CPS 2022 RFP – FINAL PROJECT REPORT

Project Title:
Testbeds for microbial source tracking using microfluidic paper-based analytical devices

Project Period:
January 1, 2023 – December 31, 2024

Principal Investigator:
Mohit Verma, PhD
Purdue University
Agricultural and Biological Engineering
ABE 3016, 225 S University St.
West Lafayette, IN 47907-2093
T: 765-496-3687
E: msverma@purdue.edu

Objectives:

1. Characterize testbeds for distribution of aerosolized fecal contamination and generate a risk heatmap that incorporates bioaerosol distribution and environmental metadata.
2. Test the ability of microfluidic paper-based analytical devices (µPADs) to rapidly characterize bioaerosol distribution.

Funding for this project was provided partly through the CPS Campaign for Research.

FINAL REPORT

Summary of Findings and Recommendations

This project evaluated the distribution of fecal contamination around animal operations using *Bacteroidales* as an indicator by collecting more than 2,500 field-based samples. Based on these evaluations, the team found that *Bacteroidales* are an appropriate quantitative indicator of determining the distribution of fecal matter deposited by bioaerosols. By combining previous findings of baseline levels of *Bacteroidales* (0–2 copies/cm^2) and high-risk levels in close proximity to animal operations (10^3–10^4 copies/cm^2), this project determined that a rough threshold of 100 copies/cm^2 of *Bacteroidales* warrants further investigation or intervention in fresh produce operations around animal operations. Thus, the team recommends using *Bacteroidales* as a quantitative metric of fecal contamination and using these levels as indicators of risk.

The project determined that measurements of dust such as PM2.5 and PM10 levels are sometimes inversely related to the levels of fecal contamination and thus should not be solely relied upon for determining levels of fecal contamination. Also, mature trees could lead to local turbulent winds and thus, potential increase in concentration of bioaerosols (instead of solely serving as a vegetative barrier). The project demonstrated that direction is also important, in addition to distance from an animal operation, in determining the dispersal of fecal contamination.

The project demonstrated that it is possible to measure the levels of *Bacteroidales* in the field within one hour, in situations where rapid response is necessary and valuable for taking action. In cases where more time is available and shipping is a possibility, lab-based tests for *Bacteroidales* are also appropriate for determining spatial distribution of fecal contamination.

Abstract

This project aimed to utilize testbeds for microbial source tracking to monitor bioaerosol-associated fecal contamination in agricultural environments. Using *Bacteroidales* as an indicator, over 2,500 field samples were collected across diverse sites, including stone fruit farms, and lettuce fields in proximity to animal operations. Analytical methods combined environmental data such as wind patterns, particulate matter (PM2.5/PM10), and other meteorological parameters to determine their relationship with contamination levels.

Key findings revealed that *Bacteroidales* levels serve as a robust metric for fecal contamination, with a rough threshold established at 100 copies/cm^2 for actionable risk. Studies demonstrated a strong correlation between contamination levels and proximity to animal operations, while weather conditions, particularly temperature and wind direction, significantly influenced dispersal patterns. Contrarily, PM2.5 and PM10 levels were inversely correlated with contamination, suggesting complex interactions affecting microbial deposition.

Innovative tools were tested for rapid field detection, including microfluidic paper-based analytical devices (µPADs) integrated with loop-mediated isothermal amplification (LAMP) technology, providing contamination results within one hour. Predictive models utilizing machine learning techniques such as Random Forest Regression integrated with Fast Fourier Transform preprocessing achieved moderate accuracy ($R^2 = 0.50$), highlighting critical environmental predictors like temperature and humidity.

Outcomes included actionable tools and recommendations for improving contamination monitoring. The study recommends incorporating *Bacteroidales* thresholds into fresh produce safety protocols and highlights the importance of accounting for wind direction and localized weather patterns. It advises against sole reliance on PM2.5 and PM10 levels for contamination assessment and encourages the integration of portable analytical tools for real-time monitoring. These findings provide a foundation for sustainable, data-driven approaches to mitigate microbial risks in agricultural contexts.

Background

One of the factors affecting the safety of fresh produce is the use of adjacent and nearby land. California Leafy Greens Marketing Agreement (LGMA) and Arizona LGMA maintain dynamic documents that provide guidelines for assessing risk in fresh produce operations in the pre-season, pre-harvest, and harvest stages.[1,2] These guidelines are driven by ongoing research and development for continued alignment with the Food Safety Modernization Act Final Rule on Produce Safety.[3] LGMAs also provide additional resources such as examples of standard operating procedures (SOPs) for assessing risk from concentrated animal feeding operations (CAFOs)[4] and pre-harvest testing guidelines.[2,5] These guidelines provide minimum distances for various animal operations but also recognize that distance alone is not sufficient to ensure the safety of fresh produce. The guidelines also acknowledge the importance of bioaerosols as one of the key factors for the dispersal of contaminants from animal operations. Yet, we lack tools to quantitatively characterize the risk from the bioaerosols at a specific site.

Our previous studies have shown that bacterial order *Bacteroidales* can serve as a reliable quantitative marker for fecal contamination and can be effectively collected using a simple plastic sheet attached to a wooden skewer, referred to as a "collection flag."[6] Background levels of *Bacteroidales* near leafy vegetable operations have been found to be low (0–2 copies/cm²).[7] However, significantly higher levels (10^3–10^4 copies/cm²) have been observed in areas closer to experimental animal operations.[6] Thus, there is a three to four orders of magnitude difference in concentration between baseline levels of *Bacteroidales* and high-risk levels of *Bacteroidales.* This project aimed to fill in the knowledge between these two extremes to determine the level of *Bacteroidales* in field operations where fresh produce is being grown in proximity to animal operations. To fill this gap, we selected three testbeds—1) Animal Science Research and Education Center (ASREC) in Indiana, 2) Stone fruit operations, California, 3) Leafy greens operations, Arizona—and characterized the distribution of *Bacteroidales* at these testbeds by collecting more than 2,500 data points from these sites.

While the lab-based approach of measuring fecal contamination is valuable, it can be further enhanced using portable analytical tools that allow the determination of levels of *Bacteroidales* in the field. We developed such a tool using a molecular technique known as loop-mediated isothermal amplification (LAMP) and implementing it in the field.[8] Such devices enable the user to get a response within an hour instead of waiting for several days to ship samples to a lab and await processing. The current project aimed to test these devices in field conditions and obtain feedback from potential users on the user-friendliness of these devices.

Research Methods and Results

Objective 1: Characterize testbeds for distribution of aerosolized fecal contamination and generate a risk heatmap that incorporates bioaerosol distribution and environmental metadata.

1. Animal Science Research and Education Center (ASREC) at Purdue University (Testbed 1)

1.1 10-week longitudinal study

This project aimed to incorporate environmental metadata into the analysis of fecal contamination risk around animal operations. Thus, the team used ASREC as a site for testing the installation of mini-weather stations and for conducting a longitudinal study. The team implemented a 10-week study at Testbed 1 in 2023 to evaluate the dynamics of contamination near a swine feeding operation. The study involved the deployment of eight mini-weather stations (Ambient WS-5000) and eight Ambient AQIN air quality monitors across the field. However, due to technical issues with weather station 5, data from seven weather stations were utilized in the analysis. The weather sensors capture hyper-local weather data, including temperature, pressure, humidity, rainfall, UV index, solar radiation, wind speed, and direction. The AQIN air quality monitors measured CO_2, $PM_{2.5}$ (particulate matter 2.5 µm), and PM_{10}

(particulate matter 10 µm) levels. To capture contamination, 120 collection flags were positioned at 12 designated points, organized into three rows (Figure 1, illustrating simplified map of ASREC testing location with weather station and flag placement spots). The flags were spaced approximately 15 meters apart from north to south and 25 meters apart from east to west, in proximity to the swine feeding operation. 12 collection flags were retrieved weekly. These samples were analyzed in the laboratory using quantitative polymerase chain reaction (qPCR) to assess *Bacteroidales* levels. Testbed 1 lacked a fresh produce operation.

