U-Net For Learning And Inference Of Dense Representation Of Multiple Air Pollutants From Satellite Imagery

Jacquelyn Shelton, Przemyslaw Polewski and Wei Yao

Climate Informatics 2020

September 23rd, 2020

The Hong Kong Polytechnic University
香港理工大學

TomTom

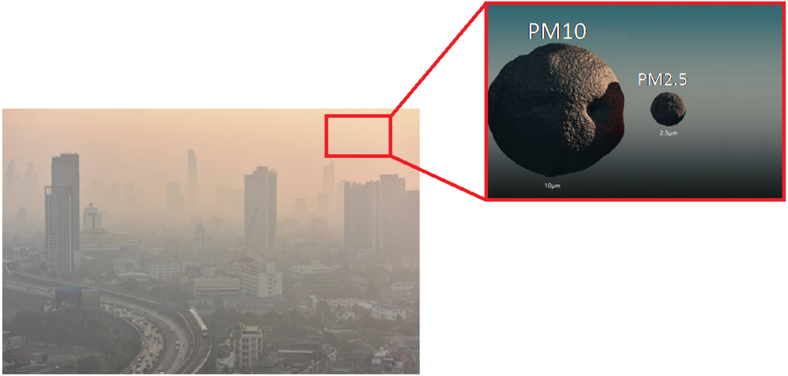
Motivation

New research linking pollution to COVID-19: lethality rate and as a transmission vector

→Wu, X., Nethery, R. C., Sabath, B. M., Braun, D., and Dominici, F. (2020) *Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study.* medRxiv.
→Coccia, M. (2020) *Two mechanisms for accelerated diffusion of COVID-19 outbreaks in regions with high intensity of population and polluting industrialization: the air pollution-to-human and human-to-human transmission dynamics.* medRxiv.

Motivation

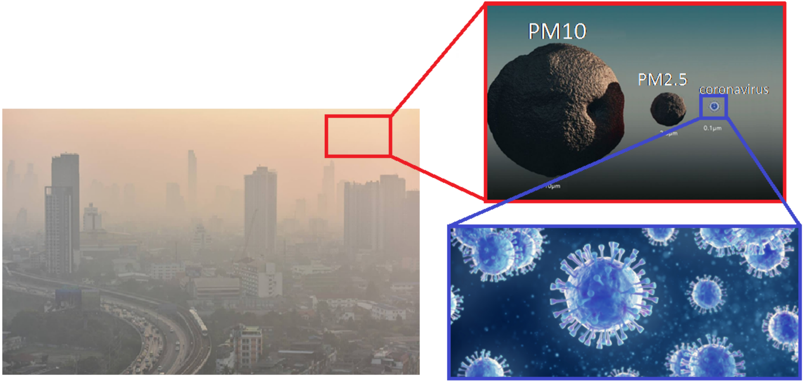
New research linking pollution to COVID-19: lethality rate and as a transmission vector



→Wu, X., Nethery, R. C., Sabath, B. M., Braun, D., and Dominici, F. (2020) *Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study.* medRxiv.
→Coccia, M. (2020) *Two mechanisms for accelerated diffusion of COVID-19 outbreaks in regions with high intensity of population and polluting industrialization: the air pollution-to-human and human-to-human transmission dynamics.* medRxiv.

Motivation

New research linking pollution to COVID-19: lethality rate and as a transmission vector

