In the Danger Zone: U-Net Driven Quantile Regression can Predict High-risk SARS-CoV-2 Regions via Pollutant Particulate Matter and Satellite Imagery

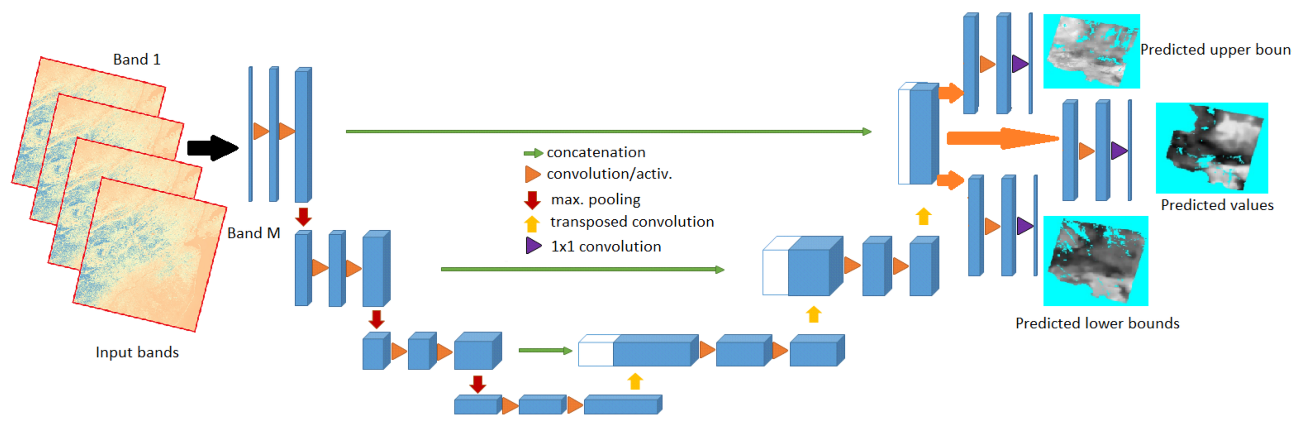
Jacquelyn A. Shelton[1], Przemyslaw Polewski[1,2], and Wei Yao[1]

[1]Hong Kong Polytechnic University [2]TomTom Location Technology Germany

Motivation and Highlight

- Since COVID-19 outbreak policy makers have been relying upon non-pharmacological interventions to control the outbreak
- Pollution linked to COVID-19:
  (1) lethality rate strongly correlated to particulate matter $\leq$ 2.5 microns in diameter, $PM_{2.5} \rightarrow 1\ \frac{\mu g}{m^3}$ increase in $PM_{2.5}$ ... 8% increase fatality rate [1]
  (2) pollution-to-human evidence that $PM_{2.5}$ a transmission vector [2]
  $\Rightarrow$ Pollution needs to be included in intervention strategies
- Goal: use U-net driven quantile regression to predict $PM_{2.5}$ air pollution concentrations based on easily obtainable satellite imagery
- Results: network can predict reasonable $PM_{2.5}$ values with their spatial distribution even for locations where pollution data unavailable
- Aim: $PM_{2.5}$ predictions aid planning efficacious COVID-19 strategies
  $\rightarrow$ e.g. understand impacts of social segregation policy on subpops – allocate medical funds to the most vulnerable populaces
  $\rightarrow$ control of COVID-19 hotspots – must pred. its spread and intensity
- Understanding an NPI's effectiveness requires understanding main transmission vectors $\rightarrow$ vital to learn how strength of air pollution affects transmission s.t. understand and design NPIs efficiently.

U-net model for pollutant particulate matter



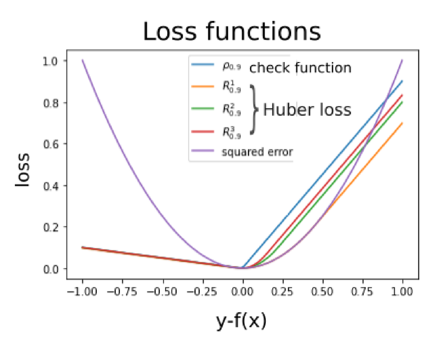
- Network structure: U-net [3] arch. with additional 2 parallel branches at top level of upsampling path for predicting confidence interval
- Loss function: weighted sum of 3 quantile losses corresponding to upper, lower bound and median prediction:

$$L_{aggr}(\theta) = \gamma_l L_{q_l}(\theta) + L_{0.5}(\theta) + \gamma_u L_{q_u}(\theta) \qquad (1)$$

- Quantile loss: each partial loss is of the form:

