Predicting What We Breathe: Machine Learning to Understand Urban Air Quality

Presenter: Jeanne Holm

PI: Jeanne Holm

Team Members: Dawn Comer, Anthony Lyons, Irene Burga, Aaron Gross @ City of Los Angeles; Mohammad Pourmahayoun @ Cal State LA; Christa Hasenkopf @ OpenAQ; Ian Brosnan and Mike Little @ NASA

Program: AIST-18

Problem to Solve

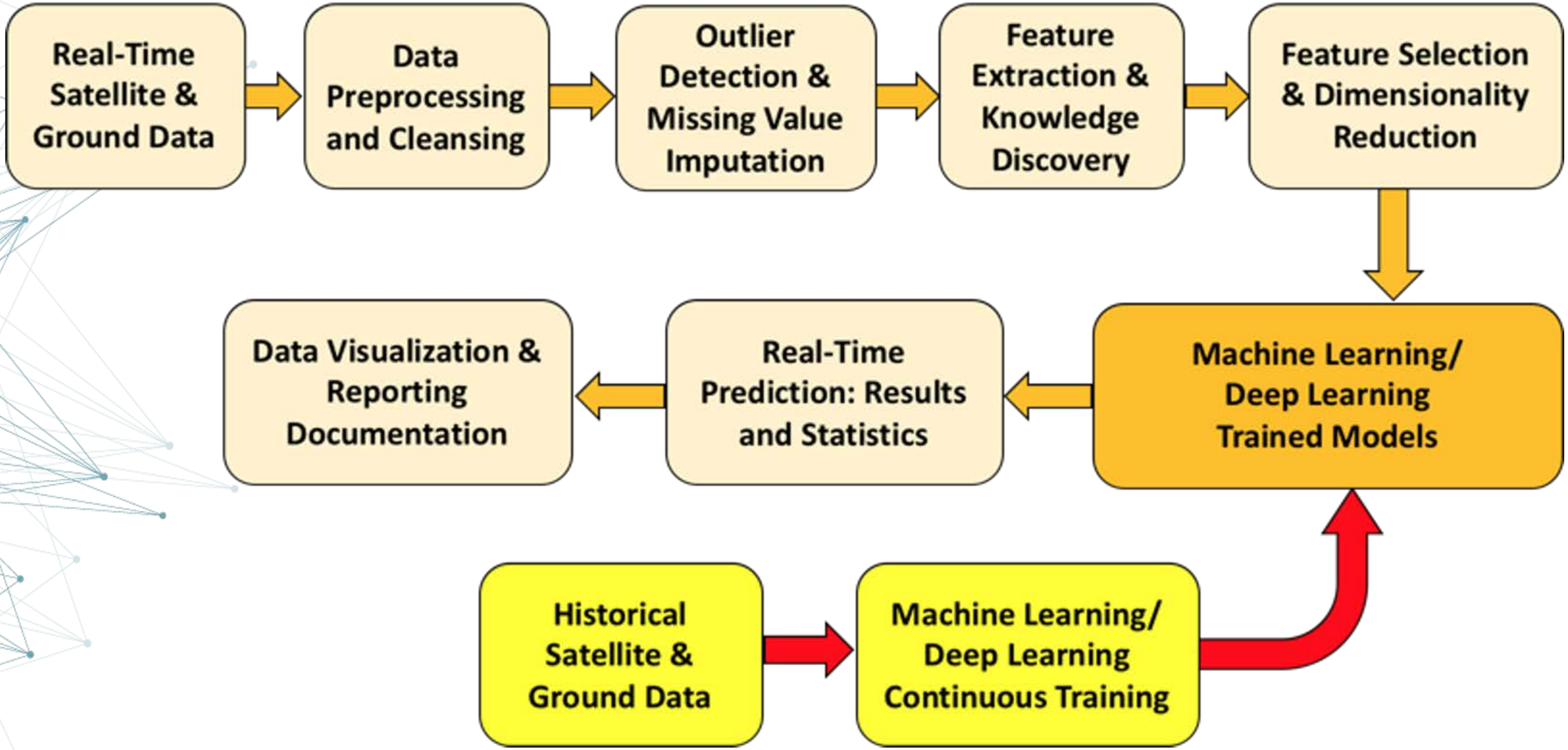
- Inform policies to improve health outcomes in L.A.
- Create an actionable global air quality index
- Increase accessibility and use of space data by using machine learning to help cities predict air quality in ways that will improve human health
- Provide tools and algorithms to future missions (such as MAIA) for rapid ground truth, conduct diverse data fusion, and support rapid use of mission data
- Current approaches lack
  - City-to-city collaboration on effective AQ control strategies
  - Accurate predictive capabilities
  - Localized urban scale data

