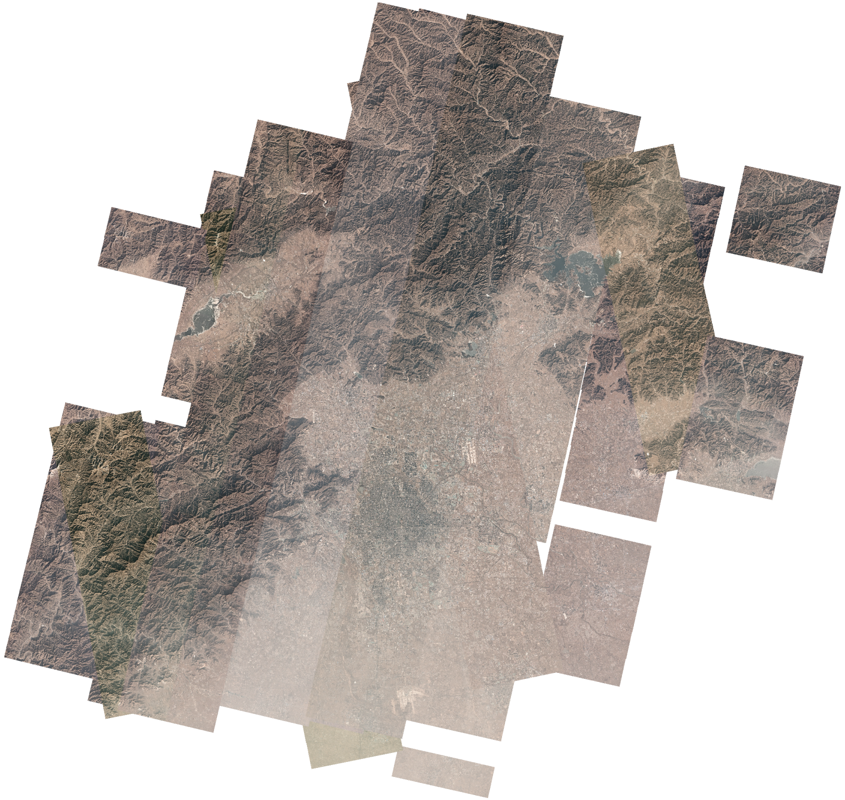
Estimating Uncertainty Intervals from Collaborating Deep Networks

David Carlson
Assistant Professor
Dept. of Civil & Environmental Engineering
Dept. of Biostatistics and Bioinformatics

Extracting Information from Remote Sensing

On the right is a map of part of Beijing stitched together from many satellite images.



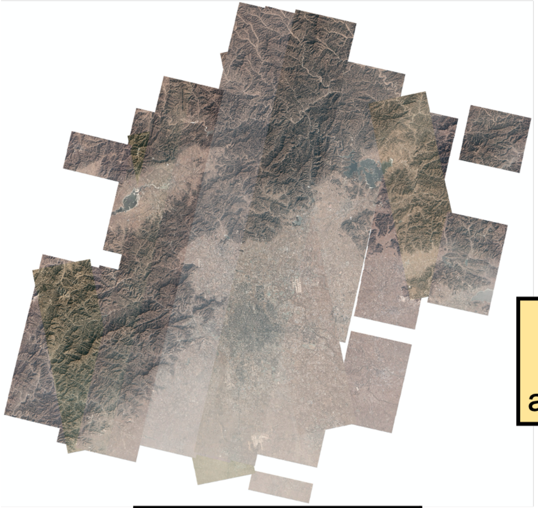
Can we predict details atmospheric properties?

If so, we could:

- Identify hot-spots for potential mitigation
- Link to epidemiological quantities (e.g., asthma, neuropsychiatric disorders, stress-pollutant interactions)

Predicting from 3m Microsatellite Data

Satellite imagery from Planet
(3m/pixel, daily resolution)