Each collection flag of a 329 cm^2 surface area was swabbed for target deoxyribonucleic acid (DNA) and its sample was resuspended in 200 µL of nuclease-free water[8]. The Luna® Universal Probe qPCR Master Mix (New England Biolabs, M3004) protocol was used on the residual swab resuspension samples taken from the collection flags on the field. Weather data and field sample data were processed using Python's open-source libraries. Field sample Ct data was first converted into $\log_{10}$(copies/cm^2) via a linear fit to log-transformed concentrations (eq 1-4). Therefore, copies per cm^2 of each collection flag were calculated by dividing the swabbed DNA copy number with the total surface area of the flag (eq 1-4)[8]. Then, the resulting $\log_{10}$(copies/cm^2) of each flag sample for 10 weeks was output as a heatmap, and overlayed onto a simple map generated with Microsoft PowerPoint created for the field of interest for easier visual analysis. The concentration heatmaps increase in color from yellow to red, indicating higher concentrations of *Bacteroidales* in the field samples (0 to 3 $\log_{10}$(copies/cm^2)) (Figure 2).

$$log_{10}\ Copies\ per\ reaction = \frac{Ct\ of\ qPCR\ reaction - 35.08}{-3.7617} \quad (1)$$

$$Copies\ per\ reaction\ in\ 1\ \mu L = 10^{log_{10}\ Copies\ per\ reaction} \quad (2)$$

$$Copies\ per\ cm^2 = \frac{200\ \mu L}{329\ cm^2} * (Copies\ per\ reaction\ in\ 1\ \mu L) \quad (3)$$

$$log_{10}\ Copies\ per\ cm^2 = log_{10}(Copies\ per\ cm^2) \quad (4)$$

Using Python's open-source libraries, wind (Figure 2), PM2.5, and PM10 data (Figure 3) were overlaid onto a simplified map of ASREC along with heatmaps to visualize environmental conditions. Wind speed and direction were represented using directional arrows, with the arrow length indicating wind speed and orientation reflecting wind flow. PM2.5 concentrations were shown as ovals, with sizes varying from less than 1 to 20 µg/m³, while PM10 concentrations were displayed as triangles using the same size scale for consistency. This integrated visualization provides a clear depiction of wind patterns and particulate matter distribution across the field.

This study allowed the team to determine the consistency and accuracy of the weather stations across different equipment. In addition, it provided data to build a prediction model of the levels of *Bacteroidales* in the field depending on weather parameters.

1.2 300-meter distance study

The team also performed a distance study at Testbed 1 where 20 rows of flags were laid out, flags spaced 15 meters part (for a total of 300 meters), extending away from the swine unit (south to north). Two weather stations were used**.** The heatmap is displayed on the simplified map of ASREC, with wind speed and direction overlayed in Figure 4. The missing flags in the ASREC distance study are represented by the gray colored areas on the heatmap (Figure 4). As seen from the results of the distance study, there is a correlation between distance from the swine unit, or source of contamination, and the concentration of fecal contamination on the field as the collection flag samples further from the swine unit tend to express lower $\log_{10}$(copies/cm^2) values.

1.3 Prediction model and user interface

We developed a predictive model to estimate *Bacteroidales* concentrations based on weather data, integrating a Python-based graphical user interface (GUI). To enhance the model's predictive accuracy, the collected weather data underwent Fast Fourier Transform (FFT) analysis.

Incorporation of Fast Fourier Transform (FFT) in Data Processing: The ASREC 10-week experiment collected environmental data at 5-minute intervals from eight weather stations positioned across the field. Among these, data from weather station 8 was selected for FFT processing due to its central placement and alignment with True North, which provided a comprehensive and unbiased representation of field conditions. This station's alignment also eliminated the need for wind direction conversions, ensuring data integrity. The FFT, conducted using the SciPy library in Python, transformed time-domain signals into the frequency domain, revealing key frequency components that capture recurring patterns in weather parameters over time. For each frequency-domain signal, the magnitude (indicating the strength of a frequency component) and phase (representing its temporal shift) were calculated. Low-frequency components, characterized by larger magnitudes, were prioritized as they represent dominant, consistent trends. This process identified the 12 most significant low-frequency values for each weather parameter per week, aligning the weather data length with the 120 field sample concentration values.

Model Construction and Feature Selection: The predictive model was constructed using machine learning techniques, with datasets from on-field experiments split into 80% training and 20% testing. The FFT-derived weather features were utilized as predictors, including parameters such as temperature, solar radiation, wind speed, CO_2 levels, and particulate matter (PM10). Several regression models—linear regression, LASSO, decision trees, support vector machines, and random forest regression (RFR)—were evaluated.[9–19] The *random_state* was defined as '42' for each of the models to ensure reproducibility and consistency in the data splitting for comparable results.[20] The RFR model demonstrated the best performance prior to optimization (MSE = 0.857; R^2 = 0.082) (Table 1 for comparison). Further refinement through hyperparameter tuning and recursive feature elimination (RFE) identified the 10 most significant predictors (Table 2), ensuring the inclusion of only the most relevant weather parameters for contamination risk assessment.

Model Performance and Insights: The optimized RFR model achieved an MSE of 0.693 and an R^2 of 0.258, with a Pearson correlation coefficient of 0.512, indicating moderate predictive accuracy (Table 3). The incorporation of FFT-transformed weather data played a pivotal role in capturing patterns related to *Bacteroidales* concentrations. While discrepancies between predicted and actual values were observed (Table 3 and Figure 5), the model identified key environmental factors influencing contamination risks.

Implications and Future Directions: This study underscores the utility of FFT in transforming complex time-series weather data into actionable features for machine learning models. By focusing on low-frequency components, FFT preprocessing enhanced the model's ability to detect relationships between environmental factors and contamination levels. Although the current model explains approximately 26% of the variance in *Bacteroidales* concentrations, it provides a strong foundation for further optimization. Future efforts will focus on refining the model through advanced methodologies, exploring non-linear relationships, and integrating additional environmental datasets. These enhancements will aim to improve predictive accuracy and broaden the model's applicability across diverse agricultural contexts, ultimately contributing to real-time contamination risk assessment in fresh produce operations.

Visualization of Predicted *Bacteroidales* Concentrations via Python GUI: A Python-based graphical user interface (GUI) was developed to enable visualization of predicted *Bacteroidales* concentrations on agricultural fields. Built with Tkinter, the interface integrates several open-source Python libraries, including OpenCV (image processing), Pillow (image handling), NumPy (large array support), Matplotlib (plotting), and Pandas (data analysis). The GUI has the following functionality:

- Input Parameters: Users specify weather parameters, the number of flags or regions for analysis, and the desired output grid layout (rows and columns).
- Field Image Selection: Users upload a field image and outline the area of interest by drawing a boxed region.
- Output Visualization: The program generates a visual overlay of the field with predicted *Bacteroidales* concentrations and provides numerical values for the selected region.

This interactive tool (Figure 6) combines real-time visualization and data analysis, allowing users to evaluate contamination risks in specific areas of their farm with ease and precision.

