Learn from cities, and optimize them.

—An overview of data extraction methods in urban environments

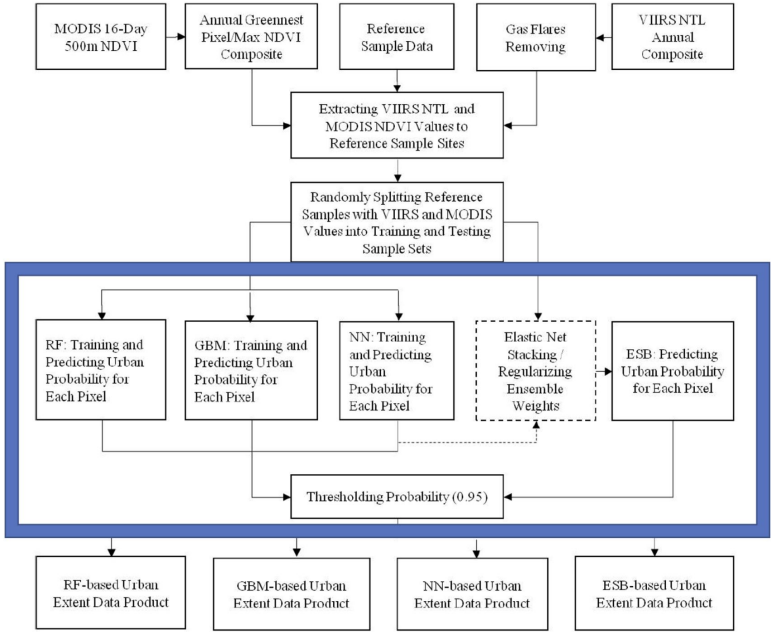
Jason Bao

The Problem

- Booming population of cities
  - In 2020, 56.2% of world population was urban
- Importance of cities
  - Economic Prosperities
  - Social Movements
- Backlash of rapid urbanization
  - Traffic congestion
  - Air pollution
  - Overcrowdedness
  - Unable to solve it efficiently due to limits of data.
- Data are useful because of Data Science.
- Utilization of data science methodologies to extract valuable urban data, to serve as reference for policy makers to improve city environment.

The Methods: Machine Learning

- Introduction
  - Automatically-improved algorithms through experience and by the use of data
- Mechanism
  - Weighted vote of various models
- Case Study: Conterminous United States
  - Objective: large-scale urban extent mapping at an intermediate resolution (500 m)
- Outcome & efficiency
  - Accuracy: >95% from past data



The Methods: Agent-Based Modeling

- Introduction
  - Models for simulating the interactions of autonomous agents
- Mechanism
  - Observe the effects at a higher level
- Case Study, Hamburg City
  - Potential human exposures to environmental stresses (pollutants)
- Outcome & efficiency
  - Offer advice on citizen behaviors after simulations

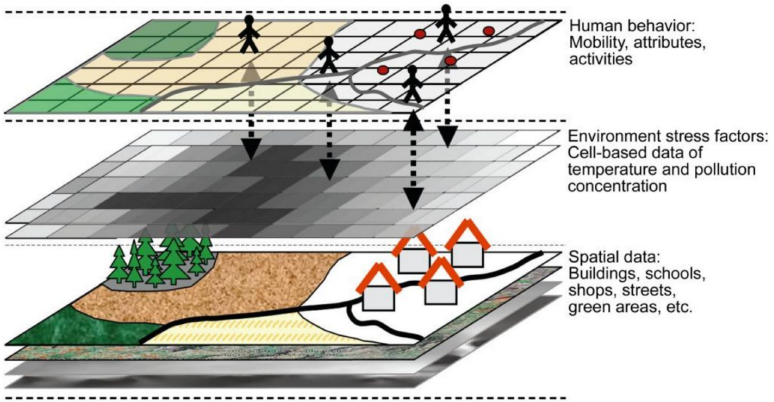


Figure 1. Illustration of the agent-based model framework for environmental exposure simulation (adapted and reedited from Leyk et al. [45]).

The Gap & Future

- Gap of current methodologies
  - Verification & Applicability?
- Presentation of a board research plan
  - Integration of the two data science methodologies
  - Predict & Observe the effects after the implementation of a hypothetical policy
  - More informed city planners, better policy & infrastructure for citizens
- A testable version of city.