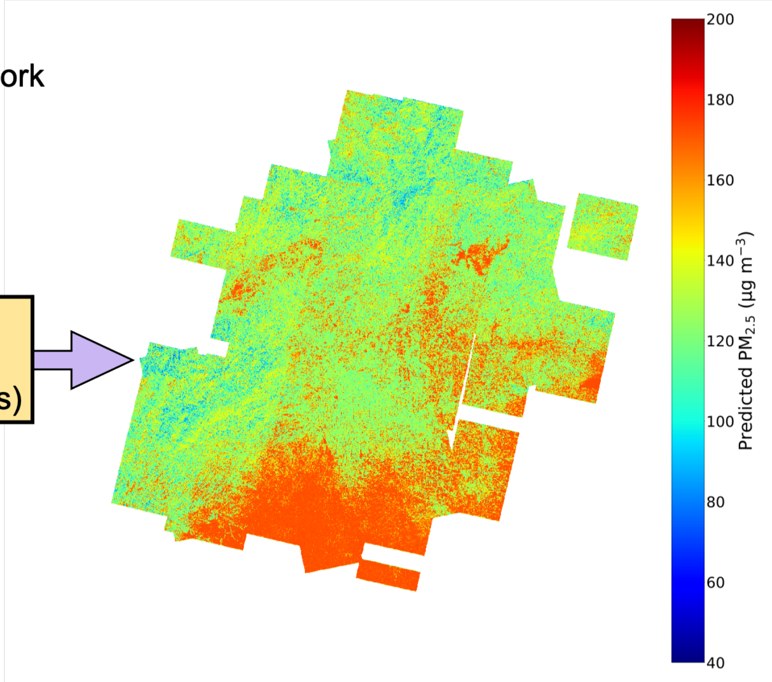


Convolutional neural network feature extractor

Random forest regressor (combines satellite image and meteorological features)

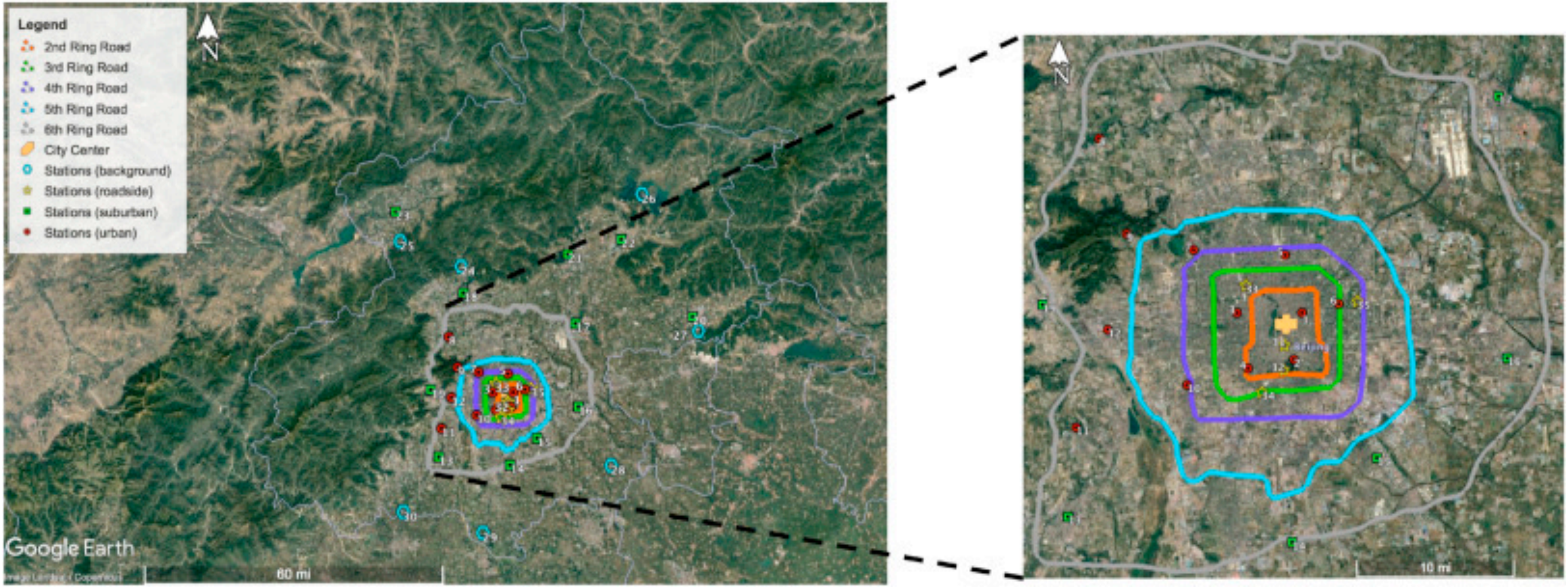
Temperature
Relative humidity
Wind speed
Sea level pressure

