Machine learning regression for calibration and prediction of low-cost stream gauges

Jeil Oh, Yeji Kim, and Dr. Matt Bartos

The University of Texas at Austin
Civil, Architectural and Environmental Engineering
Cockrell School of Engineering

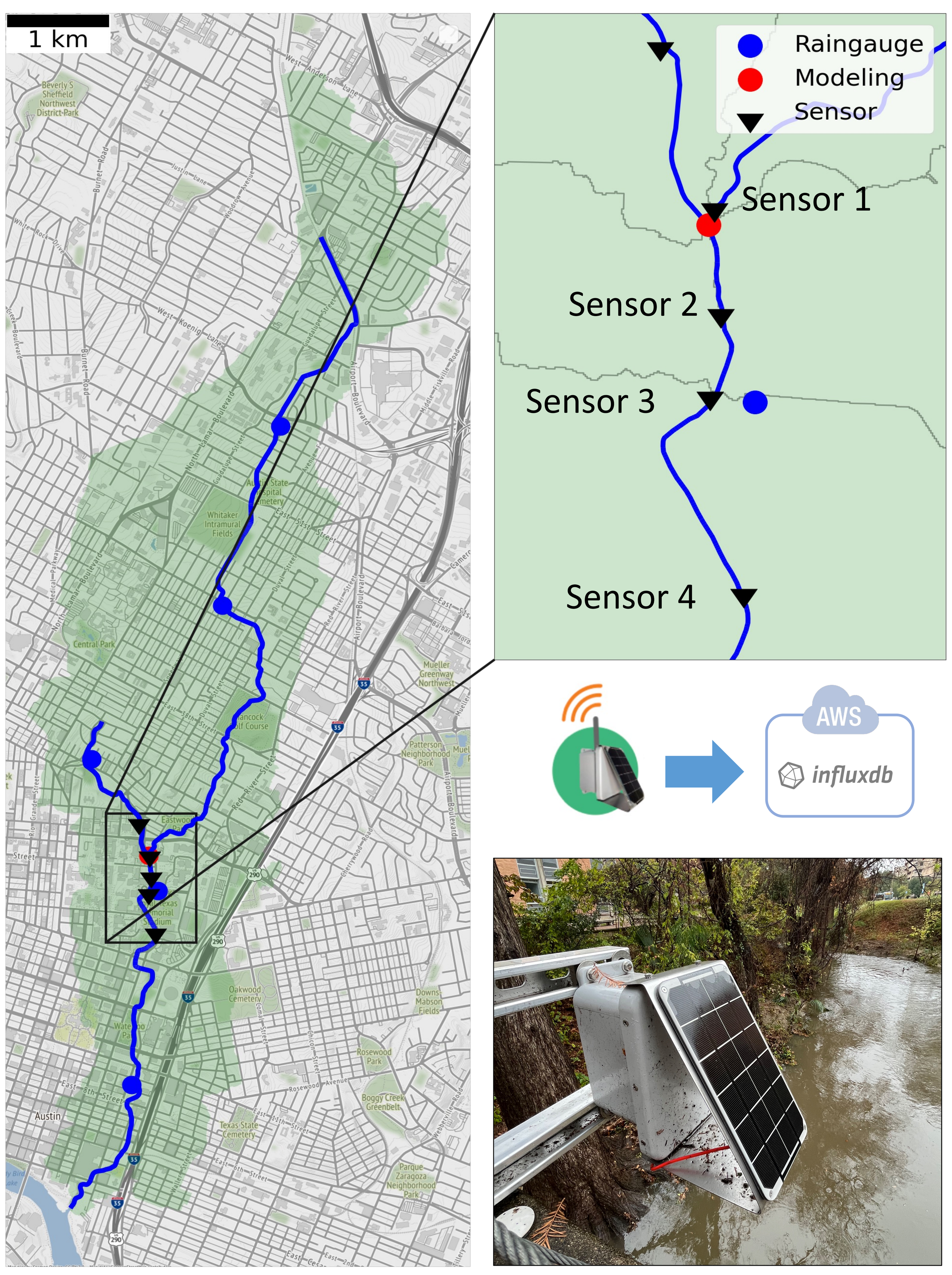
Introduction

Advances in the Internet of Things have led to the development of low-cost sensing strategies that enable denser streamflow monitoring networks, improving the management of stormwater hazards such as floods and combined sewer overflows. However, low-cost sensors are often sensitive to environmental conditions and may have faults such as spikes and erratic drifts, limiting their effectiveness for monitoring purposes. To address this issue, we investigated different calibration and prediction models for faulty sensors by utilizing machine learning (ML) regressors.

Methods

- **Wireless sensor network:**

The wireless sensor network is deployed in the Waller Creek watershed in Austin, TX. Four ultrasonic depth sensors, positioned in close proximity within a flood-prone confluence zone during periods of heavy rainfall, are utilized to rectify measurement inaccuracies and recalibrate faulty readings.