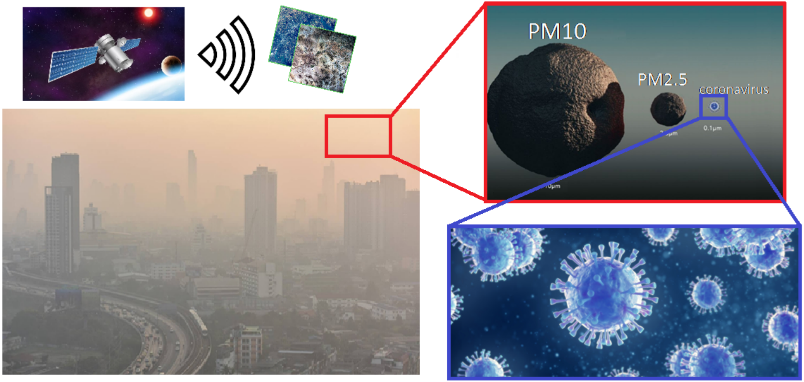


→Wu, X., Nethery, R. C., Sabath, B. M., Braun, D., and Dominici, F. (2020) *Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study.* medRxiv.
→Coccia, M. (2020) *Two mechanisms for accelerated diffusion of COVID-19 outbreaks in regions with high intensity of population and polluting industrialization: the air pollution-to-human and human-to-human transmission dynamics.* medRxiv.

Motivation

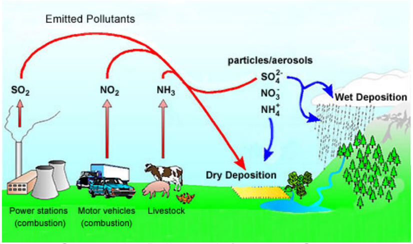
New research linking pollution to COVID-19: lethality rate and as a transmission vector

⇒ **Goal:** Machine learning to estimate air pollution concentrations at fine scale using satellite imagery



→Wu, X., Nethery, R. C., Sabath, B. M., Braun, D., and Dominici, F. (2020) *Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study.* medRxiv.
→Coccia, M. (2020) *Two mechanisms for accelerated diffusion of COVID-19 outbreaks in regions with high intensity of population and polluting industrialization: the air pollution-to-human and human-to-human transmission dynamics.* medRxiv.

Air pollution – From primary air pollutants to aerosols

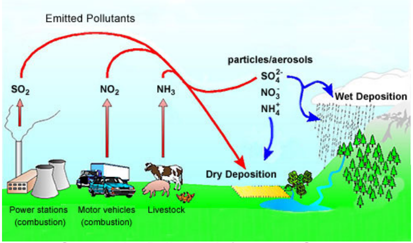


Source: Air Pollution Information System

Primary pollutants (ground emissions):

- Nitrogen oxides – NO_x
- Sulphur oxides – SOx
- Volatile organic compounds – VOC

Air pollution – From primary air pollutants to aerosols



Source: Air Pollution Information System

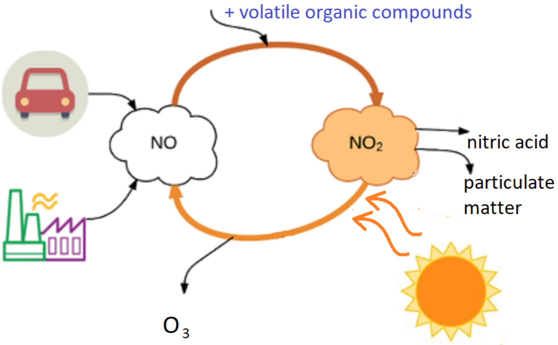
Primary pollutants (ground emissions):

- Nitrogen oxides – NO_x
- Sulphur oxides – SOx
- Volatile organic compounds – VOC

Secondary pollutants (form in atmosphere):

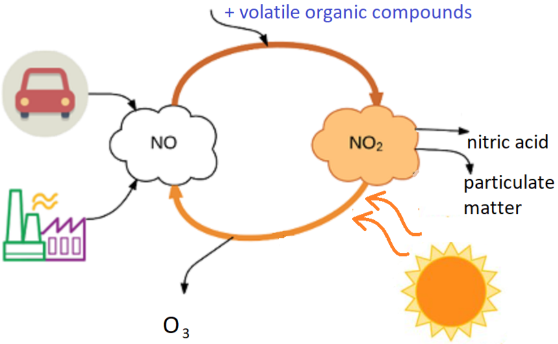
- React with each other
- Can travel long distances as aerosols
- Eventually return to the surface as wet or dry deposition