2. Experiments at Testbed 2) Installation of collection flags, weather stations, and air quality monitors – Stone fruit farm, in California

2.1 Distribution of Bacteroidales *during a seasonal study*

The Purdue team collaborated with George Nikolich of Nikolich Consulting in California, with the objective of gathering weather data and bioaerosol distribution of *Bacteroidales* for two seasons. The first study was in August 2023 and the second study was in April 2024. Both the studies took place at stone fruit production fields adjacent to a cattle feedlot (Figure 7, simplified map of the site). The team fabricated rectangular collection flags measuring 7.62 × 21.59 cm and affixed them to wooden stakes of height 4 ft (Figure 8). In the August 2023 study, a total of 390 of these collection flags were deployed across the site. Additionally, eight Ambient WS-5000 weather stations and eight Ambient AQIN air quality monitors were strategically positioned among plum (north) and peach orchards (south), the feedlot (central area), and the vineyard (west) (Figure 9). Similarly, in April 2024, 812 collection flags were deployed across the sites along with five weather stations and air quality monitors. Figure 9 shows the relative locations of the flags (blue regions), WS-5000 weather stations (yellow circles with numbers), and air quality monitors. The collection flags and weather stations were placed on the locations for 7 days. At the end of 7 days, collections flags were collected in individual Ziploc bags for overnight shipment to the Purdue laboratory. Upon receipt, the Purdue team swabbed both sides of each flag using wet polyester-tipped swabs, subsequently resuspending each swab in 200 µL of nuclease-free water. The analysis involves the detection of *Bacteroidales* using the qPCR assay. After calculating the levels of the bacteria using equations 1-4, heatmaps were generated and overlayed on the simplified map of the testing location (Figure 10).

The levels of *Bacteroidales* in August 2023 ranged from 0 to 134.90 copies/cm², while in August 2024, they ranged from 0 to 23.49 copies/cm². This substantial difference highlights a notable decrease in *Bacteroidales* concentrations between the two years/seasons. Also, it seems that approximately 100 copies/cm^2 of *Bacteroidales* is a rough threshold of where additional investigation or intervention should be implemented if fresh produce is being grown in these areas. This number is based on previous observations of baseline (0–2 copies/cm^2)[7] and around animal operations (10^3–10^4 copies/cm^2)[6] and the observation around stone fruit operations (in proximity to ~3000–4000 cattle).

2.2 Changes in weather during a seasonal study

The average values of weather parameters for each weather station were calculated for 2023 and 2024 and summarized in Tables 4 and 5. Statistical analysis comparing these averages was performed using a t-test to compare means and a Mann-Whitney U (MWU) test to account for the small, non-parametric

dataset (Table 6). Significant differences were observed in temperature, humidity, dew point, PM2.5, and PM10 (Figure 11). Temperature was significantly higher in 2023, coinciding with increased *Bacteroidales* concentrations, likely due to the combination of higher temperatures and lower humidity (Δ = 10.53%), which may have facilitated contaminant dispersal. In contrast, PM2.5 (Δ = 14.19 µg/m³) and PM10 (Δ = 14.09 µg/m³) concentrations were higher in 2024, despite lower *Bacteroidales* levels. Significant variability in dew point data was observed in 2023, likely influenced by microenvironmental differences at weather station locations, highlighting localized impacts on atmospheric moisture.

Considering that PM2.5 (Δ = 14.19 µg/m³) and PM10 (Δ = 14.09 µg/m³) concentrations were higher in 2024 despite lower *Bacteroidales* levels, further analysis was conducted to explore this observation. Particles on 90 collection flags from the 2024 study were examined by placing the flags on a black background and imaging them using a digital scanner. Using a Python-based program, particles were counted and analyzed (Figure 12). A linear regression analysis was performed to assess the relationship between the number of particles and *Bacteroidales* concentrations, expressed as $\log_{10}$(copies/cm²) (Figure 13). The regression line revealed a slight negative trend, indicating that as the number of particles increased, *Bacteroidales* concentrations slightly decreased. The R^2 value of 0.3784 suggests that approximately 37.84% of the variation in *Bacteroidales* concentrations can be explained by the number of particles, representing a weak fit.

The inverse correlation between dust and level of *Bacteroidales* is surprising and potentially counterintuitive since dust is often considered a carrier of microbial matter. This inverse trend might reflect a dilution effect, where an increase in airborne particulate matter creates a more dispersed environment, limiting the deposition of *Bacteroidales* onto surfaces. Alternatively, it could suggest a competing effect where higher concentrations of PM2.5 and PM10 reduce the survivability or transport efficiency of *Bacteroidales*. Also, it's possible that the size of aerosols that carry *Bacteroidales* are smaller than 2.5 µm and thus, not accounted for in the particle count measures. It is also possible that the dust particles cause an inhibition of the qPCR assay and this issue could be investigated in the future.

2.3 Prediction model

Using the *Bacteroidales* concentration in $\log_{10}$(copies/cm^2) and weather data from August 2023 and April 2024, the team developed a prediction model. The model was developed to estimate fecal contamination concentrations on flags placed at different locations around a feedlot based on weather data. The model integrates weather parameters, spatial distance, and directional attributes of each flag relative to the feedlot to improve prediction accuracy.

The following three steps are used for data preprocessing:

1. Determine the Feedlot Centroid: The centroid of the feedlot was calculated using the formula for the centroid of an arbitrary polygon defined by n vertices $(x_0, y_0), (x_1, y_1), \dots (x_{n-1}, y_{n-1})$:
   Centroid, $\left(C_x, C_y\right) = (\frac{1}{6A}\Sigma_{i=0}^{n-1}(x_i + x_{i+1})(x_i y_{i+1} - x_{i+1}y_i), \frac{1}{6A}\Sigma_{i=0}^{n-1}(y_i + y_{i+1})(x_i y_{i+1} - x_{i+1}y_i))$, Area, $A = \frac{1}{2}\Sigma_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1}y_i)$
2. Calculate Euclidean Distance: The Euclidean distance of each flag from the feedlot centroid was computed as: Distance, $d = \sqrt{[(x_c - x_i)^2(y_c - y_i)^2]}$ For coordinates of feedlot centroid (x_c, y_c), and coordinates of flag (x_i, y_i)
   Weighted Averaging of Weather Data: Weather stations located at various positions among the flags recorded data at different times. The average weather data was weighted based on proximity to the time the flags were retrieved for measurement. For each measurement, an index was assigned, with the most recent measurement receiving the highest index.

The weight of a data point was calculated as:

$$w_i = \frac{2i}{n(n+1)}$$

Where i is the index of the data point and n is the number of data points.

Weighted Average $a = \frac{\Sigma_{i=0}^{n-1} w_i m_i}{\Sigma_{i=0}^{n-1} w_i}$, for measurements m_i

These values were unique to each weather station. For each flag, the weighted average data from each station was weighed by the inverse of the distance from the flag to each station.

Station weight for flag i and station j, $s_{ij} = \frac{1}{d_{ij}\Sigma_{j=0}^{n'-1}\frac{1}{d_{ij}}}$

From Euclidean distance d_{ij} between flag i and station j, and number of weather stations n'

The final value for a flag i was computed as: $\frac{\Sigma_{j=0}^{n-1} s_{ij} a_i}{\Sigma_{j=0}^{n-1} s_{ij}}$

3. Data Normalization: All data were standard normalized to ensure uniformity across features.

Supervised Learning and Model Development: The dataset was divided into training (80%) and testing (20%) subsets. The training set was used to build the regression model, while the testing set was reserved for performance evaluation. A Voting Regressor (VR) was implemented, combining Random Forest Regressor, Linear Regressor and Gradient Boosting Regressor. The VR averages the predictions from the three methods. The model achieved an R^2 value of 50.22%, indicating the proportion of variance in the data explained by the model (Figure 14). The Root Mean Squared Error (RMSE) was 0.71 $\log_{10}$(copies/cm²), reflecting the average deviation between predicted and actual values. Additionally, the Mean Absolute Error (MAE) was 0.53 $\log_{10}$(copies/cm²), highlighting the average magnitude of errors in the predictions. The actual and predicted $\log_{10}$(copies/cm²) for the 20% testing data are listed in Table 7. The team determined the top 10 parameters that contribute to the variation by first extracting the importance scores of the features from the Random Forest model using recursive feature elimination. Next, the team created a DataFrame (Table 8) that included the feature names and their corresponding importance scores. The team found that temperature had the highest importance score of 0.283, followed by direction_y with a score of 0.164, and humidity with a score of 0.154. The direction_y feature includes both the distance and whether the sample was calculated north or south of the centroid of the feedlot. PM10 surprisingly made at the bottom of the list with an importance factor of 0.033.

