QUARTERLY PROGRESS REPORT
March 2024 – May 2024

PROJECT TITLE: Evaluating and Optimizing the Value of Anaerobic Digestion of Food Waste using Sensitivity Analysis and Machine Learning

PRINCIPAL INVESTIGATOR(S):

PI: Ana Martin-Ryals, Assistant Professor, admartin@ufl.edu, 352-294-6708

Co-PI: Nikolay Bliznyuk, Associate Professor of Statistics, nbliznyuk@ufl.edu, 352-294-6734

Affiliation: Agricultural & Biological Engineering (ABE), University of Florida, Gainesville, FL

PROJECT WEBSITE: Hinkley Project - UF/IFAS Agricultural and Biological Engineering

WORK ACCOMPLISHED DURING THIS REPORTING PERIOD:

Work accomplished during this reporting period included continuing work on Objectives 2 and beginning work in Objective 3 as described below.

Objective 2 is to develop a data-driven, machine-learning based model that can predict anaerobic digestion performance as a function of feedstock characteristics and operating conditions.

During this reporting period, the expanded dataset (n=302) that was established during Quarter 2 was used to further develop the machine-learning model. The dataset contained 16 predictors variables (Table 1), encompassing eight feedstock characteristics and eight digester operating conditions, and one response variable which was experimental methane production.

Table 1. Summary of predictor variables used in Random Forest model development.

Variable Type	Categorical	Numerical
Feedstock Characteristics	• Feedstock Source	• Volatile Solids (VS) (% w/w) • Total Solids (TS) (% w/w) • VS/TS (%) • Carbohydrate Content (% TS) • Protein Content (% TS) • Lipid Content (% TS) • Feedstock C/N
Operating Conditions	• Reactor Type • Pretreatment Type	• Solid Retention Time (days) • Number of Process Stages • Digester Volume (L) • Hydraulic Retention Time (days) • Organic Loading Rate (g-VS/L-d) • Temperature (C)

One-hot encoding was used to transform categorical predictors into numerical values without ordinal relationships. A single decision regression tree using the iterative imputer method in scikit-learn was used to impute all missing values while matching expectations from domain knowledge. During imputation, the GridSearchCV tool in the scikit-learn library was used to perform hyperparameter tuning. The imputed dataset was split into training and testing datasets for 10-fold cross validation. The RandomForestRegressor method available in sci-kit learn was used to build a regression tree with all predictors, without prescreening. Hyperparameter tuning was run using GridSearchCV, and identified a maximum depth of 20, minimum number of samples per leaf of 1, and number of estimators at 100.

Figure 1 shows the ranking of the top ten predictor variables that emerged from the Random Forest analysis of the expanded dataset, as well as the initial dataset (n=100). These rankings different between the two analysis, with lipid content, volatile solids content, and feedstock source emerging as eth tip three predictor variables in the expanded dataset, versus organic loading rate, protein content and volatile solids content being the top three predictor variables in the initial dataset. The regression tree built on the expanded dataset had a mean squared error (MSE) of 15,339.37, while that of the initial dataset was closer to 7,500. An MSE of 0 indicates perfect accuracy. Thus, both datasets yielded very low prediction accuracy, with the expanded dataset having an even lower accuracy than the initial dataset.