Air pollution – Nitrogen cycle in atmosphere



Source: World Air Quality Index

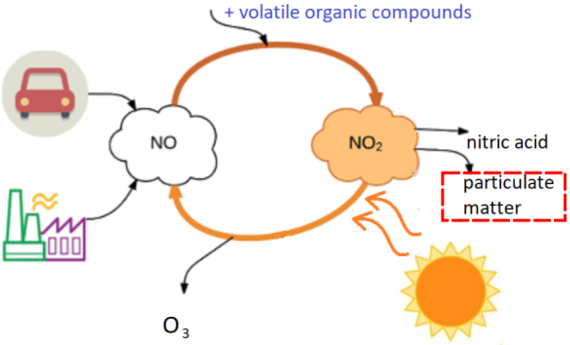
Air pollution – Nitrogen cycle in atmosphere



Source: World Air Quality Index

- Nitric acid and ammonia react to form NH_4^+, NH_3^-

Air pollution – Nitrogen cycle in atmosphere



Source: World Air Quality Index

- Nitric acid and ammonia react to form NH_4^+, NH_3^-
- Result: **aerosol** ammonium nitrate **contribute** to **particulate matter**

Air pollution – Consequences

…include the following:

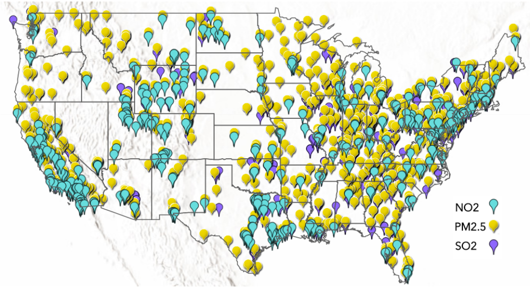
- NO_2 linked to respiratory problems in humans
- Nitric acid formation leads to acid rain
- Affects visibility by altering the way light is absorbed and scattered in the atmosphere
- Some constituents of the ambient PM mixture promote climate warming (e.g., black carbon) or cooling influence (e.g., nitrate and sulfate)
- Adversely affect ecosystems, including plants, soil and water through deposition of PM and its subsequent uptake by plants or its deposition into water, affecting water quality/clarity

Worldwide monitoring of air pollution

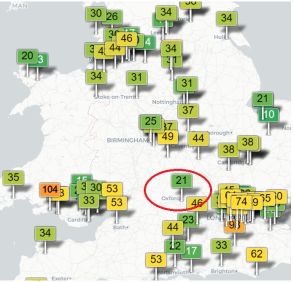
- Dedicated measurement stations for NO_x, SO_x, O_3, PM
- More than 30,000 stations in the World (source: World Air Quality Index)

Worldwide monitoring of air pollution

- Dedicated measurement stations for NO_x, SO_x, O_3, PM
- More than 30,000 stations in the World (source: World Air Quality Index)
- Low spatial resolution



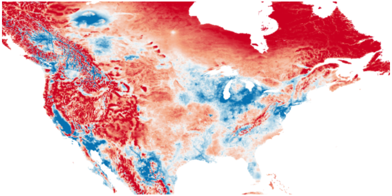
Unites States - locations of pollution sensors (epa.gov)

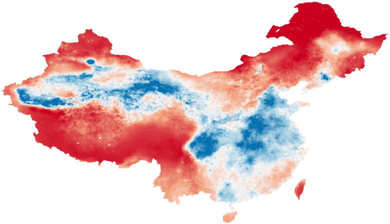


Daily pollution values in UK (waqi.info)

Pollution data – Ground-truth

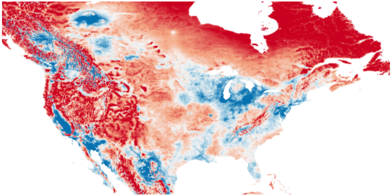
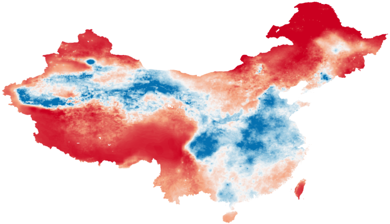
- Pollution database: (Aaron van Donkelaar et al., 2019) monthly estimates of $PM_{2.5}$ for North America and China, from 2001-2017 at 0.01° resolution (no further updates)





