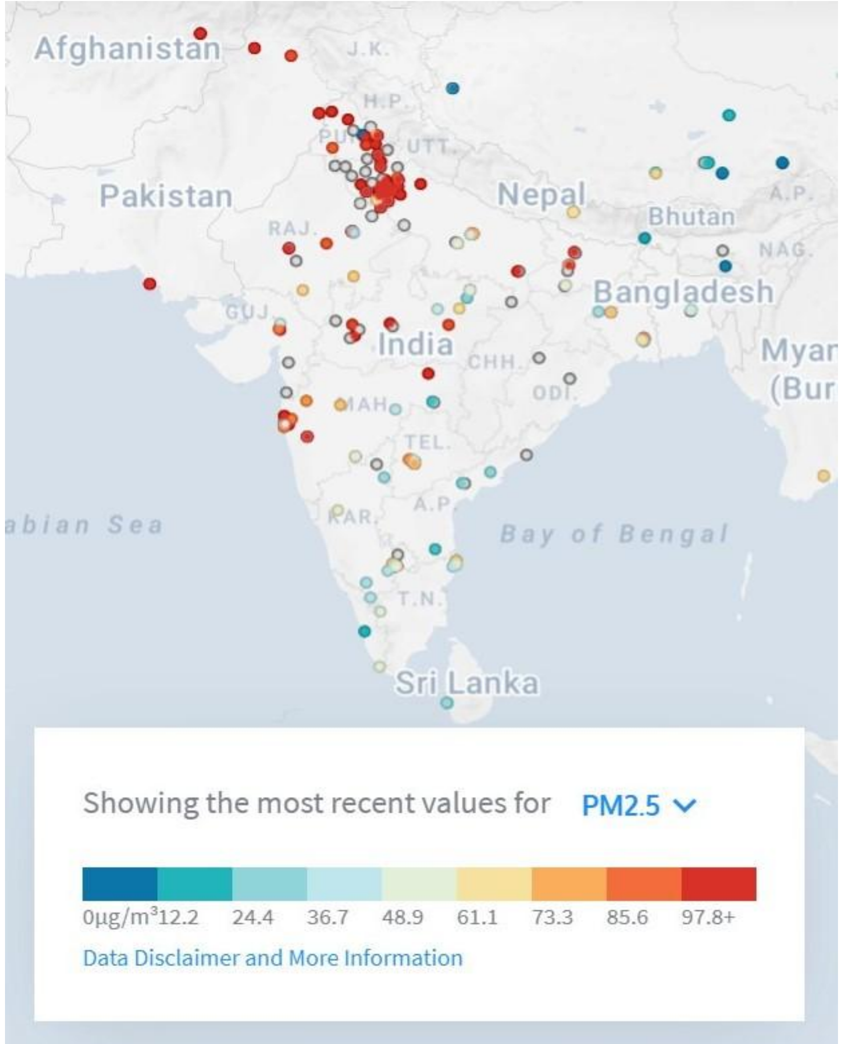
Indian Institute of Technology (IIT) Gandhinagar

Motivation

- 7 million people die every year due to air pollution

Motivation

- 7 million people die every year due to air pollution
- Extremely sparse deployment



Problem Statement

Where do we install the next set of air quality monitoring stations/sensors so that we can best infer the air quality at unknown locations?

