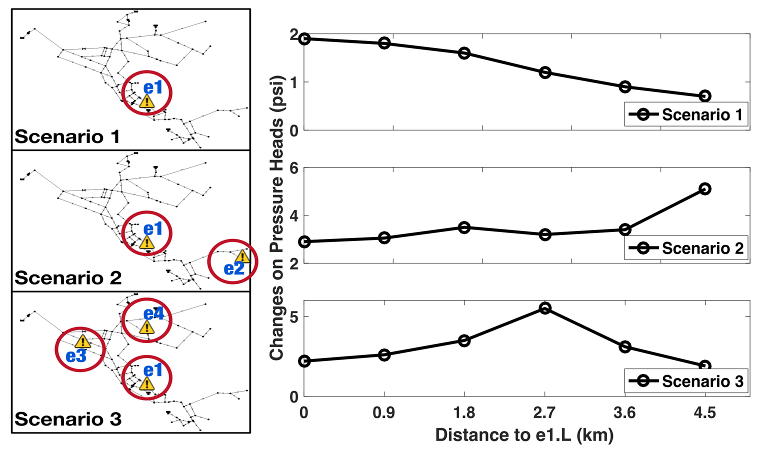
AquaSCALE: Exploring Resilience of Community Water Systems

Nalini Venkatasubramanian(PI), Sharad Mehrotra(Co-PI), Qing Han, Phu Nguyen, Kuo-Lin Hsu, Mehdi Sadri, Soroosh Sorooshian, UC Irvine
Ronald T. Eguchi, Georgiana Esquivias, ImageCAT Inc.; Daniel Hoffman, Montgomery County, MD

"A framework to improve the outcomes of water related emergencies"

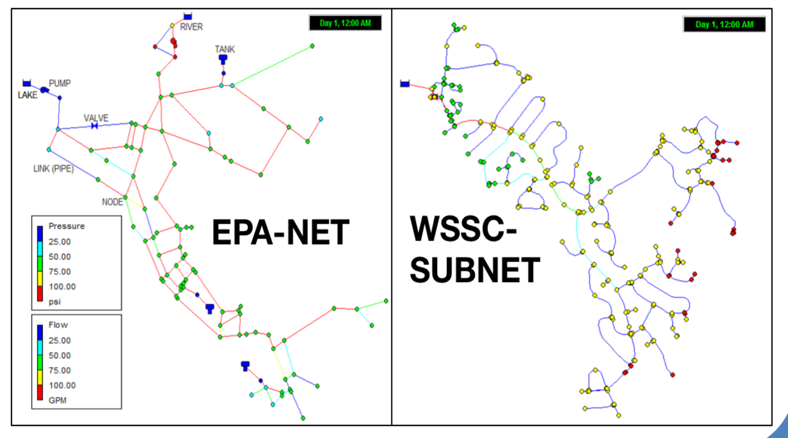
Key Resilience Issue – Pipe Leaks

- Pipe break/leak is one of the most frequent types of failures in water networks worldwide.
- Recent reports from agencies (Los Angeles Dept. of Water/Power, Washington Suburban) indicate that communities are experiencing an unusual increase in pipe breaks and leaks.
- Extreme weather and heavy rainfall (e.g. Hurricane Sandy, El Nino 2016) can stress already weakened pipes to the point of causing major breaks.



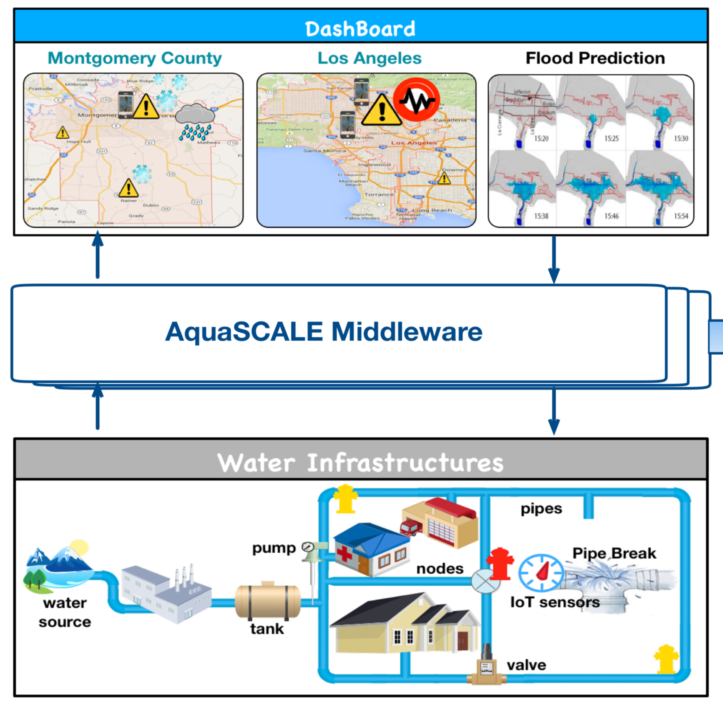
- Compared with the single leak detection, multiple pipe failures is much more complex to locate.
- Multiple events interact with each other resulting in a set of highly correlated observations.
- Difficult to follow a changing pattern.