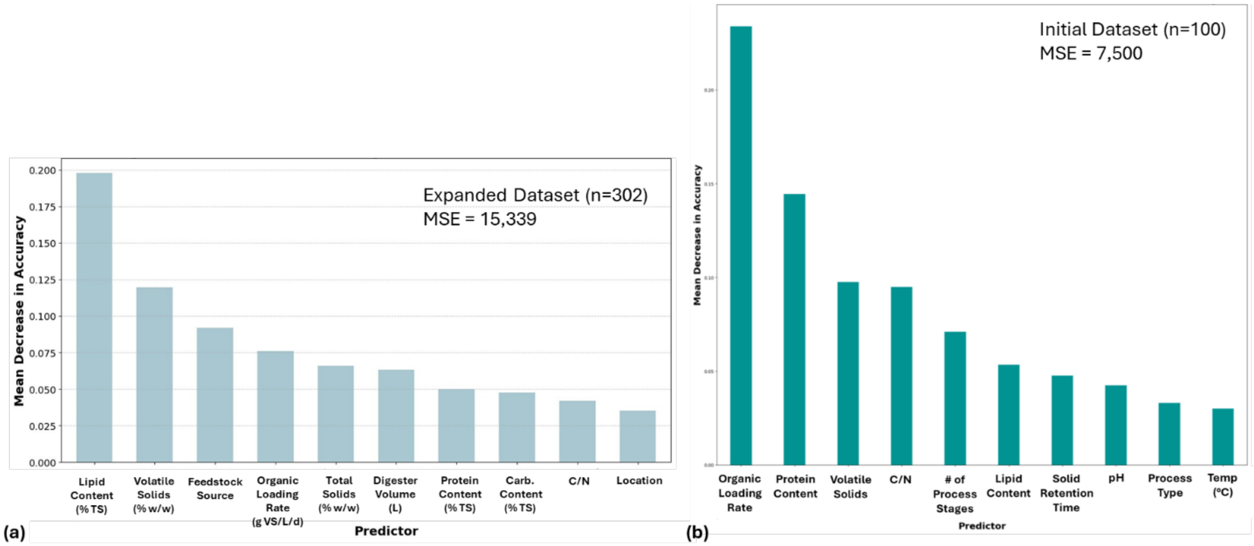

Figure 1. Ranking of most important predictor variables for Random Forest model using (a) Quarter 2's expanded dataset, and (1) Quarter 1's initial dataset.

Based on the results of this quarter's analysis, it is suspected that scale may play a factor in methane production and predictability. The initial dataset had been expanded with the aim of including more observations with values for feedstock macromolecule content (i.e., carbohydrate, protein, and lipid content). However, most of these new observations came from small-scale (< 3 liter) batch experiments (refer to Quarter 2 Progress Report for more detail).

In the last quarter of this project, we will use the expanded dataset to investigate the impact of scale on performance predictability, by diving the dataset into "small" and "large" scale datasets. We will also expand our dataset to include co-digestion of food waste with wastewater biosolids, and/or fats, oils, and grease. Finally, we will investigate the application of Extreme Gradient

Boosting (XGBoost) as a means for improving prediction accuracy. XGBoost is an increasingly popular machine learning approach that involves stage-wise regression tree construction by combining multiple poorly performing models into a single model with strong predictive ability (i.e., trees are stacked so that each new tree reduces the residuals of the previous group of trees.).

Objective 3 is to conduct an economic sensitivity analysis to evaluate the impact of feedstock and process variation on carbon intensity and overall cost.

During this reporting period, an initial economic analysis was carried out. An anaerobic digester process was modelled in the software CapdetWorks (baseline assumptions: 0.1 MGD loading rate, 3% total solids in feed, 88% volatile solids, 30-day minimum retention time), and the corresponding capital and operating costs were obtained as summarized in Table 2. Also shown in Table 2 is the estimated value for the methane that could be produced. A low, median, and high value are shown based on the low, median, and high values for methane production that were obtained from the literature review as presented in the Quarter 2 report.

Table 2. Summary of digester costs and potential methane production value, based on initial CapdetWorks model and expanded dataset methane production values.

Anaerobic Digester Cost Summary		
Present worth	23,900,000	$
Total construction cost	17,000,000	$
Annualized Digester Costs		
Operation labor	218,000	$/yr
Maintenance labor	59,300	$/yr
Material cost	127,000	$/yr
Chemical cost	0	$/yr
Energy cost	72,700	$/yr
Construction cost	1,510,000	$/yr
Total Annual Cost	**1,987,000**	**$/yr**
Methane Production Value (assuming $22.90 per thousand cf)		
Low (45 mL/g VS)	130,330	$/yr
Median (425 mL/g VS)	1,230,894	$/yr
High (801 mL/g VS)	2,319,872	$/yr

Figure 2 shows the potential net cost for the digester based on this initial analysis. Considering the cost of the digestor and the value of the methane that could be produced, the net cost of the system assuming a median methane production rate is estimated to be $756,106 per year, assuming a 40-year lifetime. This is a very preliminary analysis that will be revised and expanded on during the next reporting period to include consideration of renewable energy credits, landfill tipping fees, co-digestion, digestate management strategies, and sensitivity analysis.