Approach

- Gaussian Processes (GPs)
    - Custom Kernels help encode domain knowledge

Approach

- Gaussian Processes (GPs)
    - Custom Kernels help encode domain knowledge

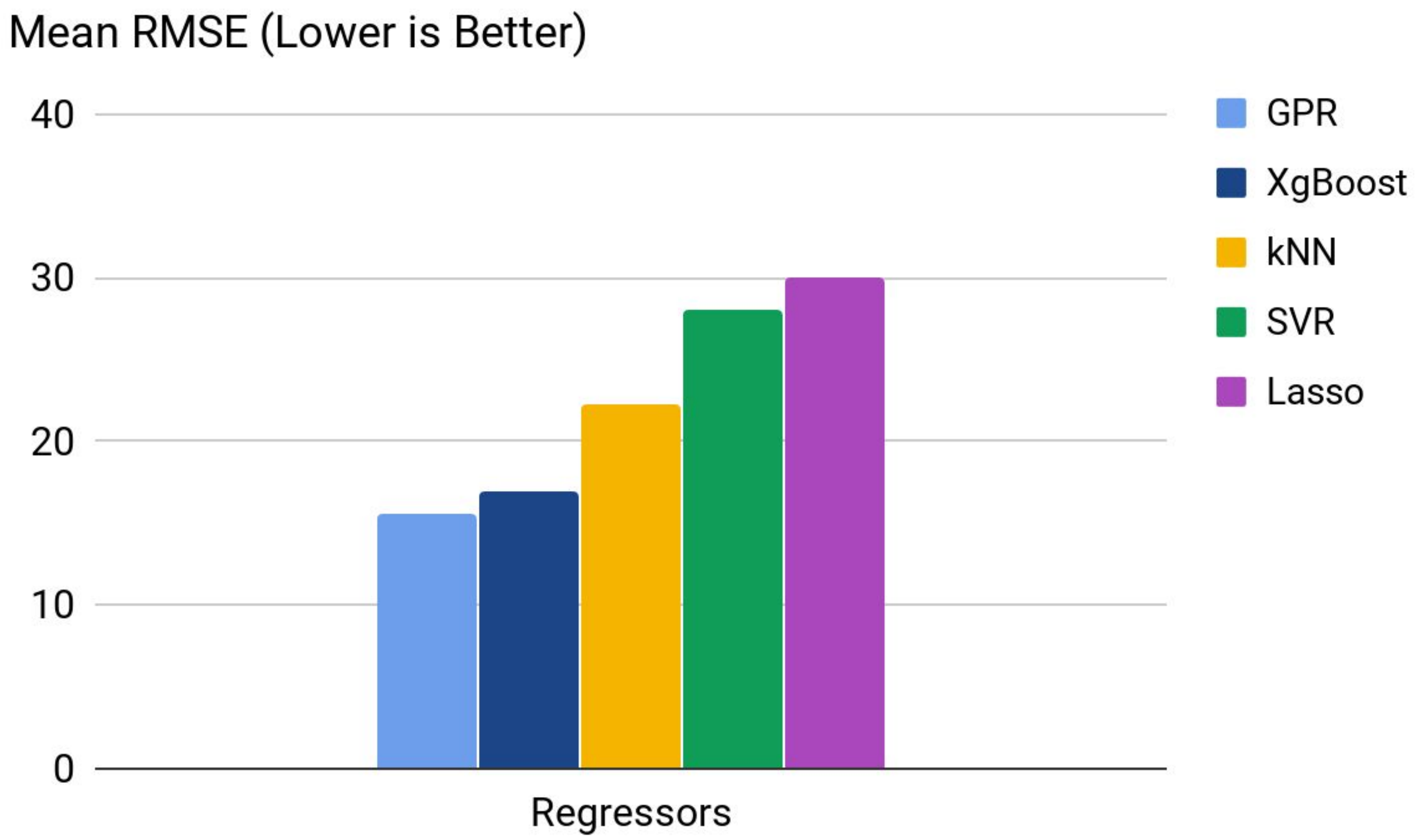
- Uncertainty Sampling
    - Actively choosing locations using predictive posterior variance

