Mapping onsite systems across the United States:

Lessons learned and advancements in machine learning based methods

Nelson da Luz PhD*

Civil and Environmental Engineering, Manhattan University, NY

Contact: nelson.daluz@manhattan.edu

*formerly at University of Massachusetts, Amherst

Background

- The last national data collection of the types of wastewater infrastructure serving the US population was the1990 Census.
- The American Community Survey (US Census Bureau) added a question about wastewater service in 2025, but it only collects data from a random subset of households in the US.
- Machine learning has emerged as a useful tool for addressing questions relating to civil infrastructure systems.
- Most communities in the US are either served by centralized sewers or onsite wastewater treatment systems (OWTS) (e.g., septic tanks).

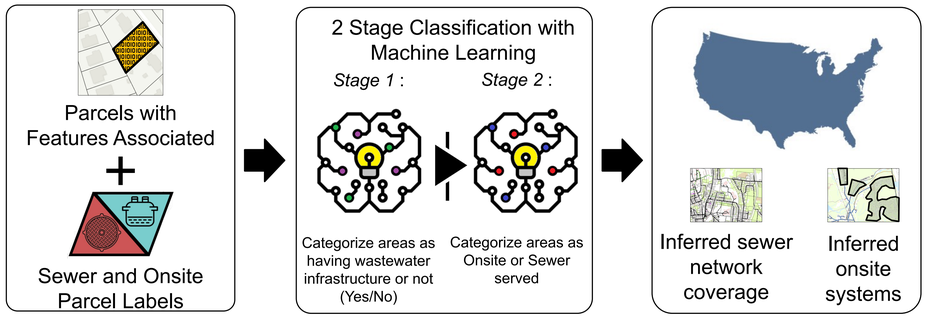
Objectives

- Predict types and locations of buried wastewater infrastructure at the land parcel scale using machine learning.
- Support states and other agencies in inventorying onsite wastewater systems

Main Machine Learning Model



1. Data Sources and Data Processing

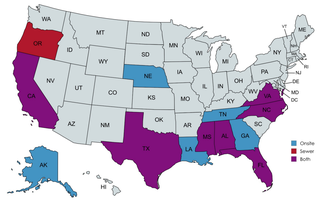


2. Machine Learning (2 Stage Random Forest) Model

3. Predictions Across the US

Model Training Data Collection

- Massive collection of land parcel level wastewater service labels
- Data from 13 states
- More than 2.5 million onsite-served parcel labels
- More than 6 million sewer served parcel labels
- More labels on the way
- Collected from online searches and from research partnerships with other universities and other agencies



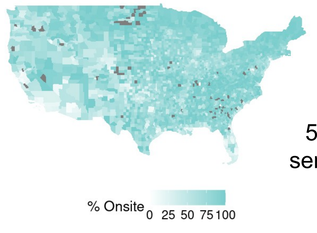
Acknowledgments

- Gulf of America Alliance
- California State Water Resources Control Board

- Florida Department of Health
- Virginia Department of Health
- North Carolina Department of Health and Human Services

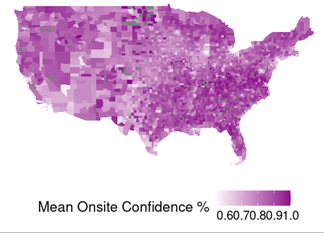
- Collaborators:
- UMass: Emily Kumpel, Jay Taneja, Kenny Hernandez, Kyle Taubert, Sylvia Imanirakiza
- Auburn University: Mallory Jordan, Stephanie Rogers
- UCLA/OWP: Greg Pierce, Dakota Keene, Khalil Lezzaik + more

Proportion of Parcels served by Onsite in each county



52M Onsite served parcels

Model Confidence in Predictions for Onsite in each county



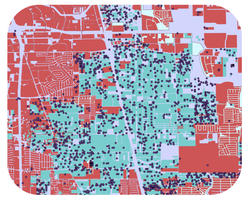
Project Highlight: Gulf of America Alliance

"Building connections, databases, and methods to understand the relationship between OWTS prevalence and water contamination in the Gulf Region"

- Predictions made across Gulf states
- Results being used as part of an OWTS pollution potential model being built by Auburn University

- The model achieved 91.2% accuracy across the Gulf states.
- Overall confidence across Gulf States was 79.2%.

State	Average Confidence
Alabama	82%
Florida	84.1%
Louisiana	78%
Mississippi	82.7%
Texas (coastal plain)	75.5%
Overall	79.2%



- The model can identify an enclave of the Houston Galveston Area that is mostly served by onsite systems amidst an area with lots of sewer

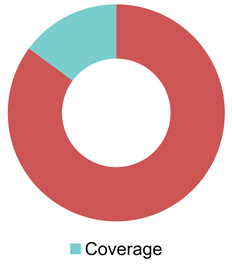
- Dots are onsite system permit locations

Project Highlight: California State Water Resources Control Board

"Wastewater Needs Assessment" (WWNA)

85.1%	14.9%	81.1%
Parcels served by sewer	Parcels unsewered	Average model confidence

- **AHS Comparison:** The 2023 American Housing Survey estimates 11% of California households are not connected to sewer systems.
- **Confidence scoring:** Each parcel's classification (Sewer or Onsite) was assigned a confidence score from 0–100.
- **44 California counties** had an average confidence score of 80% or higher.



- The model serves as a screening and planning tool for CA.
- Within the WWNA: To inform OSTS Connection Opportunities of unsewered areas into larger treatment and collection systems (e.g., septic to sewer, septic to community cluster)