Pollution data – Ground-truth

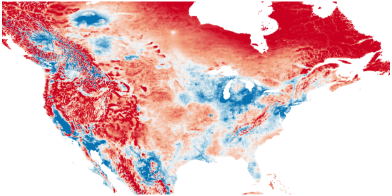
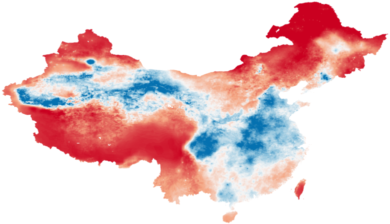
- Pollution database: (Aaron van Donkelaar et al., 2019) monthly estimates of $PM_{2.5}$ for North America and China, from 2001-2017 at 0.01° resolution (no further updates)

- Source: ground sensors, sat. imagery, chem. transfer model
- Partial masses of component pollutants $(NO_3^-, NH_4^+, \ldots)$

Pollution data – Ground-truth

- Pollution database: (Aaron van Donkelaar et al., 2019) monthly estimates of $PM_{2.5}$ for North America and China, from 2001-2017 at $0.01°$ resolution (no further updates)

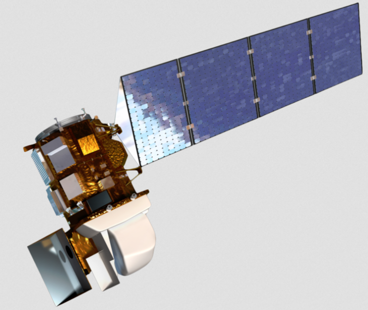



- Source: ground sensors, sat. imagery, chem. transfer model
- Partial masses of component pollutants $(NO_3^-, NH_4^+, \ldots)$

$\Rightarrow$ **Goal:** learn from this historical pollution data to predict new pollutant concentrations at fine scale

Multispectral satellite imagery - Landsat 8

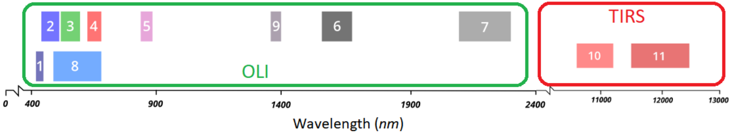
Landsat series: longest continuous satellite mission in operation (since 1972)



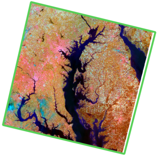
Source: US Geological Survey

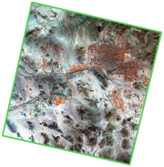
- Carries two independent optical instruments
  - Operational Land Imager
  - Thermal Infrared Sensor
- Worldwide coverage with 16 day revisit period
  - 2 images per month
- Ground sampling distance: 30 m (OLI) and 100 m (TIRS)

Multispectral satellite imagery - Landsat 8

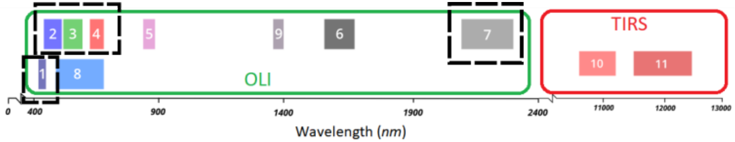


- Covers various parts of the electromagnetic spectrum from violets to thermal infrared with 11 spectral bands
- Publicly available from US Geological Survey
  - Mirrors hosted by Google/Amazon (for bulk download)





Multispectral satellite imagery – Pollution



Aerosol optical depth (AOD): measures attenuation of light rays passing through atmosphere

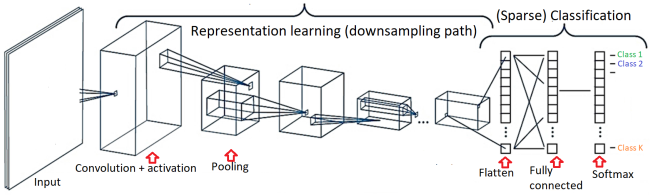
- Proxy for $PM_{2.5}$ content (light scattering on aerosol)
- Explicit estimation methods based on deep blue (#1), blue (#2), red (#4) bands
- Bands #3 and #7 have also been linked to PM levels