Evaluation - Data set

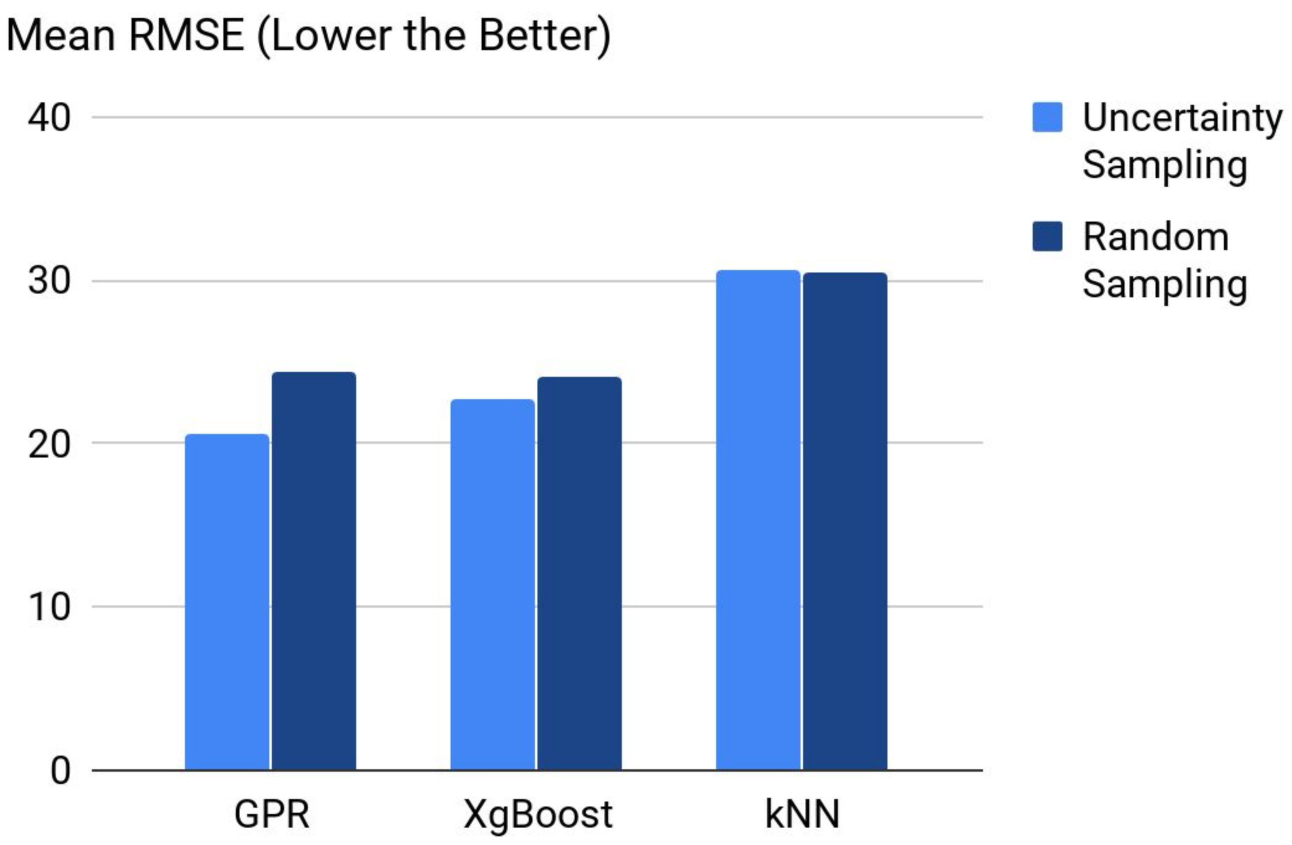
- Publicly available Beijing dataset from 36 stations from 2014-2015
- Features - locations, humidity, wind speed, wind direction, pressure, and temperature

Evaluation - Spatiotemporal Interpolation

- Efficacy of the Gaussian Process Regressor



Evaluation - Active Learning



Future Work

- Different costs for different stations
- Imposing budget constraints
- Automate the process of selecting kernels while modeling air quality

Summary

- Air Quality monitoring is an important problem to solve.
- Spatial monitoring of air quality is very sparse
- We proposed a GP to actively deploy air quality monitors and demonstrate promising results

Contact:{**deepak.narayanan, apoorv.agnihotri, nipun.batra**}@**iitgn.ac.in**