Characterize, understand, and improve the quality of air in urban areas across the planet

Solution

1. Create a model for cities to integrate air quality data from ground and space-based measurements
2. Apply machine learning to large datasets to predict air quality and relate to on-the-ground interventions and activities
3. Improve decision making for local governments

Structure for Machine Learning Models

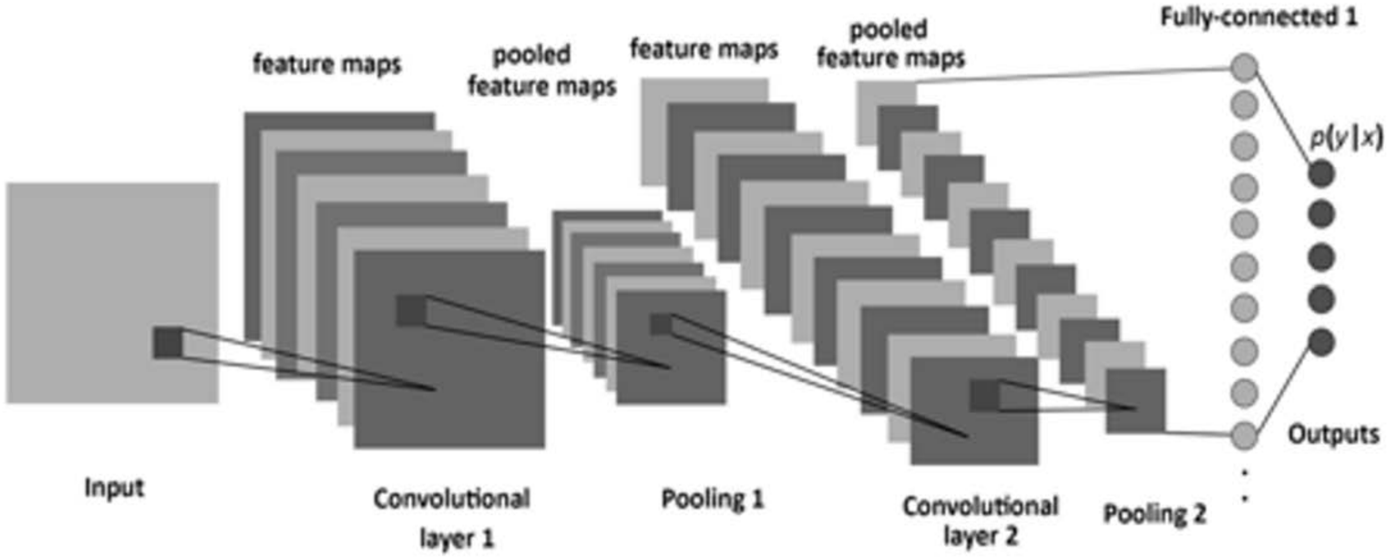


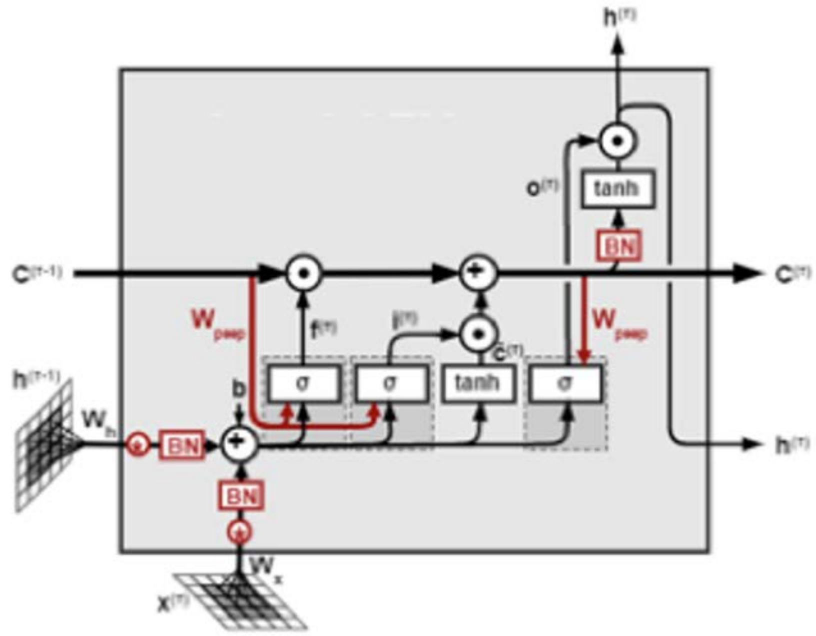
[Ref]: P. Muthukumar, E. Cocom, J. Holm, D. Comer, A. Lyons, I. Burga, Ch. Hasenkopf, and M. Pourhomayoun, “Real-Time Spatiotemporal Air Pollution Prediction with Deep Convolutional LSTM through Satellite Image Analysis,” The 16th Int. Conference on Data Science (ICDATA’20), 2020.