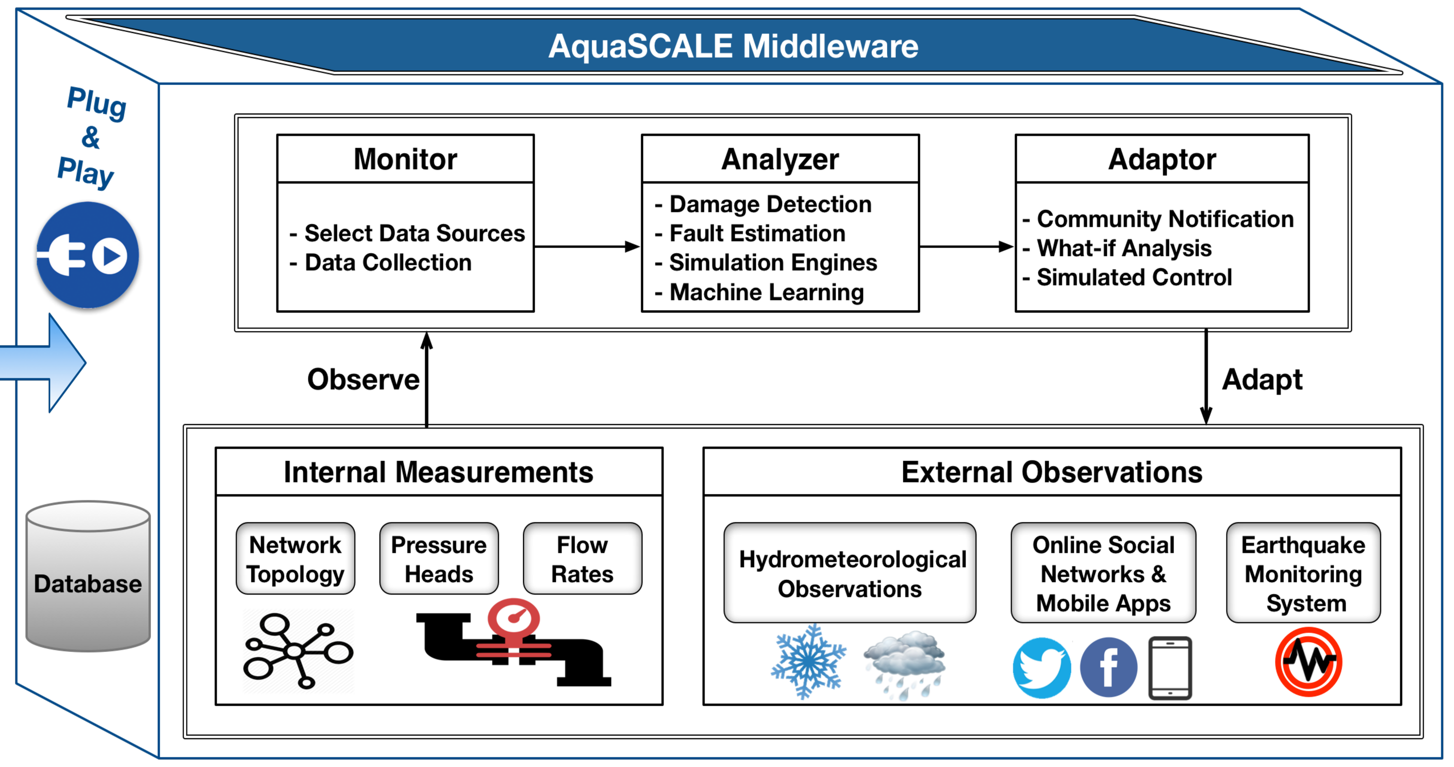
- AquaSCALE is evaluated using two networks.
- EPA-NET: a canonical network from EPANET. (96 nodes and 118 pipes)
- WSSC-SUBNET: a real subzone of the WSSC water service area. (299 nodes and 316 pipes)