Unsupervised Learning: The team analyzed the *Bacteroidales* concentration in $\log_{10}$(copies/cm²) and pre-processed weather data from August 2023 and April 2024 using unsupervised learning. The clustering process began with hierarchical clustering (Figure 15), which revealed four distinct clusters within the dataset. This initial analysis provided a foundation for determining the optimal number of clusters. Based on this insight, K-Means Clustering (Figure 16) was applied, leveraging the elbow method to analyze inertia scores and validate the selection of four clusters. The K-Means clustering algorithm was then utilized to assign each data point to one of these clusters, enabling a more structured and interpretable segmentation of the dataset.

The K-cluster-based Analysis (Table 9 showing average of each parameter in the clusters): The cluster-based analysis reveals distinct trends in the weather parameters and their relationship with $\log_{10}$(copies/cm²). Cluster 1 demonstrates the highest mean value for $\log_{10}$(copies/cm²) (1.41), coupled with higher dew point (55.24 °C), elevated solar radiation (134.56 W/m²), and a moderate distance (507.48 m). In contrast, Cluster 0 and Cluster 2 exhibit similar log10(copies/cm²) values (0.54), with Cluster 2 having slightly higher dew points (13.87 °C vs. 14.03 °C) and gust speeds (3.89 m/s). These clusters also have higher PM10 and PM2.5 levels compared to others. Cluster 3 shows the lowest

log_{10}(copies/cm²) (0.41), higher distances (584.56 m), and solar radiation values (127.23 W/m²), indicating a potential dilution effect at greater distances. Wind directions and speeds also vary significantly across clusters, with Cluster 2 having the highest wind speed (2.70 m/s), while Cluster 3 has the lowest (1.32 m/s). These trends suggest that clusters differ significantly in environmental conditions, which could influence contamination levels

The hierarchical clustering grouped the data into four distinct clusters (Table 10, showing average of each parameter in the clusters) based on similarities in contamination levels and environmental parameters. The cluster-based analysis reveals distinct trends across environmental and contamination parameters. Cluster 1 exhibits the highest contamination level, with a mean log_{10}(copies/cm²) of 1.408, coupled with elevated dew point (55.24 °C) and temperature (27.64 °C). This cluster also shows moderate wind speeds (1.381 m/s) and lower particulate matter concentrations (PM10 = 7.348 µg/m³, PM2.5 = 7.026 µg/m³). In contrast, Cluster 2 has the lowest contamination level (log_{10}(copies/cm²) = 0.405), higher distances (584.56 m), and moderate temperatures (27.44 °C), with similarly low particulate matter concentrations. Cluster 3 demonstrates moderate contamination levels (log_{10}(copies/cm²) = 0.549) and elevated particulate matter concentrations (PM10 = 18.912 µg/m³, PM2.5 = 18.724 µg/m³), along with higher wind speeds (3.890 m/s) and moderate solar radiation (188.884 W/m²). Lastly, Cluster 4 shows similar contamination levels to Cluster 3 (log_{10}(copies/cm²) = 0.539), with slightly lower wind speeds (2.702 m/s) and comparable particulate matter levels (PM10 = 17.523 µg/m³, PM2.5 = 17.237 µg/m³). These variations highlight the influence of environmental factors such as wind speed, particulate matter, and temperature on contamination levels across clusters.

3. Experiments at Testbed 3) Installation of collection flags, weather stations, and air quality monitors – Lettuce fields, Arizona

Spatial Distribution of Bacteroidales

The Purdue team established a collaborative effort with Dr. Channah Rock, The University of Arizona, and an industrial fresh produce operation in Arizona, with the objective of gathering data on the bioaerosol distribution of *Bacteroidales* and weather.

The Purdue team was provided access to one lettuce field at an experimental site (1.5 miles from a concentrated animal feeding operation (CAFO), spanning 2.5 acres) where we placed 225 flags 66 meters apart, with one weather station and air quality monitor in the north (located behind the bush) and one in south (Figure 17). For non-research locations, we used the boundaries of a canal system around the CAFO; mostly >1200 feet from the CAFO) for placing flags (445 flags) and weather stations. The lettuce field, sites along the canal, and weather station's locations are marked on the simplified map in Figure 17. Upon concluding the 7-day experiment, the collection flags were meticulously gathered and placed in individual resealable plastic bags for overnight shipment to the Purdue laboratory. Upon receipt, the Purdue team swabbed both sides of each flag using wet polyester-tipped swabs, subsequently resuspending each swab in 200 µL of nuclease-free water. The analysis involves the detection of *Bacteroidales* in all samples using a qPCR assay. The resulting data was used to generate heatmap plots based on the log_{10}(copies/cm^2) of *Bacteroidales*. The heatmap for *Bacteriodales* levels, wind speed and wind direction are overlayed on the simplified maps for the tested sites (Figure 18). The weather station 3 had considerably different wind speed because it was surrounded by bushes. The manual for the weather stations suggests that for accurate sensing it is advisable to install weather stations in locations free from obstructions. The average of each weather parameter from stations 1 and 2 are provided in Table 11.

The *Bacteroidales* levels are higher near the feedlot (Figure 18), with elevated PM2.5 and PM10 concentrations also observed in proximity. However, this trend contrasts with the data collected in

California, suggesting either site-specific differences in contamination dispersal or a minimal role of PM2.5 and PM10 in spreading fecal contamination, emphasizing the need to focus on other weather parameters. This deduction agrees with the importance score established for PM10 (0.033, Table 8) in the prediction model of seasonal study.

Similarly, for the industrial partner fields, we followed a similar protocol. The fields spanned around 7 acres of iceberg lettuce in total. There were no animal operations around, however, one of the lots (Field-3) had been problematic in the past that enhanced our interest in testing it for *Bacteroidales*. We placed 425 flags in one field and 225 flags in the other, with weather stations. The heatmap of *Bacteriodales* levels and wind speed and direction are in Figure 19. We did not see *Bacteriodales* at these sites except of one sample showing very high concentrations which is likely contamination in the qPCR assay (in Figure 19).

4. Software application for data visualization

To develop customizable and downloadable visualizations, a web application was created using a Python-based WSGI framework to enable users to add qPCR and weather station data specific to user farms. Figure 20 presents an overview of the web platform functionalities.

First, users are prompted to upload a map of their site of interest (either a satellite image or an anonymized map), depending on their desired level of anonymity. After receiving an input map, the platform then prompts the user to mark any qPCR testing sites and weather station locations as "zones of interest" at their respective locations on the map. The user can then upload data corresponding to these locations.

Once user data has been uploaded for each location, the algorithm sorts the qPCR and weather data by processed and raw formats, relying on keywords within the uploaded data files. To assign raw qPCR data, a user must upload all well plate files related to the qPCR zone of interest and then further assign specific wells to the location within the zone of interest. Processed qPCR data uploads need no further user inputs. Processed weather data, which summarizes all data for multiple weather stations, is split by weather station, and the user assigns each weather station to a specific mapped zone. Uploading raw weather data requires two file uploads: weather-specific data and sensor data. Raw weather data needs no further user input.