→Vermote, E., Justice, C., Claverie, M., Franch, B. (2016) *Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product.* *Remote Sens. Env.*
→Zhang, B.; Zhang, M.; Kang, J.; Hong, D.; Xu, J.; Zhu, X. (2019) *Estimation of PM_x Concentrations from Landsat 8 OLI Images Based on a Multilayer Perceptron Neural Network.* *Remote Sens.*

Fully Convolutional Networks (FCNs)

Classic Convolutional Neural Networks

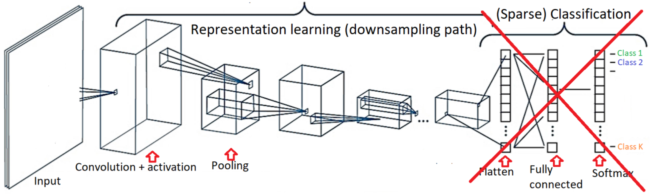
- Sparse representation learning (size independent)
- Classification with Fully Connected Layers (size-bound)



Fully Convolutional Networks (FCNs)

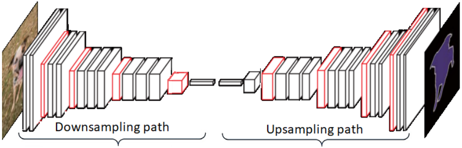
Classic Convolutional Neural Networks

- Sparse representation learning (size independent)
- Classification with Fully Connected Layers (size-bound)

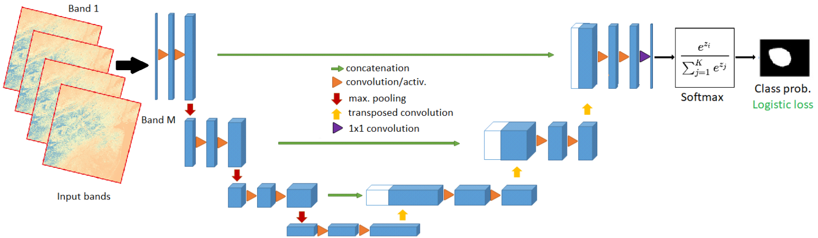


Fully Convolutional Networks

- Only size-independent operations
- Dense (per-pixel) predictions (non-sparse)

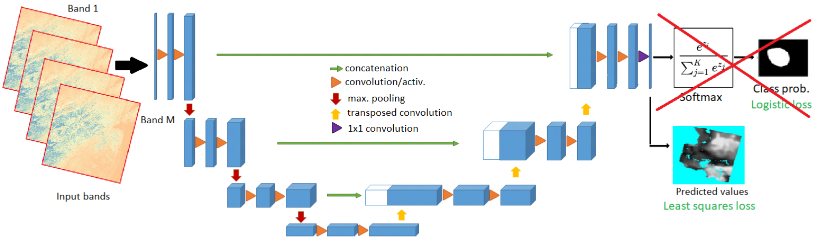


U-Net: a state-of-the-art Fully Convolutional Neural Networks

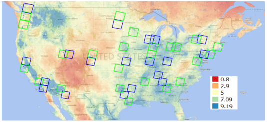


- Introduced by Ronneberger et al. (2015) in the context of classification
- To alleviate loss of detail, includes high-resolution original feature maps in upsampling

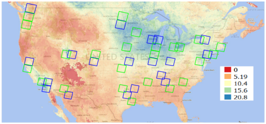
U-Net: a state-of-the-art Fully Convolutional Neural Networks



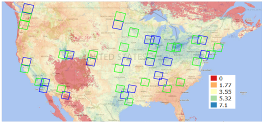
- Introduced by Ronneberger et al. (2015) in the context of classification
- To alleviate loss of detail, includes high-resolution original feature maps in upsampling
- We remove softmax layer and trained with least-squares loss directly on the 1x1 convolution output (final layer)