Machine Learning Models

<u>Deep Neural Networks:</u>

- <u>Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM)</u>: For the **temporal** correlation in the data
- <u>Convolutional Neural Network (CNN)</u>: For the **spatial** correlation in the data
- <u>Convolutional RNN/LSTM</u>: For the **spatiotemporal** correlation in the data





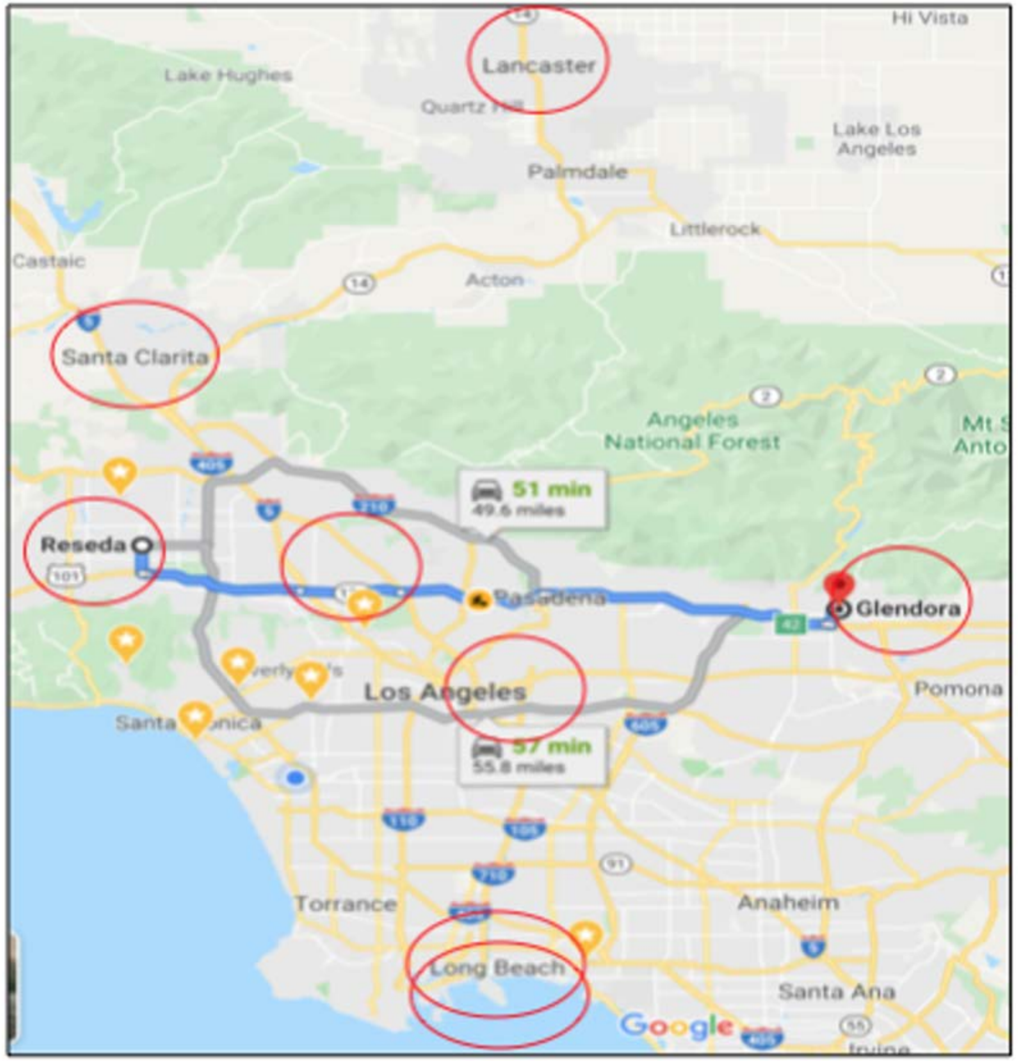
Sample Ground-Based Data

AQMIS Dataset (www.arb.ca.gov)