Goals and Overview

- Water is a precious resource and lifeline service to community worldwide.
- The critical infrastructures have been developed over decades (centuries sometimes) and become complex and vulnerable to failures.
- Pipe breakage is one of the most frequent types of failures that often causes community disruptions.
- Develop methodologies to understand operational performance and resilience issues for real-world community water infrastructures.
- Explore solutions to problems in cyberspace before instantiating them into a physical infrastructure.
- Prevent water service failures by identifying operational degradation in aging infrastructures.
- Improve speed and accuracy of damage estimation in natural disasters and human-made hazards.
- Reduce service restoration time in the event of large hazards.

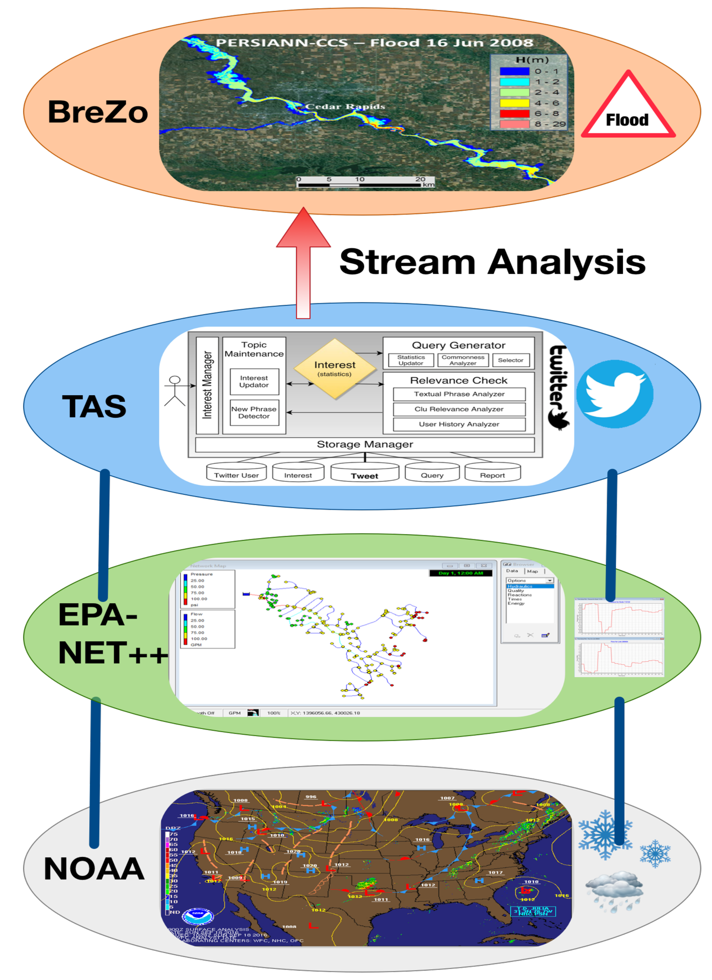




Prototype Implementation

Integration of Multiple Resources

- Leverages dynamic data from multiple information sources including IoT sensing data, geophysical data, human inputs, and simulation engines.
- Create a sensor-simulation-data integration platform.
- Pipe failures often lead to changes in pressure heads and flow rates.
- The chance of water main breaks rises significantly in the event of extremely cold weather.
- Human reports on leak events can identify the potential damaged region.
- A large-scale pipeline failures or a pipe burst may cause severe flooding.



Cascading Events – Incorporate flood modeling and prediction to capture the impacts of pipe failures.