$PM_{2.5}$



NO_3



NH_4

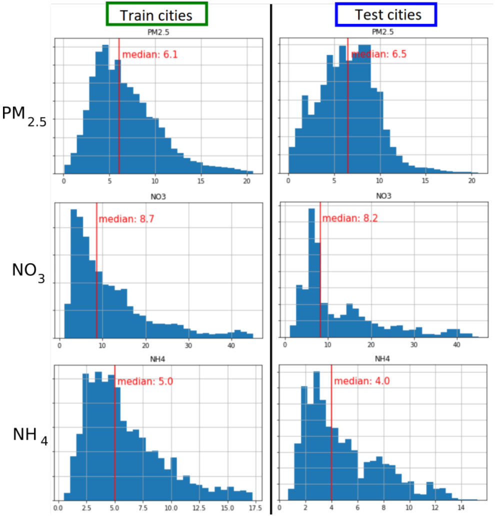
Satellite data:

- $N = 114$ Landsat images from $2013 - 2017$ →24 training cities
  - Additional $N_T = 24$ images from 20 test cities
- Locations of US cities chosen to reasonably represent the diverse US geography
- Rectangles indicate bounding boxes of the Landsat images' locations:
  green boxes = training and validation, blue boxes = testing

Pollution data:

- Pollutant concentration maps and regions defining training and testing data
- Background: average concentrations of $PM_{2.5}$, NO_3 , NH_4 in 2017 $[\mu g/m^3]$

Experiments – Setting

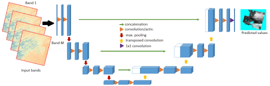


Data properties:

- Eliminated outliers beyond 1st and 99th quantile
- Training and testing cities show similar medians and distributions

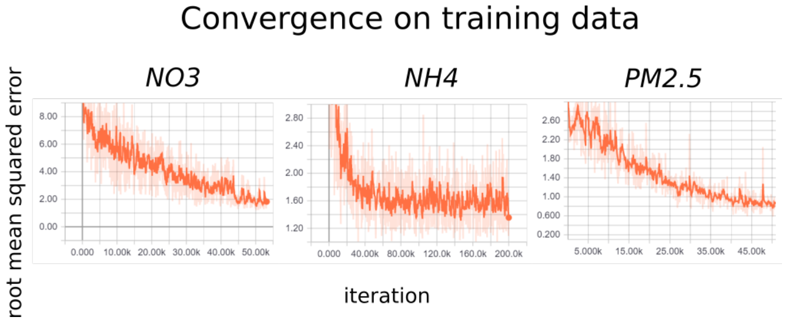
Experiments – Setting

U-net set-up:



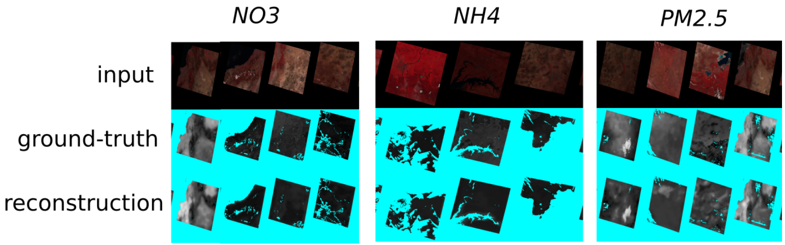
- Optimization: network trained separately for each pollutant – $PM_{2.5}$, NO_3 , NH_4 – by optimizing with ADAM over 500 epochs w. minibatch size of 15 and 100 internal iterations
- Parameterization: dropout ratio $= 0.5$, learning rate $= 0.00005$
- Structure: 3 layers, 32 feature maps at top level
- Input: 10/11 Landsat bands (high-resolution panchromatic dropped), 200x200 tiles
  - Save download time (4x size), no new information

Experiments — 1. Sanity check with ground-truth



- Trained on 93 of 114 images from training cities
- Shown: Convergence of U-net performance to low predictive error indicated by root mean squared error loss function (y-axis) on training set NO_3, NH_4, and $PM_{2.5}$ over iterations (x-axis)

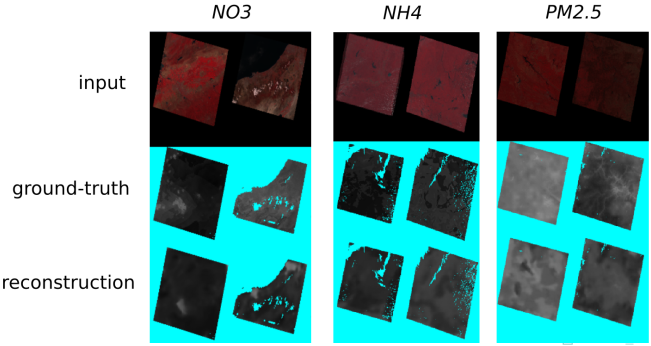
Experiments — 2. Generalize to temporally novel data