The qPCR files are then processed by copies/cm^2 to generate heat maps on a $\log_{10}$ scale. Weather stations are filtered through to isolate key visualization features like wind, particulate concentrations, hourly rainfall, solar radiation, and carbon dioxide concentrations. These qPCR and weather station visualizations are then able to be downloaded for the user to visualize or analyze.

Objective 2: Test the ability of microfluidic paper-based analytical devices (µPADs) to rapidly characterize bioaerosol distribution.

In order to implement the analysis of *Bacteroidales* in the field, we have incorporated LAMP assay for *Bacteroidales* on µPADS. To run these assays, we have developed a heating/imaging device and a sample distribution device. We used two different approaches for sample distribution and tested them with potential users.

1. Approach 1, Field-Applicable Sampling Tool (FAST):

1.1 Overview

The Field-Applicable Sampling Tool (FAST) is a biodegradable, microfluidic device designed to facilitate the resuspension and distribution of swab samples onto µPADs under field conditions. The device

incorporates a case and slider mechanism that ensures precise volume delivery from a resuspension chamber to the µPADs (Figure 21). The liquid is guided through y-junction channels to rehydrate the paper pads effectively (Figure 22). This design provides an easy-to-use, cost-effective alternative to laboratory-grade micropipettes and supports processing up to 10 samples simultaneously.

To complement FAST, a heating-imaging platform was developed for conducting Loop-mediated Isothermal Amplification (LAMP) reactions in field settings. This portable system integrates hardware and software to enable precise temperature control, automated data collection, and real-time monitoring of LAMP reactions (Figure 24). Together, these innovations offer a solution for on-site diagnostics, especially for detecting *Bacteroidales* in environmental samples.

1.2 Materials and Methods

The FAST device was designed and fabricated using 3D-printing and laser-cutting technologies. The top and slider components were manufactured using a Bambu Lab X1-Carbon Combo 3D printer with Flashforge PETG Pro filament. These components were printed with a sparse infill density of 15% and default PETG settings to ensure structural integrity and precision.

The bottom part with y-channels was fabricated using an Anycubic Photon D2 resin printer with plant-based resin. Detailed slice settings included a layer thickness of 0.050 mm and exposure times of 2.3 seconds. Postprocessing involved a 15-minute isopropyl alcohol wash, followed by air drying and ultraviolet curing for six minutes to ensure the component's durability and cleanliness.

The cartridge housing the µPADs, was laser-cut from medium black acrylic latex sheets using a Glowforge Plus machine. Adhesive PCR plate seals were also laser-cut to create pre-cut adhesive layers, ensuring the µPADs were securely affixed within the cartridge.

Assembly of the FAST device involved bonding the top and bottom components using Gorilla super glue, followed by precise alignment of pegs and holes under weighted pressure for 30 minutes. After a 24-hour curing period, the slider was prepared by attaching nitrile butadiene rubber O-rings and a 10 mm x 10 mm mesh to its openings using adhesive. All assembled components were cleaned with RNAse Away and nuclease-free water in a PCR workstation to ensure sterility. The µPADs were added to the cartridge under sterile conditions and secured with pre-cut adhesive layers. The steps for the fabrication of the µPADs and collection flags are documented in Wang *et al.*[6]

The heating-imaging platform was constructed using 3D-printed parts and integrated electronics to maintain a stable temperature of 65 °C during LAMP reactions. The embedded camera captured time-lapse images every minute, and Python-based algorithms analyzed the colorimetric changes on the paper pads, producing sigmoidal curves akin to qPCR results. The entire system was designed for portability, operating efficiently with minimal power consumption.

1.3 Results

We tested the Field-Applicable Sampling Tool (FAST) and the heating-imaging platform with three independent users in California, including one experienced participant and two novice users. These individuals operated the system under guided conditions, and their feedback emphasized the device's ease of use, particularly for those unfamiliar with laboratory techniques (Figure 23).

Field validation extended to stone fruit farms, where the FAST device and heating-imaging platform were tested to confirm their functionality in real-world scenarios (Figure 24). We also deployed the system at a commercial lettuce farm to detect *Bacteroidales* contamination. During these experiments, the integrated platform was operated from the trunk of a car, powered by a 1000W battery, demonstrating its portability and adaptability.

The heating-imaging platform successfully reduced both the cost and turnaround time for LAMP reactions, delivering reliable results within 60 minutes detection (Figure 25). The system maintained precise temperature control with minimal fluctuations (± 0.2 °C), ensuring optimal conditions for accurate.

In laboratory settings, the FAST device achieved an average accuracy of 89% for equal flow-splitting and 70% for volume hydration (Figure 22). While these results validated the device's functionality, the accuracy was not sufficient for precise quantitative applications. Consequently, we pursued another approach (Approach 2), which involved using glass capillaries instead of micropipettes to enhance accuracy and improve quantitative performance.

2. Approach 2, Glass capillaries for fluid distribution:

To enhance the accuracy of the system, we incorporated glass capillaries (VWR Catalog Number 53507-148) into the design. The glass capillaries utilize capillary action to draw sample liquid and transfer it to the paper pads via reverse capillary action. This method ensures precise volume delivery to the µPADs, addressing limitations observed with the initial approach.

Several modifications were made to accommodate the use of glass capillaries. The size of the µPADs was reduced from 5x6 mm to 3x3 mm. The smaller µPADs require only 7.5 µL of reagents to be dried, compared to 25 µL for the larger pads, significantly reducing the cost of each pad. The glass capillaries, capable of holding 7 µL of sample, efficiently hydrate the smaller pads during reverse capillary action. Also, the design of the chips and the cartridge holding the µPADs was updated to align with the new pad dimensions and capillary-based workflow. These design changes are detailed in Figures 26 and 27.

To evaluate the performance of the updated system, we deployed it in the field office of a swine research unit. A worker from the swine unit was requested to run the assay using the glass capillary system and provide feedback on its usability. Additionally, we engaged another startup's employee to perform the assay and offer their input (Figure 28). The results obtained using glass capillaries were comparable to those achieved with traditional micropipettes (Figure 29), demonstrating that the updated system-maintained precision while providing a more cost-effective and field-friendly solution.

The users provided the following feedback:

1. Handling Chips and Glass Capillaries:
   Challenge: User-2 initially faced difficulties handling the small-sized chips and glass capillaries.
   Observation: After working with 1-2 chips, User-2 appeared more confident.
   Suggestion: Users recommended providing 2-3 practice chips to allow for familiarization before starting the actual experiment.
2. Instruction Manual Preference:
   Feedback: User-2 found the process straightforward but expressed a preference for a video tutorial over a written manual, as he identified himself as a visual learner.
3. Glass Capillary Design:
   The rubber bulb on top of the dropper has a hole that was not immediately noticeable. Like the testers, User-2 initially struggled, attempting to press the bulb without realizing it needed to be pressed from the top to function.
4. Sample Visibility:
   Challenge: Since water is used to resuspend the swabs, the colorless nature of the sample made it difficult to discern whether the glass capillary was fully filled.
   Recommendation: Add a slight color to the water to improve visibility. This can be achieved using a non-reactive dye that does not alter the solution's pH or a very low concentration (e.g., 0.01 or 0.001 mM) of phenol red, matching the pH of the micro-pads.

5. Glass Capillary Visibility Against Background:
   Issue: The transparent glass capillaries were hard to see against the black tabletop surface.
   Suggestion: Provide a colored sheet as a background to enhance visibility. While black surfaces offer high contrast for transparency, alternative contrasting colors could improve usability.