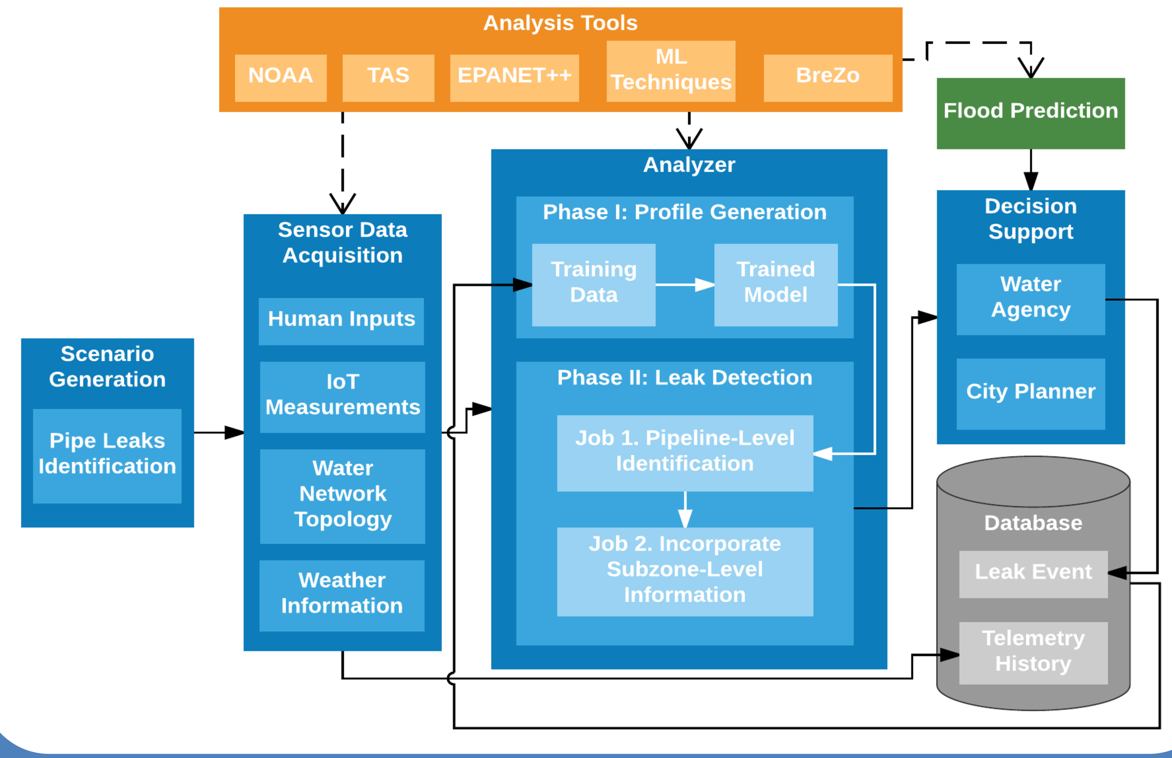
Human Inputs – Integrate a novel Tweet Acquisition System to extract leak-related tweets in an online adaptive manner.

IoT Measurements – Using a commercial grade hydraulic simulator EPANET enhanced with the support for IoT sensors and failure modeling.

Geophysical Data – Integrate NOAA to supply information pertaining to the state of the atmosphere including weather warning and forecasts.

Leaks Identification Workflow

- The initial implementation of AquaSCALE is designed as a workflow based system comprised of multiple components.
- Scenario Generation Module enables providing meaningful and diverse water contexts to the framework by generating a range of situations.
- Sensor Data Acquisition Module enables gathering of real-time field information and projecting the effects on the simulation outcomes.
- Analytics Module is used to plug and unplug specific information at will depending on the specific context of applications.
- Decision Support Module enables users/operators/analysts to interact.



A Composite Leak Identification Algorithm

- A two-phase approach enables an accurate and timely leak events identification.
- Phase I – the profile model is generated offline. Phase II – additional observations are integrated with the predicted results from the profile model when the live data coming in.
- Proposed a hybrid approach HybridRSL, a combination of RF and SVM via LR.
- Remain robust with decreasing number of IoT devices

Algorithm 1 Training Profile Models

Input water network topology T, IoT measurements X, true leak events Y_v and classifiers f_v for $v \in \mathcal{V}$
Output the profile model $f = \{f_v : v \in \mathcal{V}\}$
Objective update f_v to best fit training samples