- Goal: Test on 23 previously unseen images from training cities
- Results: Mean absolute deviations: $NO_3 = 4.75$, $NH_4 = 2.24$, and $PM_{2.5} = 2.23\ \mu g/m^3$
- Shown: Ex. of pollutant concentrations predicted by the U-Net for new temporal data at locations used during training for all pollutants

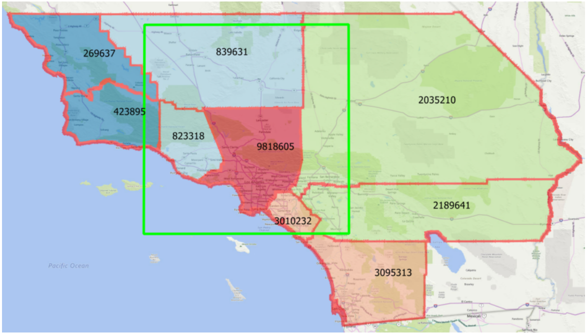
Experiments – 3. Cities with similar pollution profile

- Goal: Evaluate performance of predicting pollutant concentrations at cities not seen during training
- Results: Mean absolute deviations: $NO_3 = 2.46$, $NH_4 = 4.07$, $PM_{2.5} = 2.27\ \mu g/m^3$
- Shown: Ex. of predicted concentrations of 20 unseen cities

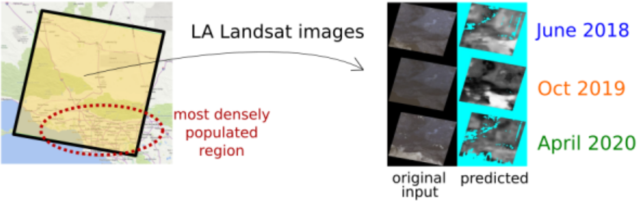


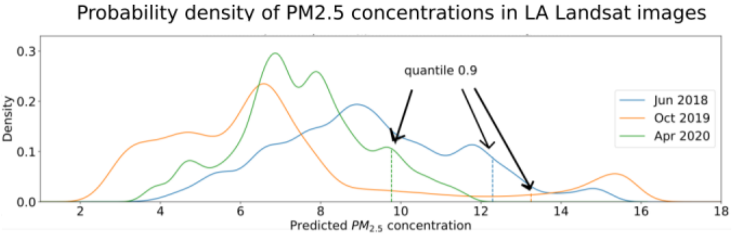
Experiments – 4. Predicting $PM_{2.5}$ for *COVID-19*

- Goal: Predict $PM_{2.5}$ concentration structure around Los Angeles before/during *COVID-19* lockdown and analyze shift in concentration distributions – years 2018-2020
- **No ground-truth** $PM_{2.5}$ concentrations for *time or city*



Experiments – 4. Predicting $PM_{2.5}$ for *COVID-19*





Shown: Landsat images w. predicted $PM_{2.5}$ concentration maps and density of predicted $PM_{2.5}$ concentrations for before/during *COVID-19* outbreak

→ Densely populated LA has high $PM_{2.5}$ concentrations – maps capture this

Summary

To **summer**-ize...

- U-net adapted to perform dense regression to learn three pollutants: NO_3 , NH_4 , and $PM_{2.5}$ from publicly available satellite imagery
- Shows generalization capabilities in both temporal and spatial domains on US cities and to data without *any* ground-truth
- Potential predictive benefits regarding *COVID-19* by inferring $PM_{2.5}$ concentrations, movement, and structure

Future/current:

- Multi-task learning - train a single network to predict multiple pollutants simultaneously
- Integrate ground measurement stations
- Large-scale experiments

Thanks!

Thanks for your attention! Questions?
Feedback or suggestions?