We calculated the time-to-peak derivative for each sample and compared the performance of the glass capillary and pipette methods. The mean, standard deviation, and median for the glass capillaries were 39.71, 5.30, and 39.10 minutes, respectively, while the corresponding values for the pipette were 38.05, 4.48, and 37.60 minutes. The data for this comparison is visualized in Figure 30. To assess whether the differences between the two methods were statistically significant, we performed the Wilcoxon Signed-Rank Test, a non-parametric test appropriate for paired data that does not assume normality. The test yielded a test statistic of 53.0 and a p-value of 0.284. Since the p-value is greater than the conventional significance level of 0.05, we failed to reject the null hypothesis, indicating no statistically significant difference between the values of glass capillaries and pipette across all samples. However, to further explore the magnitude of the difference, we calculated Cohen's d, which measures the effect size. The Cohen's d value for this comparison was approximately 0.338, indicating a small to medium effect size. While the Wilcoxon test suggested no significant difference, the Cohen's d value implies that the difference between the two methods has practical relevance, despite not being statistically significant. This result is because we observed that users encountered issues such as hand shakiness, causing the sample to splash on the side instead of on the paper pads. Therefore, future work will include designing a guiding rack to assist users in accessing the sample delivery point more accurately and efficiently.

Outcomes and Accomplishments

1. **Environmental Monitoring at Testbed 1 (ASREC)**:
   - Successfully conducted a 10-week proof-of-concept study to characterize the dynamics of *Bacteroidales* contamination near a swine feeding operation.
   - Deployed 120 collection flags, eight weather stations, and air quality monitors to monitor environmental conditions and contamination levels over 10 weeks.
   - Generated detailed heatmaps overlaying *Bacteroidales* concentration data, wind speed/direction, and particulate matter (PM2.5, PM10) distributions for visual analysis.
   - Demonstrated a clear correlation between the distance from the swine unit and *Bacteroidales* concentrations, with contamination decreasing at greater distances.

2. **Predictive Modeling and GUI Development**:
   - Developed a predictive model using Random Forest Regression (RFR) with FFT-transformed weather data, achieving an R^2 of 0.50 (based on the data from stone fruit operations).
   - Identified critical weather predictors (e.g., temperature, dew point, wind speed, solar radiation) influencing contamination patterns.
   - Created an interactive Python-based GUI for visualizing predicted *Bacteroidales* concentrations and environmental data, enabling user-friendly risk assessments.

3. **Field studies in Testbed 2 (stone fruit operations, California) and Testbed 3 (leafy greens operations, Arizona)**:
   - Conducted seasonal studies in California around ~3000-4000-head feedlot and found substantial differences in *Bacteroidales* concentrations between seasons, with higher concentrations observed during warmer months.
   - Identified significant environmental variability between seasons, influencing concentration trends.
   - Observed an inverse relationship between dust count (according to PM2.5 and PM10) and level of *Bacteroidales* in California studies.
   - Determined the levels of *Bacteroidales* around a large concentrated animal feeding operation in Arizona and noticed a dependence on distance as well as direction of observation.

4. **Portable Detection System Development**:
   - Designed and tested the Field-Applicable Sampling Tool (FAST) and heating-imaging platform for LAMP-based detection of *Bacteroidales*.
   - Enhanced usability by transitioning to a glass capillary system for sample delivery, achieving comparable accuracy to traditional micropipettes.
   - Validated field operability with portable setups and user feedback, emphasizing system simplicity and cost-effectiveness.

References

(1) California LGMA. LGMA Metrics (Food Safety Practices), 2021. https://lgmatech.com/wp-content/uploads/2020/08/Current-version_August-2021-CA-LGMA-Metrics_FINAL-v20210818-clean-1262021.pdf (accessed 2022-01-20).
(2) Arizona LGMA. Commodity Specific Food Safety Guidelines for the Production and Harvest of Lettuce and Leafy Greens, 2021. https://www.arizonaleafygreens.org/_files/ugd/cdf4b0_00a96ec5bbcd4975b7e4236145659920.pdf (accessed 2022-01-26).
(3) Nutrition, C. for F. S. and A. FSMA Final Rule on Produce Safety. *FDA* **2021**.
(4) California LGMA. CAFOs: Pre-Season & Pre-Harvest SOPs. https://lgmatech.com/wp-content/uploads/2020/07/SOP_CAFOs_PreSeason_PreHarvest_Assessment_Final_Accessible.pdf (accessed 2022-01-20).
(5) California LGMA. *Pre-Harvest Testing Guidance*. https://lgmatech.com/wp-content/uploads/2021/04/Pre-Harvest-Testing-Guidance-20210416.2.pdf (accessed 2021-11-04).
(6) Wang, J.; Ranjbaran, M.; Ault, A.; Verma, M. S. A Loop-Mediated Isothermal Amplification Assay to Detect Bacteroidales and Assess Risk of Fecal Contamination. *Food Microbiol* **2023**, *110*, 104173. https://doi.org/10.1016/j.fm.2022.104173.
(7) Wang, J.; Ranjbaran, M.; Verma, M. S. Bacteroidales as a Fecal Contamination Indicator in Fresh Produce Industry: A Baseline Measurement. *J Environ Manage* **2024**, *351*, 119641. https://doi.org/10.1016/j.jenvman.2023.119641.
(8) Wang, J.; Kaur, S.; Kayabasi, A.; Ranjbaran, M.; Rath, I.; Benschikovski, I.; Raut, B.; Ra, K.; Rafiq, N.; Verma, M. S. A Portable, Easy-to-Use Paper-Based Biosensor for Rapid in-Field Detection of Fecal Contamination on Fresh Produce Farms. *Biosensors and Bioelectronics* **2024**, *259*, 116374. https://doi.org/10.1016/j.bios.2024.116374.
(9) Belanche-Muñoz, L.; Blanch, A. R. Machine Learning Methods for Microbial Source Tracking. *Environmental Modelling & Software* **2008**, *23* (6), 741–750. https://doi.org/10.1016/j.envsoft.2007.09.013.
(10) Wu, J.; Song, C.; Dubinsky, E. A.; Stewart, J. R. Tracking Major Sources of Water Contamination Using Machine Learning. *Front. Microbiol.* **2021**, *11*, 616692. https://doi.org/10.3389/fmicb.2020.616692.
(11) Bourel, M.; Segura, A. M.; Crisci, C.; López, G.; Sampognaro, L.; Vidal, V.; Kruk, C.; Piccini, C.; Perera, G. Machine Learning Methods for Imbalanced Data Set for Prediction of Faecal Contamination in Beach Waters. *Water Research* **2021**, *202*, 117450. https://doi.org/10.1016/j.watres.2021.117450.
(12) Tanui, C. K.; Benefo, E. O.; Karanth, S.; Pradhan, A. K. A Machine Learning Model for Food Source Attribution of Listeria Monocytogenes. *Pathogens* **2022**, *11* (6), 691. https://doi.org/10.3390/pathogens11060691.
(13) Weller, D.; Belias, A.; Green, H.; Roof, S.; Wiedmann, M. Landscape, Water Quality, and Weather Factors Associated With an Increased Likelihood of Foodborne Pathogen Contamination of New York Streams Used to Source Water for Produce Production. *Front. Sustain. Food Syst.* **2020**, *3*, 124. https://doi.org/10.3389/fsufs.2019.00124.
(14) Weller, D.; Brassill, N.; Rock, C.; Ivanek, R.; Mudrak, E.; Roof, S.; Ganda, E.; Wiedmann, M. Complex Interactions Between Weather, and Microbial and Physicochemical Water Quality Impact the Likelihood of Detecting Foodborne Pathogens in Agricultural Water. *Front. Microbiol.* **2020**, *11*, 134. https://doi.org/10.3389/fmicb.2020.00134.
(15) Bedell, E.; Harmon, O.; Fankhauser, K.; Shivers, Z.; Thomas, E. A Continuous, in-Situ, near-Time Fluorescence Sensor Coupled with a Machine Learning Model for Detection of Fecal Contamination Risk in Drinking Water: Design, Characterization and Field Validation. *Water Research* **2022**, *220*, 118644. https://doi.org/10.1016/j.watres.2022.118644.