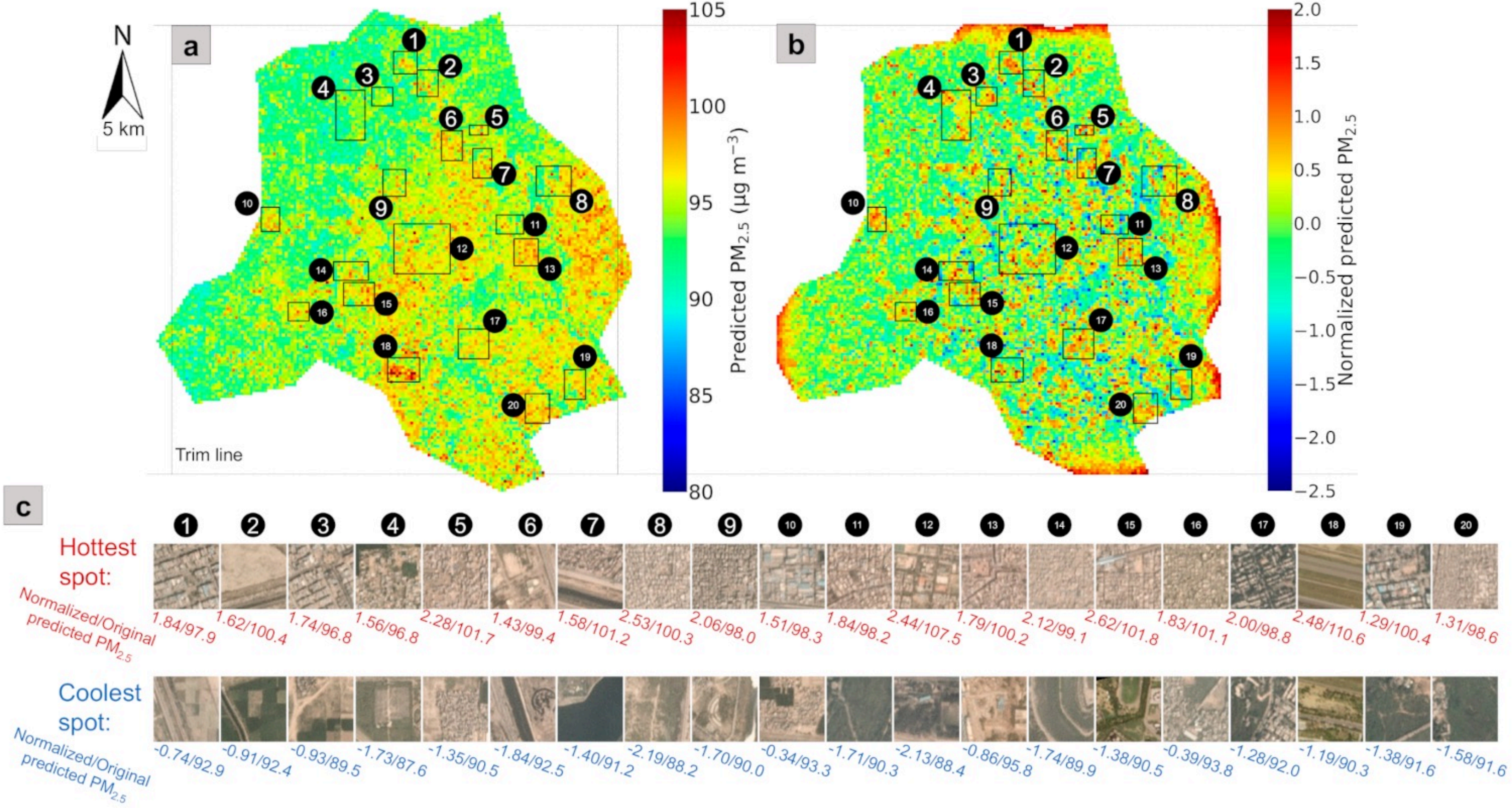
Map of estimated $PM_{2.5}$
(200m spatial resolution)



Zheng et al, Atmospheric Environment 2020

Can we find unusual local areas?





Zheng et al, Remote Sensing 2021

When are the data trustworthy?