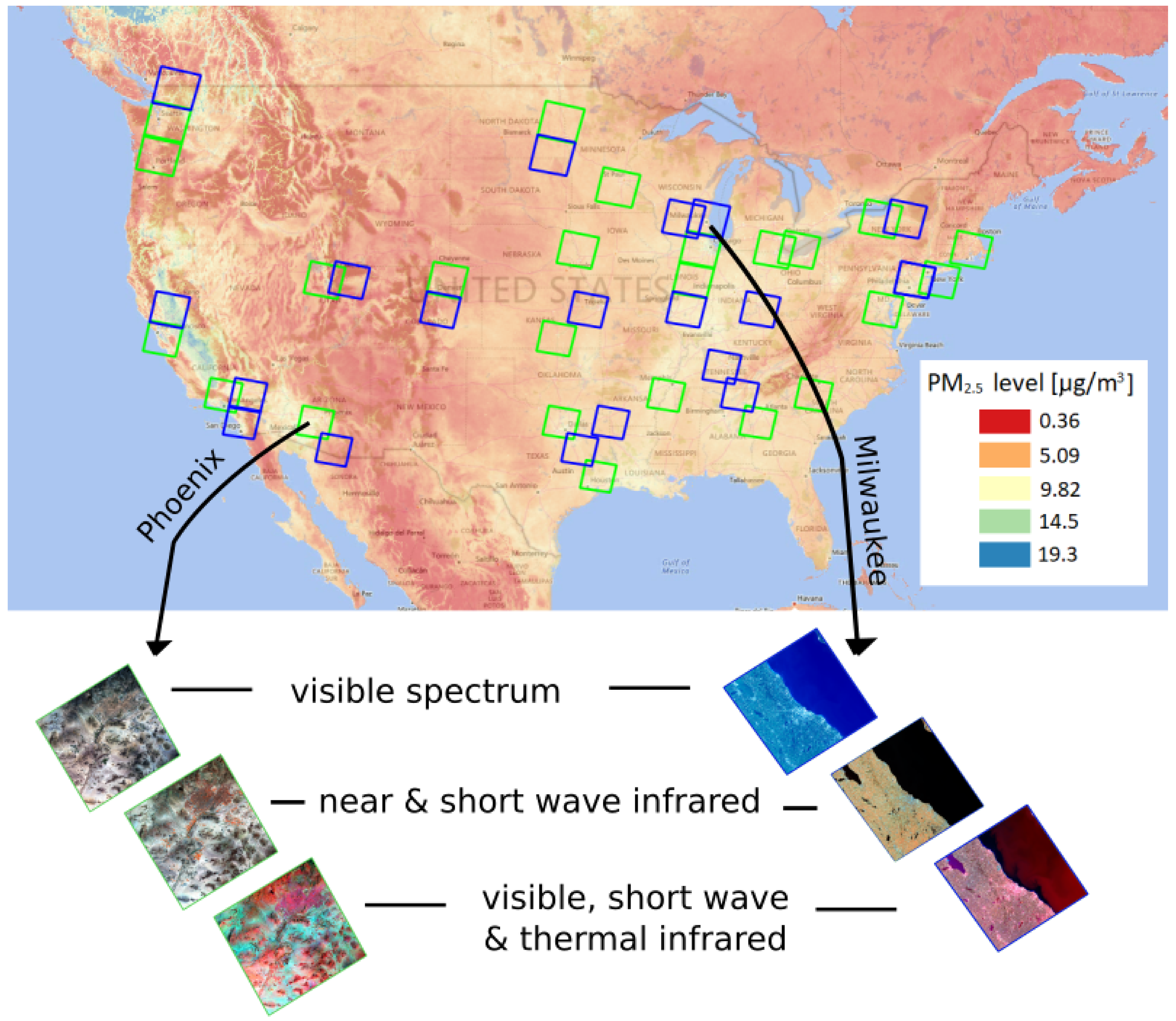
$$L_q(\theta) = \frac{1}{|S|} \sum_{(x,y)\in S} \rho_q(y - f(x|\theta))$$
$$\rho_q(r) = r[r \geq 0] - (1-q)r \qquad (2)$$

- Smooth approximation: ρ_q (Eq.2) is not differentiable everywhere; we approximate it with the (smooth) asymmetric Huber loss:



$$H(r|\delta_l, \delta_u) = r^2 - (r-\delta_l)_+^2 - (-r-\delta_u)_+^2$$
$$\rho_q \approx R_q^{\alpha}(r) = \alpha H(r|q/2\alpha, (1-q)/2\alpha) \qquad (3)$$