(16) Chen, F.; Zhou, B.; Yang, L.; Zhuang, J.; Chen, X. Assessing the Risk of E. Coli Contamination from Manure Application in Chinese Farmland by Integrating Machine Learning and Phydrus. *Environmental Pollution* **2024**, *356*, 124345. https://doi.org/10.1016/j.envpol.2024.124345.
(17) Buyrukoğlu, G.; Buyrukoğlu, S.; Topalcengiz, Z. Comparing Regression Models with Count Data to Artificial Neural Network and Ensemble Models for Prediction of Generic Escherichia Coli Population in Agricultural Ponds Based on Weather Station Measurements. *Microbial Risk Analysis* **2021**, *19*, 100171. https://doi.org/10.1016/j.mran.2021.100171.
(18) Xu, X.; Rothrock, M. J.; Mohan, A.; Kumar, G. D.; Mishra, A. Using Farm Management Practices to Predict Campylobacter Prevalence in Pastured Poultry Farms. *Poultry Science* **2021**, *100* (6), 101122. https://doi.org/10.1016/j.psj.2021.101122.
(19) Pang, H.; McEgan, R.; Mishra, A.; Micallef, S. A.; Pradhan, A. K. Identifying and Modeling Meteorological Risk Factors Associated with Pre-Harvest Contamination of Listeria Species in a Mixed Produce and Dairy Farm. *Food Research International* **2017**, *102*, 355–363. https://doi.org/10.1016/j.foodres.2017.09.029.
(20) scikit-learn. Train_test_split. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html.

APPENDICES

Publications and Presentations

Published:

- Wang, J., Kaur, S., Kayabasi, A., Ranjbaran, M., Rath, I., Benschikovski, I., Raut, B., Ra, K., Rafiq, N., Verma, M.S. (2024). A portable, easy-to-use paper-based biosensor for rapid in-field detection of fecal contamination on fresh produce farms. Biosensors and Bioelectronics, 259, 116374.
- Wang, J., Ranjbaran, M., Verma, M.S. (2024). Bacteroidales as a fecal contamination indicator in fresh produce industry: A baseline measurement. Journal of Environmental Management, 351, 119641.

In Preparation:

- Kaur, S., Sarathy, N., Ahmed B., Ra, K., Kayabasi, A., Wang, J., Puerres, D., Bran, L., Verma, M.S.* (2023). Predicting fecal contamination around feedlots using weather data, distance and direction. *In Preparation*.

Presentations:

- Verma, M.S. (2024) Deciphering One Health microbiomes using field-deployable biosensors. *Auckland Bioengineering Institute Seminar Series* Auckland, New Zealand. DOI: 10.52843/cassyni.fcpkj7
- Verma, M.S. (2024) Field-deployable biosensors for One Health microbiomes. *Front Range Microbiome Symposium 2024* Fort Collins, USA.
- Verma, M.S. (2024) Field-deployable biosensors for rapid characterization of microbiomes. *Pittcon 2024* San Diego, USA.
- Verma, M.S. (2023) Field-deployable Biosensors for One Health: Applications in animals, humans, plants, and the environment. *Purdue Applied Research Institute University Consortium for Health, Food, and Agriculture Resilience* Purdue University, West Lafayette, USA (Virtual).

Budget Summary

This project was awarded $394,516 in research funds; all funds awarded were spent.

Category	**Budgeted $**	**Spent $**	**Spent %**
Personnel	252,256	272,109	108%
Supplies and Expenses	122,868	102,952	84%
Indirect Costs	19,392	19,454	100%
Total	394,516	394,516	100%

Tables and Figures (see below)

TABLES

Table 1: Comparison of MSE and R^2 conditions using different machine learning models for ASREC weeks 1-10 (without tuning)

Machine Learning Model	**Mean Squared Error (MSE)**	**R-Squared (R^2)**
Linear Regression	1.020	-0.094
LASSO Regression	0.937	-0.004
Decision Tree	2.970	-2.183
Random Forest Regression (RFR)	0.857	0.082
Gradient Boosting	1.031	-0.106
Support Vector Regression (SVR)	1.028	-0.102
Neural Network (NN)	1.602	-0.718

Table 2: Significant predictors in the random forest regression model for ASREC weeks 1-10 with tuning

Feature	**Importance**
Temperature * CO_2	0.245
Solar Radiation Values	0.157
Temperature	0.108
log (Wind Speed)	0.089
Wind Speed / Gust	0.077
Gust Values	0.076
Temperature / Humidity	0.074
PM10 Values	0.061
Wind Direction Values	0.058
CO_2 Values	0.056

Table 3: ASREC weeks 1-10 actual versus predicted field sample concentrations after FFT, RFR, and feature engineering with reported statistical methods of comparison

Actual Concentration log_{10}(copies/cm^2)	**Predicted Concentration log_{10}(copies/cm^2)**
0.656	-0.691
0.823	0.994
1.929	0.919
1.033	0.564
1.009	0.759
-0.261	0.546
1.429	1.159
0.775	0.736
1.225	0.868
1.924	1.323
0.193	0.899
0.369	0.733
0.999	0.748
1.100	1.269
1.615	1.416
1.254	1.771
1.371	1.581
2.099	1.269
1.004	0.985
1.658	1.310
0.241	1.068
0.916	-0.284
1.355	1.078
-2.853	0.007
Statistical Methods of Comparison	
Mean Absolute Error (MAE)	0.587
Mean Squared Error (MSE)	0.693
Root Mean Squared Error (RMSE)	0.832
R-Squared (R^2)	0.258
Pearson correlation Coefficient	0.512

Table 4: Average values of weather parameters recorded at each weather station for August 2023.

Temperature (°C)	Humidity (%)	Dew Point (°C Td)	Wind Speed (km/h)	Wind Gust (km/h)	Wind Direction (Â°)	ABS Pressure (Hg)	Solar Radiation (w/m2)	UVI	PM 2.5 (ug/m3)	CO2 (ppm)	PM 10 (ug/m3)	Weather station number
27.03632	64.57853	65.86933	0.889754	1.79197	205.2828	29.55243	169.5269	1.338897	5.64558	358.4944	5.981395	1
26.93348	64.08509	18.6784	1.042142	2.037199	154.5128	29.63118	275.1992	2.26317	4.87445	403.9998	5.165561	2
26.72184	66.44785	66.16871	0.769792	1.859599	319.569	29.66322	211.2786	1.72929	4.6316	410.5398	4.9403	3
27.22867	61.68608	65.07491	4.755913	5.935056	214.336	29.64539	199.9744	1.598515	8.1617	429.743	8.73995	4
27.0862	63.05011	18.58308	5.063246	6.259912	241.1259	29.63794	171.4575	1.326754	8.7292	487.2117	9.8904	5
26.48814	67.00561	66.42902	0.981333	1.903068	362.4261	29.61083	83.43026	0.569231	6.8094	462.2372	7.22414	6
25.99471	70.46484	19.44803	0.872289	1.948507	245.2234	29.62598	109.9166	0.810945	4.7507	425.1144	5.06169	7
26.62737	67.86463	67.24332	4.208397	5.316875	356.0661	29.66255	215.1712	1.753043	7.42087	550.5617	7.63913	8

Table 5: Average values of weather parameters recorded at each weather station for April 2024