- Understanding the uncertainty in predictions will maximize utility
- Deep learning methods do not provide uncertainties out-of-the-box
- Our proposal: make two networks work together to capture uncertainty

Approximating Quantile Regression

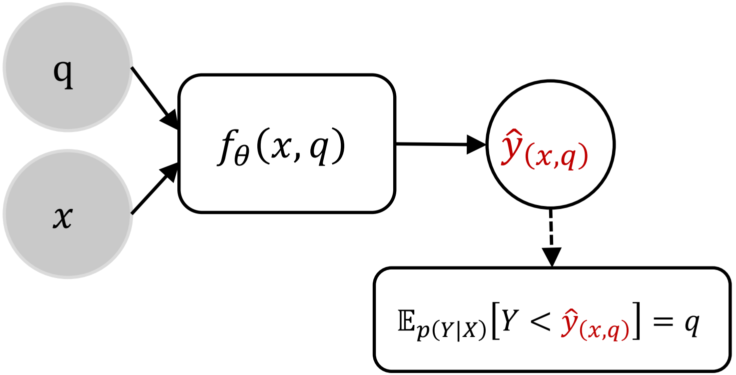
Suppose we have input feature $X \in \mathbb{R}^p$ with an associated outcome $\mathrm{Y} \in \mathbb{R}$ and define the q^{th} conditional quantile as $y_{q,x}$.

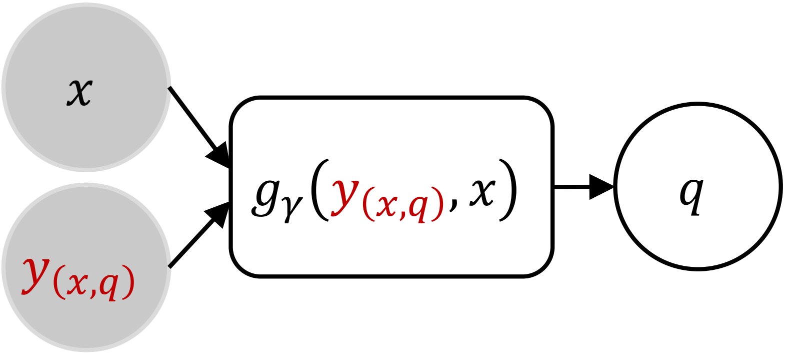
We can define the full distribution/uncertainty by the conditional CDF and inverse CDF, which we will approximate with neural networks:

$$g_\gamma\left(y_{q,X}, X\right) \simeq p\left(y_{q,x} < Y \middle| X\right) = q$$

$$f_\theta(X, q) \simeq y_{q,X}$$

Pairing Two Networks





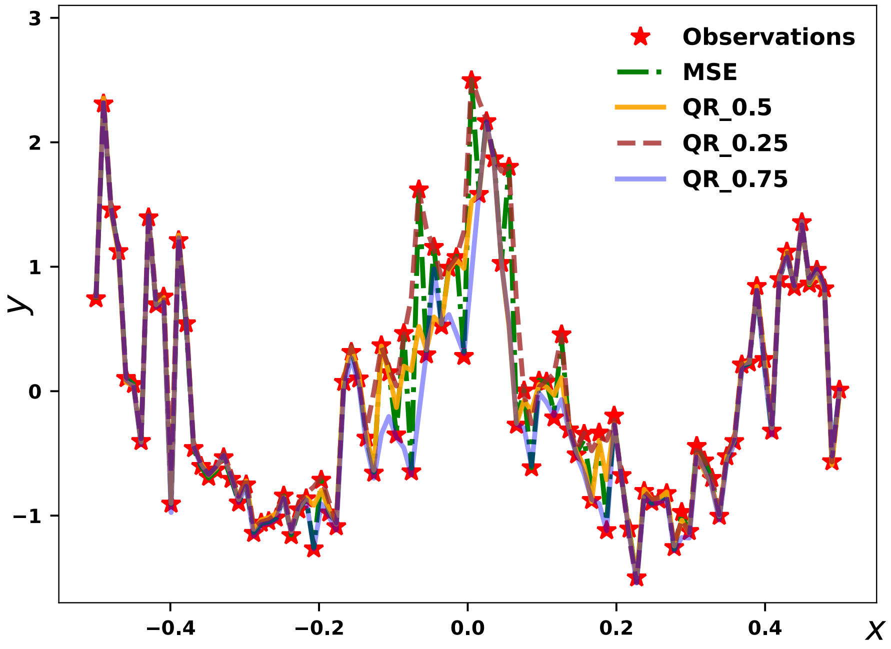
$$\text{f-loss}_\theta : \quad \mathbb{E}_{q \sim p(q), \boldsymbol{x} \sim p(X)} \left[\left(q - g_\gamma(f_\theta(q, \boldsymbol{x}), \boldsymbol{x}) \right)^2 \right]$$

$$\text{g-loss}_\gamma : \quad \mathbb{E}_{q \sim p(q), \boldsymbol{x}, y \sim p(X,Y)} \left[\ell(1_{(y < f_\theta(q, \boldsymbol{x}))}, g_\gamma(f_\theta(q, \boldsymbol{x}), \boldsymbol{x})) \right]$$