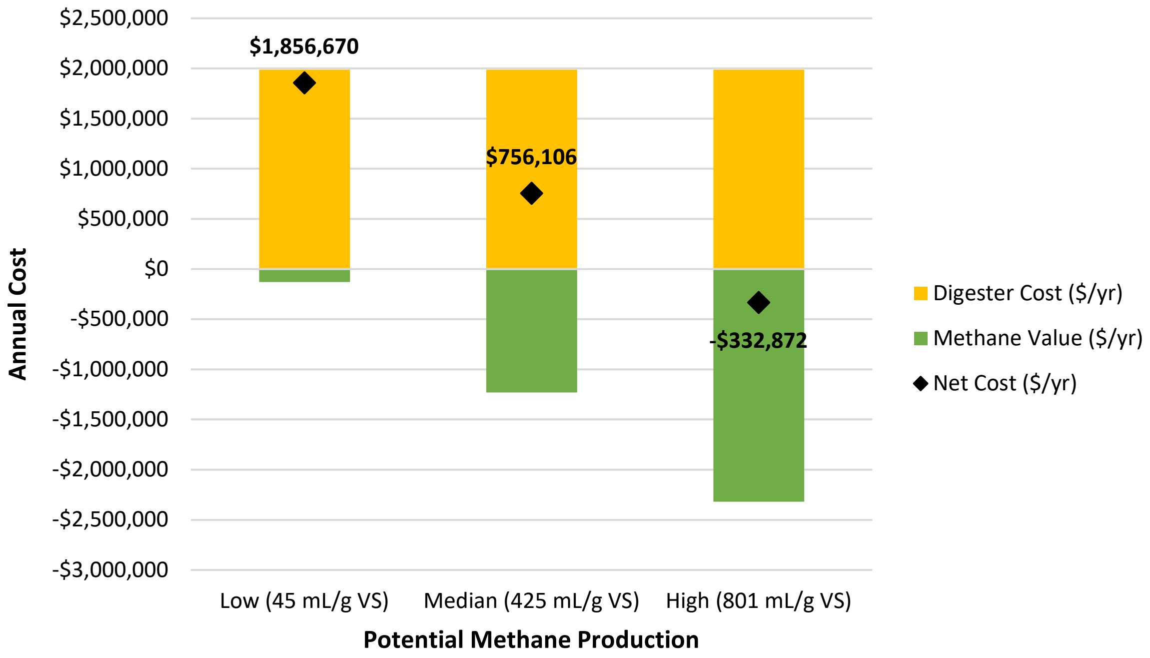

Figure 2. Estimated annual net cost for an anaerobic digester treating food waste, assuming a low, median, and high methane production rate.

TAG MEETINGS:

No official TAG Meetings were held during this reporting period. However, Quarter 2 results were shared with all TAG members via email, and with some TAG members via the Manure Lunch Seminar Series, in which PI Martin-Ryals presented Quarter 2 results in an oral presentation on March 1. TAG members in attendance included Del Bottcher and Eric Neihaus.

METRICS REPORTING:

1. Summary of input provided by the TAG during this period.

   After presenting this work at the Manure Lunch Seminar Series, questions and feedback from the audience and TAG members indicated that co-digestion of food waste with wastewater biosolids may be a more promising approach than co-digestion of food waste alone. This would likely dampen potential process instability due to feedstock variation, and by co-locating the digester at a wastewater treatment plant, would provide a convenient means for managing the digestate. This will be explored more thoroughly through literature review and economic sensitivity analysis in the last project quarter.

2. Publications resulting from **THIS** Hinkley Center project.

   None

3. Research presentations resulting from (or about) **THIS** Hinkley Center project.

- Martin-Ryals, Ana. “Advancing carbon, water and nutrient recovery from anaerobic digestion of food waste”. Manure Lunch Seminar Series. Alachua, Florida, March 1, 2024.
- Martin-Ryals, Ana. “Utilizing wet-organic waste to support a circular bioeconomy”. New Mexico State University Chemical Engineering Graduate Seminar Series. Las Cruces, New Mexico, April 26, 2024

4. List who has referenced or cited your publications from this project.

None

5. How have the research results from **THIS** Hinkley Center project been leveraged to secure additional research funding? What additional sources of funding are you seeking or have you sought?

Preliminary results from this project were leveraged to secure a 1-year research grant through the UF Space Institute. Project title: Development of an Integrated Process Model to Optimize in Situ Bioresource Utilization for Long-Duration Space Missions. PI: Ana Martin-Ryals, Co-Investigators: Amor Menezes, Nikolay Bliznyuk. Project Period: Aug. 1, 2024 – July 31, 2025. Amount: $79,204. PI Martin-Ryals will also submit an NSF Career proposal in July 2024 using preliminary results from this project.

6. What new collaborations were initiated based on **THIS** Hinkley Center project?

PI Martin-Ryals was asked to join two new teams for submission of two research grant proposals, based on results generated and meeting held related to this project. The first is a University of Florida team developing a DARPA proposal for bioregenerative space life-support systems. The second is a collaboration with colleagues at New Mexico State University, Los Alamos national Laboratory, and The University of Puerto Rico for submission of a Department of Energy proposal for anaerobic digestion and thermochemical conversion of macroalgae to biobased products, including technoeconomic analysis and life cycle assessment.

7. How have the results from **THIS** Hinkley Center funded project been used (not will be used) by the FDEP or other stakeholders?

None

PICTURES:

None at this time.