- PM 2.5
- 60,000 data samples
- Collected from 8 sensors
- On an hourly basis
- One year duration (1/1 to 12/31/2019)
- California Air Resources Board data

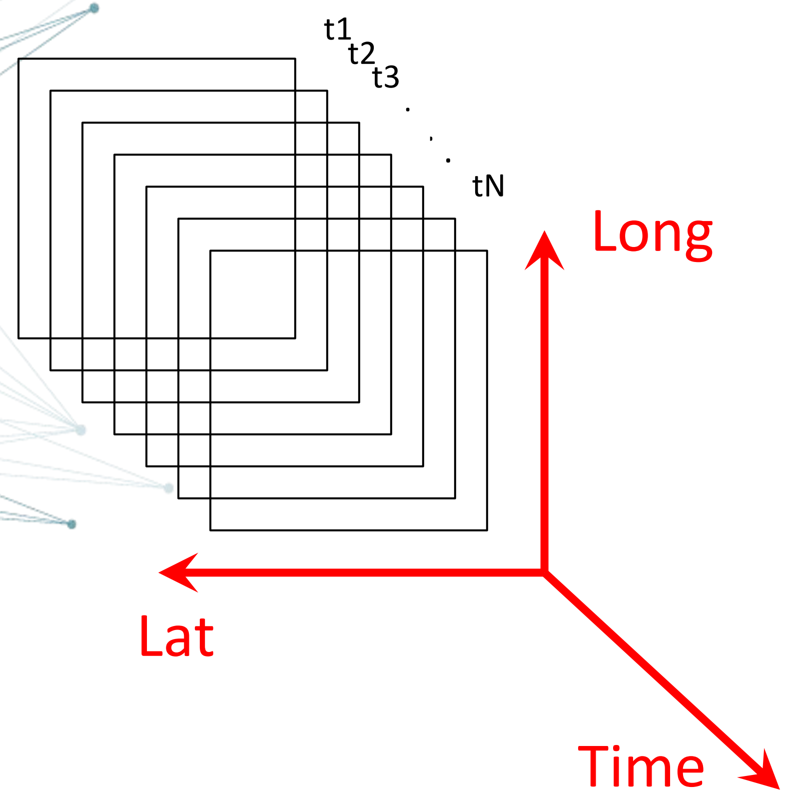
Deep Neural Network

- Several convolutional and recurrent layers
- 10 months for training
- 2 months for testing



Considering Temporal and Spatial Patterns in the Data

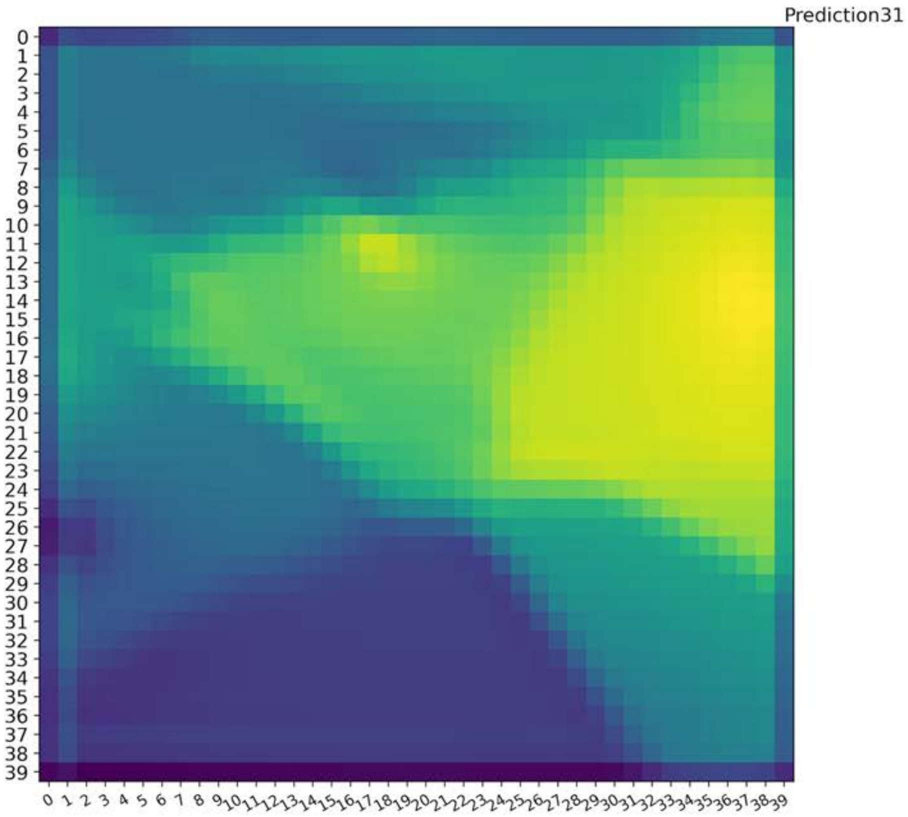
Temporal Correlation:



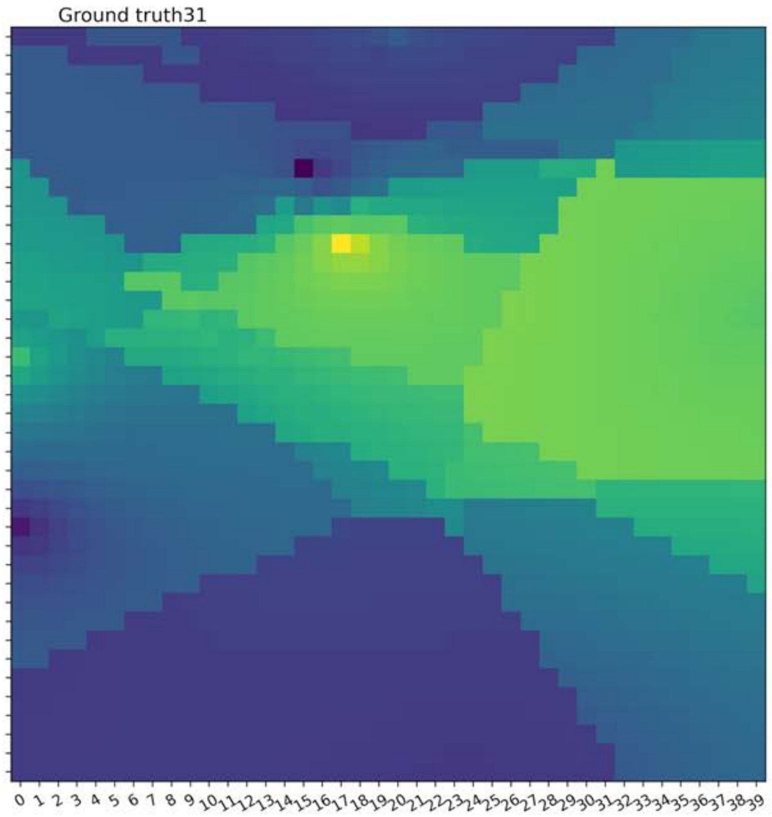
Spatial Correlation:

X	X	Lan-caster	X	X
X	X	X	X	X
Santa Clarita	X	X	X	X
Reseda	North Hollywood	X	X	Glendora
X	X	LA City Hall	X	X
X	X	Long Beach 2	X	X
X	X	Long Beach 1	X	X