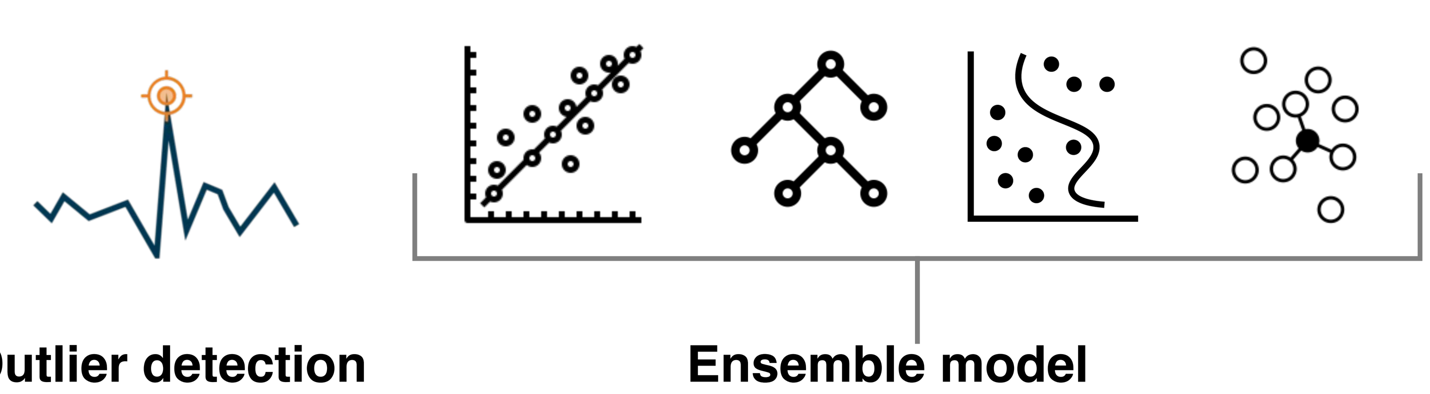


- **ML Regression Framework:**

First, we apply Random Sample Consensus (RANSAC), an iterative approach that enables outlier detection by estimating a mathematical model from a dataset, notwithstanding the presence of outliers.

Following the exclusion of outliers, we implement multiple ML models including Multiple Linear Regression, K-Nearest Neighbor (KNN) Regression, Random Forest (RF), and Support Vector Regression (SVR) to learn the dynamics from proximate sensors.

Lastly, we create an ensemble of estimates from ML models to build a calibration model, ensuring reliable sensor estimates even in the case of sensor malfunction.