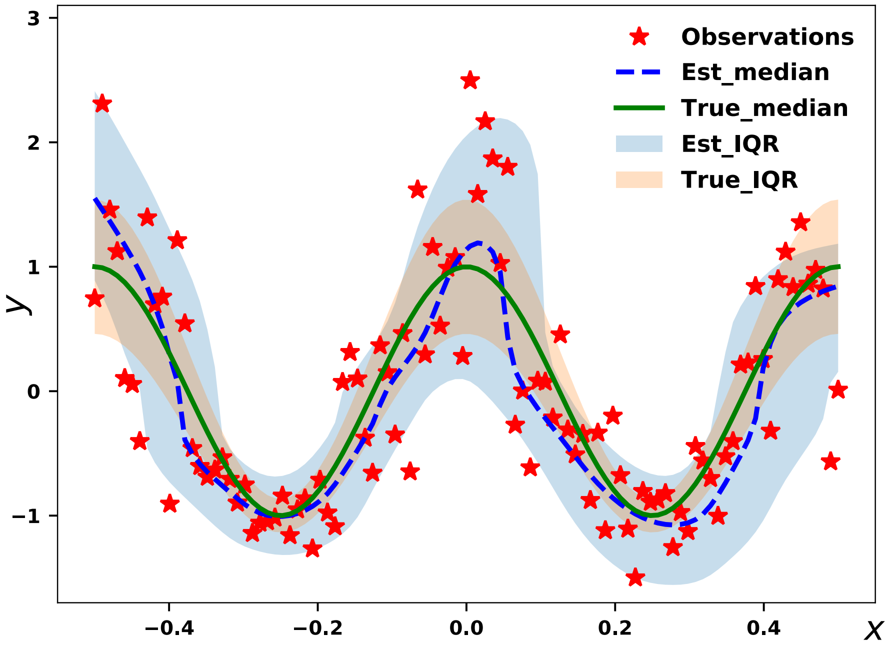
Overview of Theoretical Analysis

- Prop 1: If $f(\cdot)$ satisfies mild conditions, then a fixed point of $g(\cdot)$ is at the ideal solution.
- Theorem 2: Under a few assumptions, $g(\cdot)$ asymptotically captures the correct distribution.
- Prop 3: If $g(\cdot)$ is optimal, then a fixed point of $f(\cdot)$ is at the ideal solution.

Robust to overfitting



Quantile Regression with Deep Networks



Collaborating Networks

Real-World Experiments

- Evaluated on 6 real-world datasets of various sizes
  - Last dataset is on forecasting future A1c in a diabetic patient population from Duke Medical Records (18,335 patients)
- Compare on calibration metrics and fit metrics
  - Mean Absolute Error (MAE) and goodness-of-fit (discrete approximation of log-likelihood)