Data: Satellite imagery and pollutant concentrations



Satellite images of US cities for training and testing & $PM_{2.5}$ **concentrations** from 2018 shown in background color map

Satellite data:

- Landsat 8 multispectral imagery [4], publicly available from USGS
- Using 9 bands from violet (0.435μm) to thermal infrared (12.5μm)
- Revisit period 16 days $\rightarrow$ 2 images per month, mission start 2013
- Ground sampling distance 30-100 m depending on spectral band

Pollution data:

- *Pollutant Particulate Matter 2.5 ($PM_{2.5}$)* $\rightarrow$ only available 2001-2018 *monthly*, overlap with Landsat 8: 2013-2018
- Fusion of ground sensor, sat. imagery data and chem. model [5]
- Resolution of 0.01$\times$0.01 degrees, $\approx$ 1.1$\times$1 km (USA)

Preprocessing:

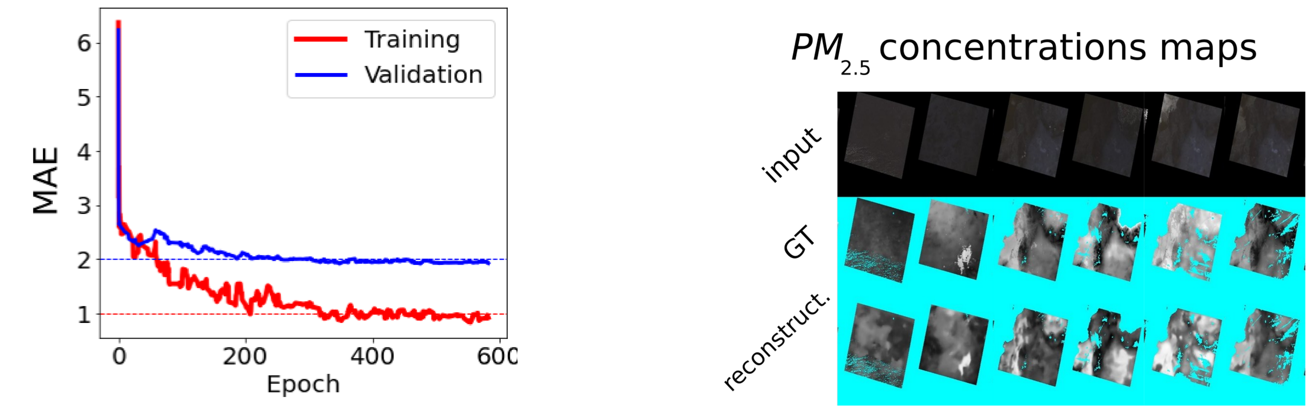
- Reproject and resample Landsat imagery to $PM_{2.5}$ resolution
- Mask out clouds to remove occlusions (band 9)
- Average sat images by month to follow GT temporal cycle

Experimental details

- Data: $N = 133$ sat imgs from Mar 2013–Apr 2018 of 24 US cities $\rightarrow$ 80:20% random split, training/testing set $\Rightarrow$ 106:27 images
- Network parameters: dropout ratio = 0.5, learning rate = 0.00005, regression quantiles $q_l = 0.1$, $q_r = 0.9$, Huber loss (3) aggregate contributions (2) set equal, function shape $\alpha = 2$

Experiments

Convergence verification and generalizability with ground-truth

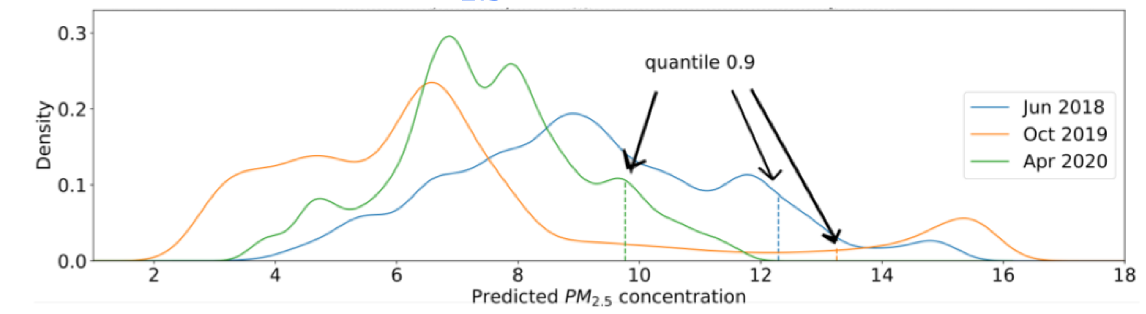


- Goal: assess generalizability of model to temporally unseen data
- Data: train = 106 & test = 26 imgs w. spatial overlap (same cities) but diff times
- Shown: U-net converges to GT $PM_{2.5}$ values w. min error & successfully reconstructs $PM_{2.5}$ concentrations and structure per pixel (dense)