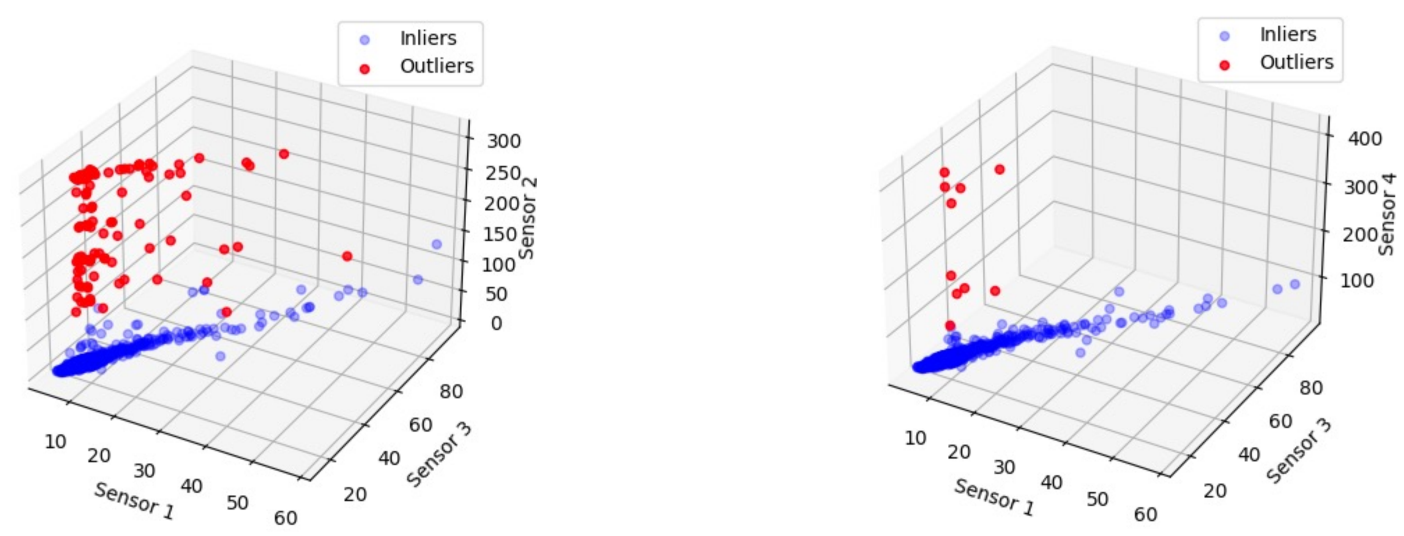


Conclusion and Future work

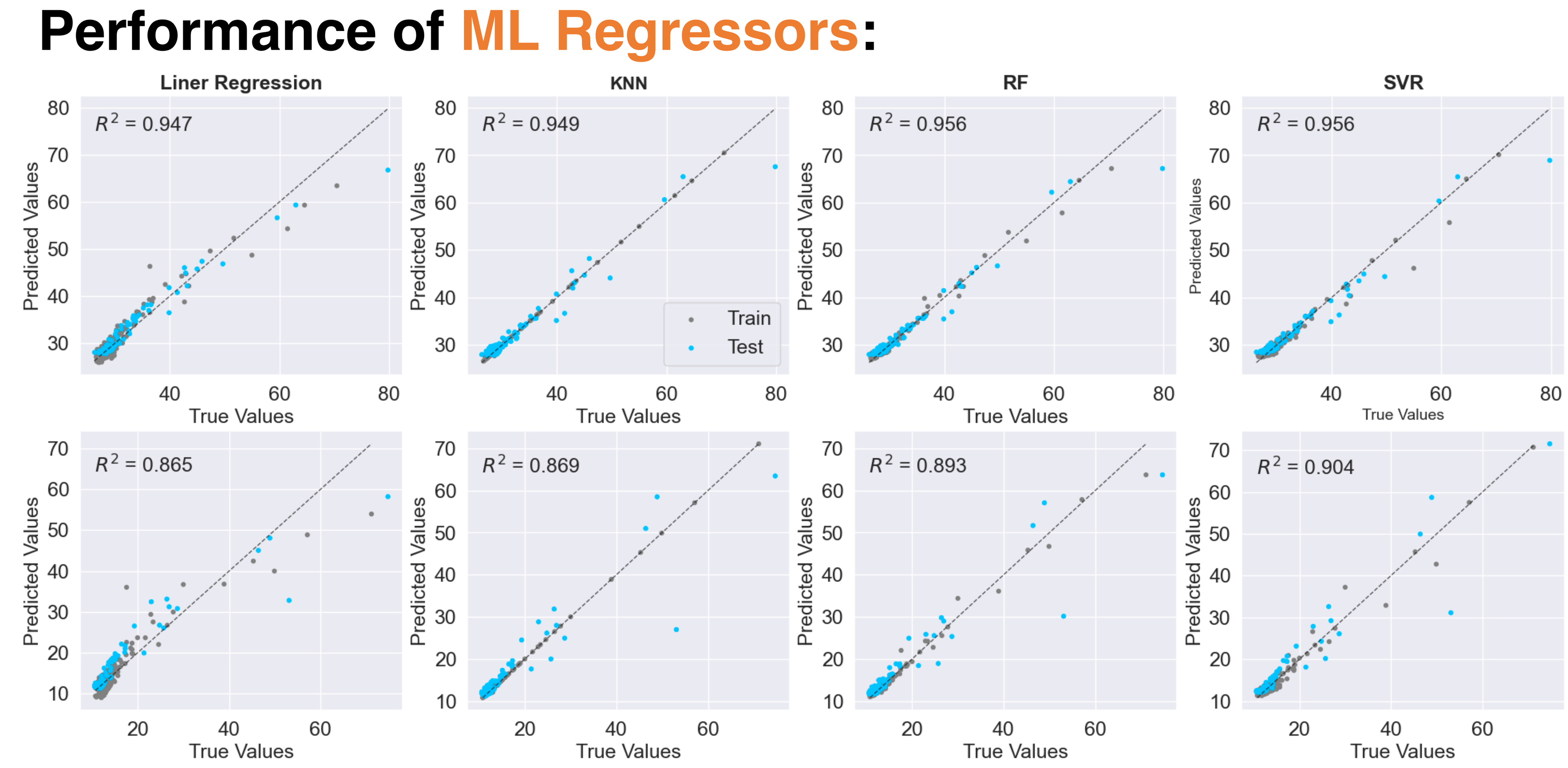
Combining ML models with distributed wireless sensors provides a promising approach to make reliable predictions from low-cost depth sensors with reduced management and ultimately to build densely distributed water monitoring systems.

Results

- **Outlier detection by RANSAC Regressor:**



- **Performance of ML Regressors:**



- **Stream gauge reconstruction by ML regressor ensembles:**

ML regressors provide not only a calibration model for a sensor network but also reliable predictions, especially under faulty sensor measurements.