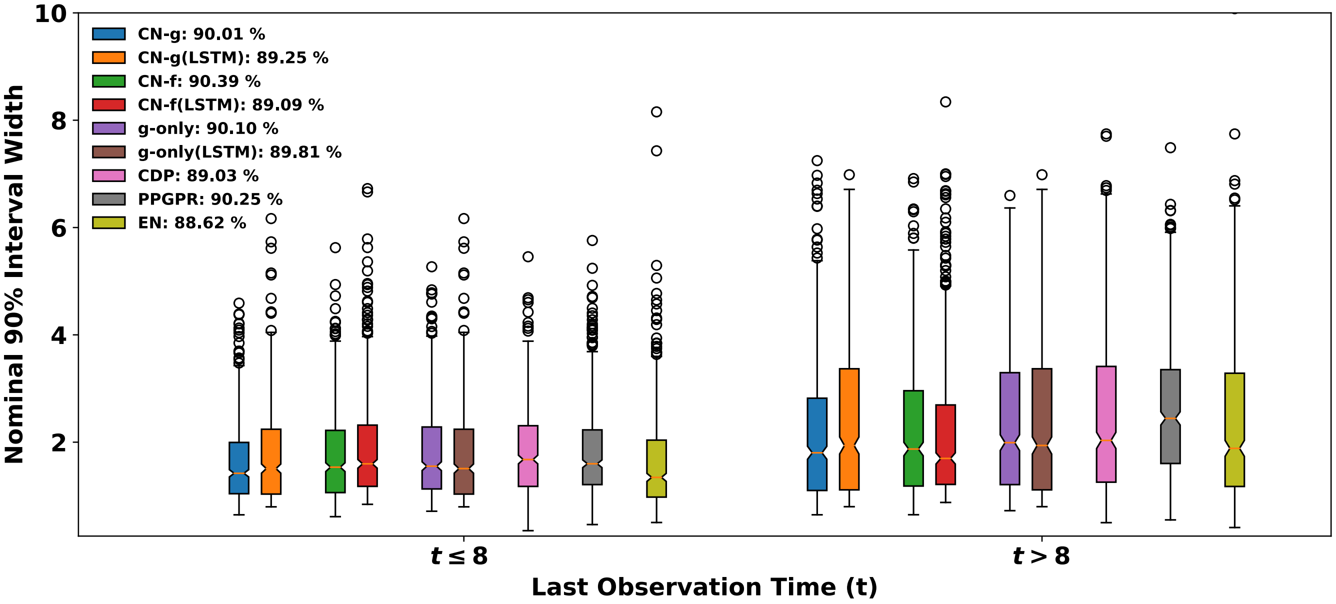
Calibration Metrics

Method/Data	CPU $\hat{cal}/\hat{90\%}$ (%)	Energy $\hat{cal}/\hat{90\%}$ (%)	MPG $\hat{cal}/\hat{90\%}$ (%)	Crime $\hat{cal}/\hat{90\%}$ (%)	Airline $\hat{cal}/\hat{90\%}$ (%)	EHR $\hat{cal}/\hat{90\%}$ (%)
CN-g	**4.62** ± **2.16** / 80.96 ± 3.64	**1.78** ± **0.63** / 88.80 ± 1.26	**3.16** ± **1.16** / 86.02 ± 2.14	2.70 ± 1.51 / 87.89 ± 1.72	**0.30 / 90.36**	**0.18 / 90.01** (0.25 / 89.25)
CN-f	7.78 ± 1.68 / 72.41 ± 3.79	2.42 ± 0.81 / 88.29 ± 1.71	6.03 ± 1.02 / 75.32 ± 2.58	2.86 ± 1.50 / 88.38 ± 1.72	0.47 / 90.89	0.47 / 90.39 (0.65 / 89.09)
g-only	**4.59** ± **2.01** / 88.67 ± 4.35	**2.00** ± **0.91** / **88.95** ± **1.74**	**3.31** ± **1.35** / 86.21 ± 2.38	2.92 ± 1.55 / 87.43 ± 2.26	**0.25** / 90.57	**0.15 / 90.10** (1.61 / 89.81)
DP	35.58 ± 1.26 / 99.64 ± 0.77	15.36 ± 0.57 / 97.55 ± 0.54	29.79 ± 0.67 / 99.93 ± 0.19	15.72 ± 0.42 / 96.32 ± 0.39	16.16 / 96.39	20.71 / 97.36
DP-CR	5.61 ± 1.81 / **89.63** ± **5.18**	2.45 ± 0.76 / 88.95 ± 2.37	3.82 ± 1.01 / 89.10 ± 3.15	**1.60** ± **0.74** / 91.21 ± 1.20	0.64 / 89.67	0.34 / 90.45
CDP	4.88 ± 1.99 / 92.53 ± 1.93	2.17 ± 0.68 / 86.44 ± 2.34	4.58 ± 1.37 / **89.94** ± **2.03**	10.65 ± 0.86 / 73.99 ± 1.18	5.03 / 91.46	3.64 / 89.03
GPR	6.82 ± 1.81 / 83.49 ± 4.73	3.53 ± 1.01 / **89.65** ± **1.56**	5.19 ± 1.27 / **90.26** ± **2.52**	7.84 ± 0.49 / **89.94** ± **0.99**	8.30 / 93.16	6.46 / 90.62
PPGPR	10.61 ± 3.21 / 74.58 ± 6.45	6.98 ± 1.17 / 77.29 ± 2.30	7.14 ± 1.93 / 77.46 ± 3.47	4.00 ± 0.87 / 83.33 ± 1.18	7.02 / 93.61	2.98 / 90.25
EN	6.17 ± 3.45 / 81.69 ± 6.66	6.58 ± 1.41 / 77.95 ± 1.97	3.64 ± 1.22 / 85.32 ± 2.90	9.08 ± 0.67 / 76.39 ± 2.04	7.89 / 93.93	1.74 / 88.62
CQR	4.81 ± 2.12 / **89.88** ± **3.24**	2.23 ± 0.94 / 91.01 ± 1.11	3.59 ± 1.29 / 91.47 ± 3.53	**1.78** ± **0.85 / 90.27** ± **1.52**	**0.38** / 90.04	-