for v in V **do**
  f_v.fit(T, X, Y_v)
end for

Algorithm 2 Inferring Leak Events

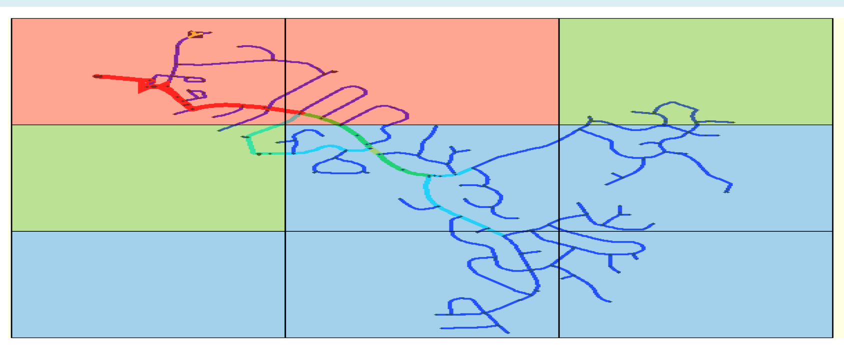
Input water network topology T, IoT measurements $\mathbf{x}$, profile model f, leak probability due to frozen $p_{v\in\mathcal{V}}(\text{leak}|\text{freeze})$ and human inputs $\mathcal{C}$
Output an updated set of leak locations $\mathcal{S}$
Objective minimize $E[\mathbf{y}]$ in (7)

```
/* Pipeline-level Identification */
```
$P = f$.predict_proba$(T, \mathbf{x})$; $\mathcal{S} = f$.predict$(T, \mathbf{x})$
for v in $\mathcal{V}$ **do**
  if v is detected to be frozen **then**
    $p_v(1) = \min\{1, p_v(1) + p_v(\text{leak}|\text{freeze})\}$
    $p_v(0) = 1 - p_v(1)$
    $\mathcal{S} = \mathcal{S} \cup \{v\}$ **if** $p_v(1) > p_v(0)$
  end if
end for

```
/* Incorporate Subzone-level Information */
```
$\mathcal{C} = \{c : c = \{v : |l_c - l_v| < \gamma \wedge v \in \mathcal{V}\}\}$; $\Phi_{c\in\mathcal{C}} = \text{Inf}$
for c in $\mathcal{C}$ **do**
  if $\exists v \in \mathcal{S}$ for $\forall v \in c$ **then**
    $\Phi_c = 0$, break
  end if
  if $\Phi_c \neq 0$ **then**
    $v^* = \arg\max_{v\in c} H(y_v)$
    if $H(y_{v^*}) > \Gamma$ **then**
      $p_{v^*}(1) = 1$, $p_{v^*}(0) = 0$, $\mathcal{S} = \mathcal{S} \cup \{v^*\}$
    end if
  end if
end for

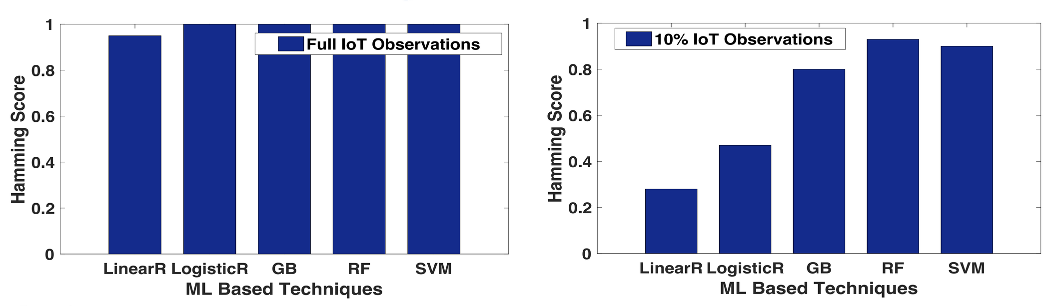
- In Phase I, the objective is to train a set of classifiers to generate a robust profile model.
- Use an extensive amount of measurements collected in water infrastructures.

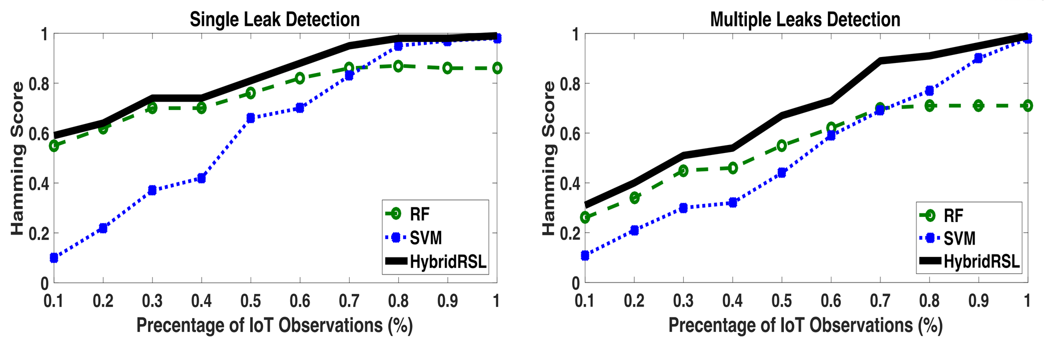
- In Phase II, multiple data sources are sequentially aggregated to infer leak locations from combinations of information.
- IoT measurements and ambient temperatures are relatively stable data sources compared with human inputs.
- Use IoT and temperature streams for pipeline-level predictions.
- Use additional human inputs for incorporating subzone-level information.