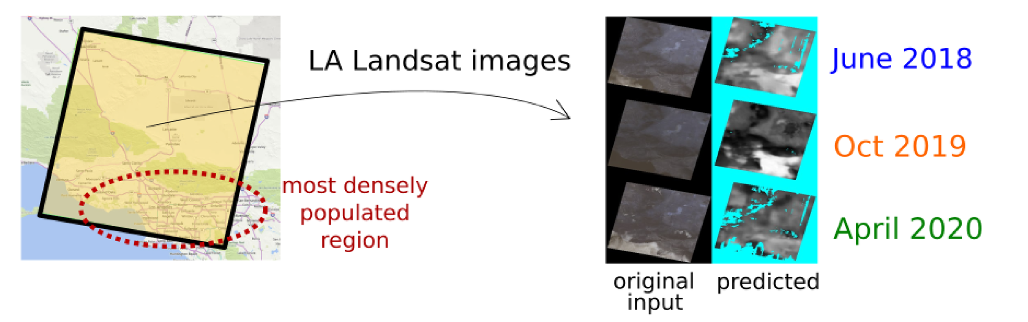
Before and after SARS-CoV-2-induced lock-down

Probability densities of $PM_{2.5}$ concentrations in LA Landsat images

A

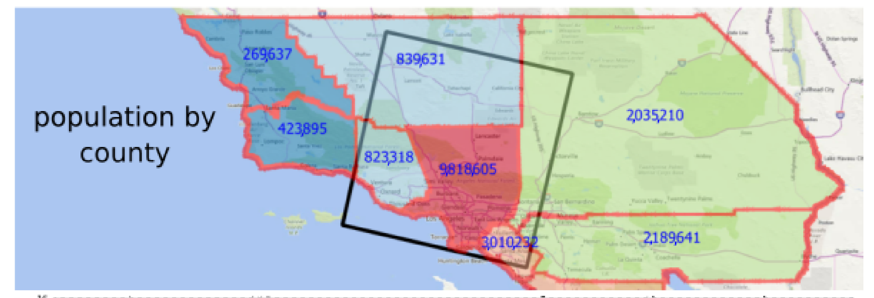


B

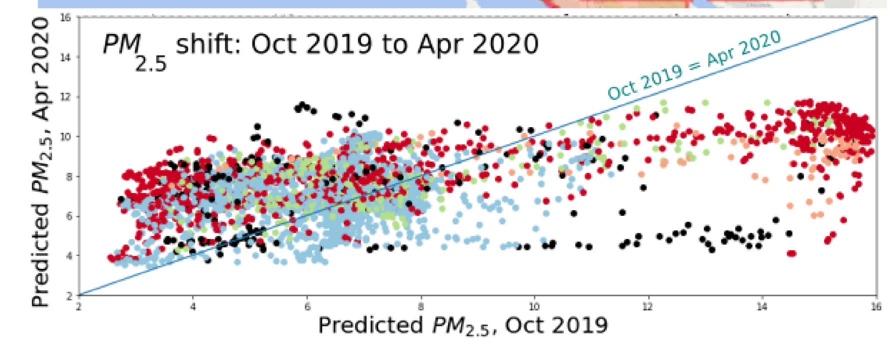


- Goal: predict $PM_{2.5}$ concentrations around LA lock-down w. no $PM_{2.5}$ GT
- Data: satellite images of Los Angeles from 2018, 2019, and 2020 *only*
- Shown: $PM_{2.5}$ concentrations distributions .9-th quantile shift from 13.2 to 9.7 $PM_{2.5}$ between Oct & Apr (**A**), matching $PM_{2.5}$ map visualizations (**B**) and drastic decrease where most densely populated

C



D



- Shown: $PM_{2.5}$ predictions from Oct 2019 vs. Apr 2020 (**D**) $\rightarrow$ greatest $PM_{2.5}$ drop in counties with largest population (red and blue points below $x = y$)
  $\Rightarrow$ make **meaningful $PM_{2.5}$ predictions**, both values and structure, where **no $PM_{2.5}$ data available** $\rightarrow$ can inform of *COVID-19 DANGER ZONES*

References

[1] 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.
[2] 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.
[3] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. CoRR.
[4] Landsat-Earth observation satellites. Technical report, Reston, VA, 2016.
[5] van Donkelaar, A., Martin, R. V., Li, C., and Burnett, R. T. (2019). Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. Environmental Science & Technology.