10-Hour Prediction of PM 2.5

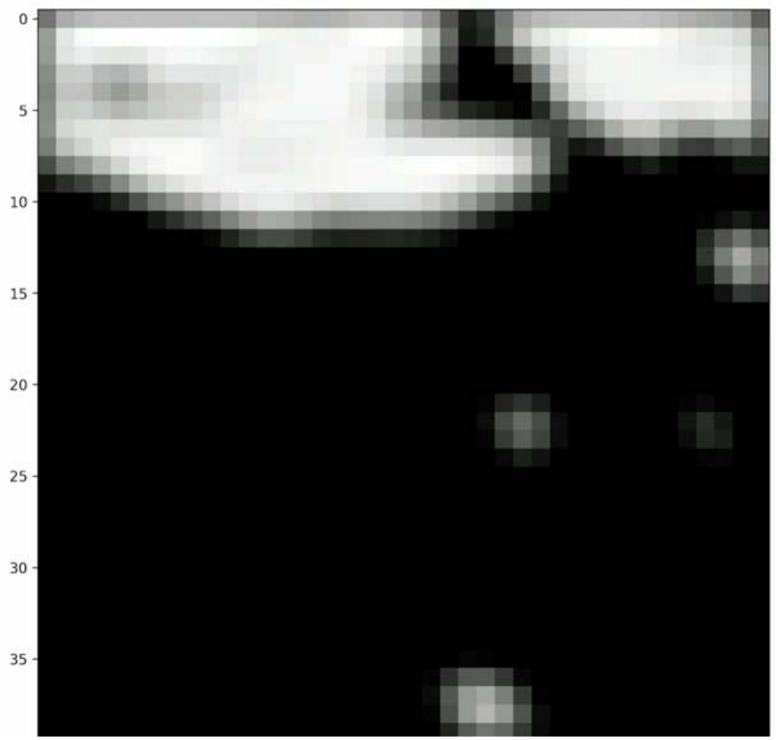


Prediction in 40x40 Grid

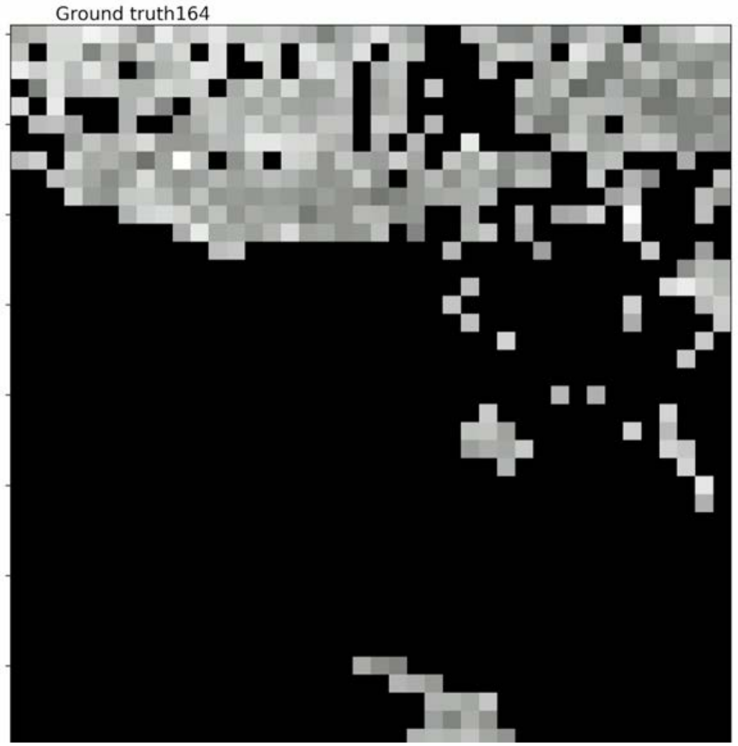


Ground Truth in 40x40 Grid

10-Hour Prediction of NO2



Prediction in 40x40 Grid



Ground Truth in 40x40 Grid

Next Steps

- Continue evolution of model, algorithms, and validation
- Identify and integrate local data (health, polluters, traffic, roads, ports) from IOT and in-situ sensors
- Identify gaps in coverage
- Engage citizen scientists (libraries, SafeCast, SmartAirLA, and more) and community partners for environmental justice for awareness and support
- Share findings via smart city air quality intervention and toolkit (C40 cities, U.N. Sustainable Development Goal Network, Climate Mayors, etc.) and identify sister cities

Opportunities for Collaboration

- Data source sharing via OpenAQ
- Machine learning models via Github
- City partners for results and lessons through C40 Global Platform
- Workshops at NASA ESTO Science Forum, AGU, IAC, and elsewhere

Jeanne.Holm@lacity.org
@JeanneHolm

Partners

- Public
  - City of Los Angeles
  - NASA/JPL
  - Southern California Air Quality Management District
  - SafeCast
- Private
  - OpenAQ
  - SmartAirLA
- Academic
  - California State University, Los Angeles
  - LA Data Science Federation
- Organizations
  - Mayor Garcetti leads the C40 Cities
  - Climate Mayors