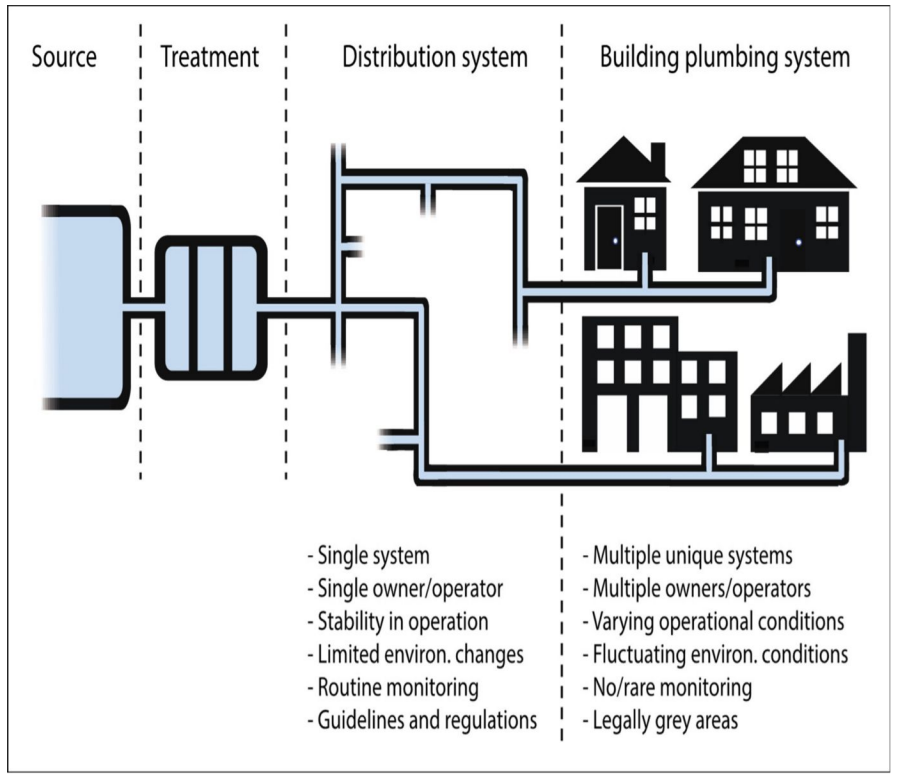
Using Data Science to Protect Residential Water Quality

R. Nerenberg, M. Sisk, M.D. Lemmon, E. Clements, Y. Duan



Motivation and Need

- Drinking water quality can degrade within homes
- Toxic metals and opportunistic pathogens
- Quality decreases with increasing stagnation time (water age)
- Often a greater problem in low-income communities
- Better information can help residents reduce risks
- Purging devices can limit water age can improve water quality

Problem Statement

- Use data science methodologies identify homes at risk for water quality problems based on easily available data on homes and water use.
- Develop community wide strategies for mitigating these problems in a fair and equitable manner.

Using Data Science to Protect Residential Water Quality

R. Nerenberg, M. Sisk, M.D. Lemmon, E. Clements, Y. Duan

Study1:Training Models for Residential Water Age.

Training local models for predicting the water age based on the home profile and water usage is a machine learning problem that contain three components:

- System: The system consists of two parts: a generator that generates features including the home's profile (square footage, number of bedrooms, number of bathrooms), occupant profile (gender, age, work class, number of occupants), and water usage pattern (water pipe distribution, water demand), and an observer that provides the water age.
- Model: The model is a deep neural network that predicts what the observer should see for the given input features.
- Loss function: Mean Square Error loss.

Study2: Mitigating Risk in a fair and equitable manner.

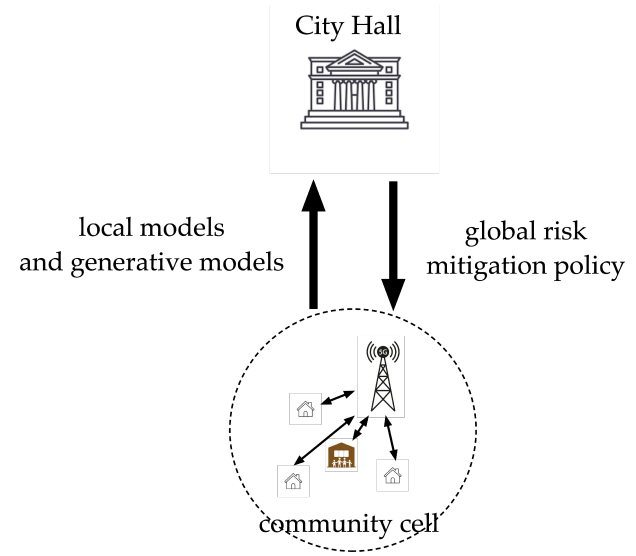
We use a fair Federated Learning framework to develop the fair mitigation policy.
The **Fair Federated Learning** framework has two parts:

- The edge devices that learn:

1. Local classifier, $\boldsymbol{\eta}_k$, that predicts water age based on residence profile.
2. Local generative neural network (WGAN), G_k , for the community's data distribution

- The community cloud server uses the data, generated by the WGAN's to train a model, $\boldsymbol{\eta}$, that minimizes MSE of local models and the statistical disparity (risk difference) between the socially advantaged community (SAC) and socially disadvantaged community (SDC)

$$\frac{1}{N}\sum_{k=1}^{N}(\eta_k(\hat{x}) - \eta(\hat{x}))^2 + \alpha\,|P(\eta(\hat{x}) = 1\,|\,\mathrm{SAC}) - P(\eta(\hat{x}) = 1\,|\,\mathrm{SDC})$$



Project Update

R. Nerenberg, M. Sisk, M.D. Lemmon, E. Clements, Y. Duan

Study1:Training Models for Residential Water Age.

We have already finished generating home profile and occupation profile. Our future work includes:

- Preparing the data of water demand and water age. (E.Clements)
- Training the deep neural network for predicting water age. (Y.Duan)
- Validating the model we trained using up to three homes in South Bend, owned by the PIs or colleagues. (R. Nerenberg, E.Clement)

Study 2: Mitigating Risk in a fair and equitable manner.

We have applied the fair federated framework to the UCI Adult database. The preliminary results demonstrate the effectiveness in terms of both utility and fairness.

- The UCI Adult Dataset has eight categorical features (work class, education, etc.) and six numerical features, and is used to predict the binary label of income (high, low).
- The dataset is split into a socially advantaged group (male) and a socially disadvantaged group (female).
- The results show that the global model trained under our framework can achieve fairness (risk difference = 0.01) while maintaining relatively good model utility (accuracy = 81%).

The future work includes:

- Clustering the data into N groups, with each group representing the dataset of an individual community. (Y.Duan, M.D.Lemmon)
- Develop a local level model for the community through transfer learning, wherein the local communities adjust only the last layer of the model previously trained in study 1. (Y.Duan)
- Developing the fair global risk mitigation policy based on all local models and the global level data distribution. (Y.Duan)
- Examining the performance of the global model from the perspectives of fairness and utility. (Y.Duan)