Model Fitting Metrics

Method/Data	CPU MAE / $\hat{gof}$	Energy MAE / $\hat{gof}$	MPG MAE / $\hat{gof}$	Crime MAE / $\hat{gof}$	Airline MAE / $\hat{gof}$	EHR MAE / $\hat{gof}$
CN-g	0.169 ± 0.022 / -1.053 ± 0.182	**0.529 ± 0.013 / -1.796 ± 0.036**	0.256 ± 0.010 / -1.289 ± 0.091	**0.384 ± 0.015 / -1.379 ± 0.041**	**0.545 / -1.824**	**0.445 / -1.525** (0.463 / -1.554)
CN-f	**0.167 ± 0.017** / -1.626 ± 0.354	**0.529 ± 0.013** / -1.957 ± 0.132	0.257 ± 0.010 / -1.780 ± 0.022	**0.384 ± 0.015** / -1.459 ± 0.047	0.546 / **-1.829**	**0.446** / -1.566 (0.463 / -1.652)
g-only	**0.155 ± 0.021 / -1.031 ± 0.147**	0.531 ± 0.014 / **-1.796 ± 0.036**	0.262 ± 0.016 / **-1.288 ± 0.071**	0.387 ± 0.016 / **-1.383 ± 0.041**	0.547 / -1.830	0.453 / **-1.539** (0.453 / -1.517)
DP	0.167 ± 0.027 / -2.265 ± 0.135	0.553 ± 0.015 / -2.009 ± 0.032	0.259 ± 0.011 / -1.928 ± 0.043	0.443 ± 0.008 / -1.898 ± 0.040	0.565 / -2.207	0.464 / -1.969
DP-CR	0.167 ± 0.028 / -1.294 ± 0.098	0.553 ± 0.015 / -1.859 ± 0.022	0.259 ± 0.013 / -1.338 ± 0.097	0.443 ± 0.009 / -1.749 ± 0.045	**0.532** / -1.905	0.457 / -1.660
CDP	0.174 ± 0.030 / **-1.020 ± 0.088**	0.549 ± 0.018 / -1.887 ± 0.043	0.252 ± 0.011 / **-1.281 ± 0.081**	0.408 ± 0.009 / -2.017 ± 0.094	0.571 / -2.122	0.462 / -1.699
GPR	0.190 ± 0.043 / -1.310 ± 0.213	0.548 ± 0.016 / -1.850 ± 0.024	**0.250 ± 0.012** / -1.293 ± 0.066	0.403 ± 0.006 / -1.717 ± 0.038	0.606 / -2.152	0.506 / -1.797
PPGPR	0.197 ± 0.042 / -1.286 ± 0.234	0.569 ± 0.016 / -2.122 ± 0.063	**0.249 ± 0.013** / -1.394 ± 0.113	0.400 ± 0.009 / -1.719 ± 0.059	0.588 / -2.100	0.472 / -1.663
EN	0.191 ± 0.039 / -1.178 ± 0.181	0.567 ± 0.014 / -2.105 ± 0.076	0.263 ± 0.017 / -1.412 ± 0.207	0.430 ± 0.010 / -1.932 ± 0.082	0.564 / -2.049	0.456 / -1.644
CQR	0.203 ± 0.050 / -	0.552 ± 0.017 / -	0.276 ± 0.018 / -	0.431 ± 0.020 / -	0.562 / -	-

Forecasting Uncertainty



Conclusions/Comments

- Collaborating Networks are a theory-backed approach to quantile regression
- Can be integrated into nearly any deep learning framework
- Moving towards multi-modal data integration with remote sensing and sensor networks

Duke UNIVERSITY Funding Acknowledgement: NIH R01EB026937