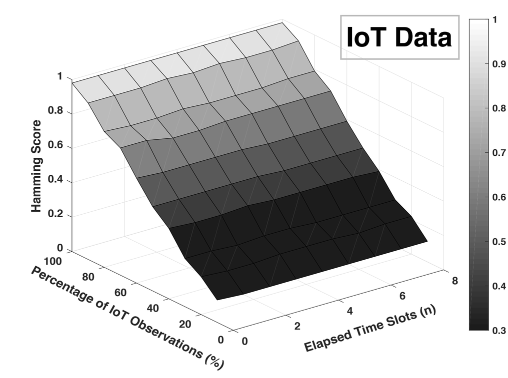
Experimental Study

- Comparison of multiple Machine Learning (ML) techniques for single leak event identifications using full and 10% IoT observations on EPA-NET water network.
- SVM and RF can keep a better performance with 10% IoT observations.

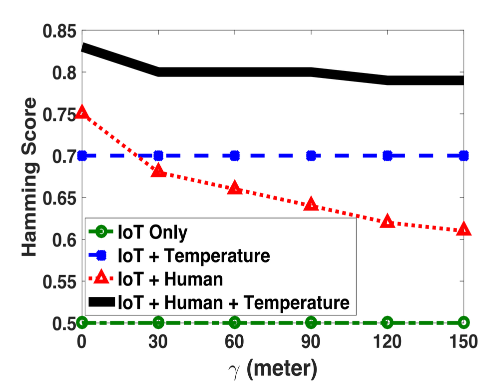




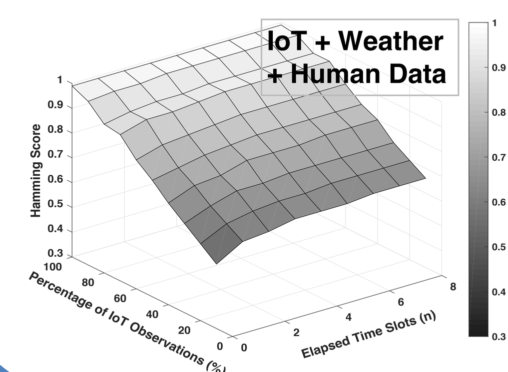
- WSSC-SUBNET is used in following results.
- With a lower percentage of IoT observations, RF yields a higher score compared with SVM.
- With more IoT data available, SVM improves the score more quickly.
- HybridRSL has the best performance.



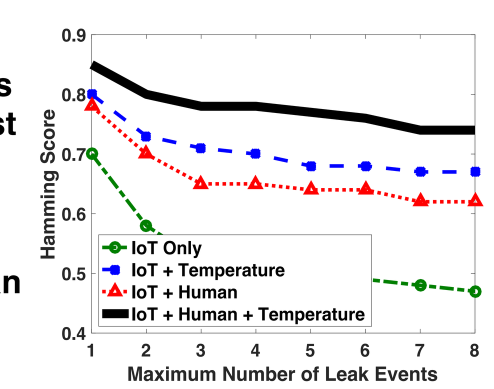
- Plotted suface shows how the hamming score varies with the percentage of IoT observations and the elapsed time slots.
- With time elapsing, more human reports may be received.



- The efficacy of incorporating with human inputs decrease with the coarser Twitter data.
- Adding temperature information can keep the score higher and relatively stable



- AquaSCALE with the integration of temperatures and human inputs is robust for detecting leak events even with a lower percentage of IoT observations or less human inputs.



- Detection using only IoT data is sensitive to the maximum number of leak events.
- The aggregation of additional information can output a better result.

CERT Center for Emergency Response Technologies · CHRS · WSSC Where Water Matters · Los Angeles Department of Water & Power · ImageCat · IRVINE SENSORIUM · IBM · thingstitute · Victory Housing · SIGFOX · Captiva · UCIrvine University of California · Montgomery County Maryland · NSF