Temperature (°C)	Humidity (%)	Dew Point (°C Td)	Wind Speed (km/h)	Wind Gust (km/h)	Wind Direction (Â°)	ABS Pressure (Hg)	Solar Radiation (w/m2)	UVI	PM 2.5 (ug/m3)	CO2 (ppm)	PM 10 (ug/m3)	Weather station number
18.41028	76.96436	13.50837	1.225981	2.180888	173.9788	n/a	169.4443	1.331612	18.96487603	463.5005165	19.19628099	1
18.6065	76.7398	13.65168	1.803046	3.051265	161.1363	29.68905	222.9151	1.825503	21.5464876	375.2174587	22.14411157	3
18.63914	76.54719	13.62811	1.54229	2.72687	139.657	29.73012	198.6903	1.607014	18.86178442	459.6410521	19.15265601	5
18.71006	74.26058	13.23529	4.183282	5.415119	251.2456	29.67956	226.4711	1.851909	21.96795866	475.1555556	22.28268734	6
18.64675	76.38596	13.59226	1.485501	2.676729	160.7446	29.705	130.5776	0.968524	21.50515996	455.5970072	21.874613	7

Table 6: Table highlights the weather parameters with statistical and practical significance. The average values of weather parameters for each weather station were calculated for 2023 and 2024 California data. Statistical analysis comparing these averages was performed using a t-test to compare means and a Mann-Whitney U (MWU) test to account for the small, non-parametric dataset

Weather Parameter	T-Test P-Value	MWU Test P-Value	Delta of Averages
Temperature (°C)	2.66E-12	0.001554002	8.161
Humidity (%)	3.28E-06	0.001554002	10.53125
Dew Point (°C Td)	0.004955002	0.001554002	34.91225
Wind Speed (km/h)	0.760977954	0.523698524	0.2745
Wind Gust (km/h)	0.85669626	0.435120435	0.1705
Wind Direction (°)	0.025708237	0.093240093	84.96475
ABS Pressure (Hg)	0.001693205	0.008246738	0.07375
Solar Radiation (W/m²)	0.725235577	0.832944833	10.125
UVI	0.717163887	0.660111294	0.09425
Hourly Rain (in)	0.373900966	0.268381627	2.20E-05
Event Rain (in)	0.373900966	0.268381627	3.60E-05
Daily Rain (in)	0.041132727	0.37356288	0.004724
PM2.5 (µg/m³)	6.18E-08	0.001554002	14.1925
CO2 (ppm)	0.863241604	0.724164724	4.83775
PM10 (µg/m³)	7.49E-08	0.001554002	14.098

Table 7: Actual vs precited (log_{10}copies/cm^2 of *Bacteriodales*) using the Voting Regression based prediction model developed using California seasonal study

Actual (log_{10}copies/cm^2)	Predicted (log_{10}copies/cm^2)
-0.592	-0.208
-0.325	-0.277
0.278	-0.435
-0.952	-0.401
0.071	-0.858
-0.389	-0.335
-0.528	-0.871
0.003	0.653

-0.376	-0.08
0.165	0.309
-0.524	-0.433
-0.771	-0.884
1.075	1.302
-0.812	-0.778
1.914	1.603
-0.592	-0.83
-1.236	-0.689
-0.871	-0.601
0.62	-0.491
2.328	1.836
-0.208	-0.223
1.26	1.677
-1.069	-0.827
-1.236	-0.131
0.377	-0.372
0.801	0.798
-1.232	0.127
-0.763	0.105
-0.263	-0.063
-1.142	-0.286
-0.997	-0.759
-0.024	1.663
-0.848	-0.762
-1.272	-0.56
-1.169	-0.545
1.156	0.162
-0.096	-0.58
-0.799	-0.511
2.134	1.783
1.787	1.122
2.134	-0.277
-0.759	0.047
-0.321	-0.315
-0.619	-0.284
-0.412	1.515
-0.502	0.248
2.071	1.122

-1.155	-0.845
-1.001	-0.55
-0.925	-0.427
0.17	-0.049
1.864	1.515
-0.948	-0.409
-0.263	-0.079
-0.997	-0.247
0.413	-0.341
2.247	1.73
0.945	-0.121
-0.222	-0.291
0.417	0.984
-0.361	0.299
-0.799	-0.126

Table 8: Top Ten Most Important Features in the Prediction Model. Feature importance was determined by extracting importance scores from the Random Forest model applied on California seasonal data using recursive feature elimination.

Feature	**Importance score**
Temperature	0.283
direction_y	0.164
Humidity	0.154
Distance	0.090
Pressure	0.084
Dew point	0.062
Direction_x	0.054
Wind speed	0.038
CO_2	0.037
PM10	0.033

Table 9: The hierarchical clustering analysis and average values for different parameters in each cluster.

Cluster	**log_{10}(copies/cm^2)**	**CO_2**	**Dew point**	**Direction x**	**Direction y**	**Gust speed**	**Humidity**	**PM10**	**PM2.5**	**Pressure**	**Solar rad**	**Temperature**	**Wind direction**	**Wind speed**
0	0.539	450.295	14.03	-0.194	-0.92	2.702	75.985	17.523	17.237	29.68	195.311	19.266	165.758	1.584
1	1.408	442.576	55.24	0.413	0.881	2.333	67.191	7.348	7.026	29.602	134.556	27.639	191.154	1.381
2	0.549	467.222	13.87	0.439	0.88	3.89	74.681	18.912	18.724	29.661	188.884	19.385	203.015	2.696
3	0.405	447.408	47.845	-0.396	-0.899	2.297	68.165	7.068	6.75	29.611	127.235	27.437	183.949	1.321

Table 10: The K-means clustering analysis and average values for different parameters in each cluster.

Cluster	log_{10}(copies /cm^2)	CO_2	Dew point	Direction x	Direction y	Gust speed	Humidity	PM10	PM2.5	Solar rad	Temperature	Wind direction	Wind speed
1	1.408	442.576	55.24	0.413	0.881	2.333	67.191	7.348	7.026	134.556	27.639	191.154	1.381
2	0.405	447.408	47.845	-0.396	-0.899	2.297	68.165	7.068	6.75	127.235	27.437	183.949	1.321
3	0.549	467.222	13.87	0.439	0.88	3.89	74.681	18.912	18.724	188.884	19.385	203.015	2.696
4	0.539	450.295	14.03	-0.194	-0.92	2.702	75.985	17.523	17.237	195.311	19.266	165.758	1.584

Table 11: The average of each weather parameter from stations 1 and 2 for experimental site in Arizona.

Temperature(℃)	Outdoor Humidity (%)	Dew Point(℃)	Wind(km/h)	Gust(km/h)	Wind Direction(°)	ABS Pressure (hg)	Solar Rad.(w/m^2)	UVI	Hourly Rain(in)	Event Rain(in)	Daily Rain(in)	Weekly Rain(in)	AQIN CO_2(ppm)	AQIN Pm2.5(ug/m^3)	AQIN Pm10(ug/m^3)	Weather station
13.980	41.974	0.020	5.523	6.932	167.753	29.820	85.762	0.607	0	0	0	0.217	568.422	6.070	9.335	1
11.856	48.784	0.341	5.065	6.284	147.973	29.788	85.792	0.594	0	0	0	0.221	383.768	2.305	3.274	2

Table 12: Time threshold values of field test done in ASREC.

Sample	Row number	Glass capillaries time threshold values	Pipette time threshold values	Delta
5000 cps/rxn	Extracted DNA	34.32	36.02	1.7
Temp_Flag1	1	41.33	34.32	7.01
Temp_Flag2	1	38.35	40.58	2.23
Temp_Flag3	1	39.1	39.85	0.75
Temp_Flag4	1	41.33	33.47	7.86
Temp_Flag5	1	42.83	32.62	10.21
5000 cps/rxn	Extracted DNA	36.02	38.35	2.33
Temp_Flag1	2	34.32	38.35	4.03
Temp_Flag2	2	41.33	36.02	5.31
Temp_Flag3	2	36.85	51.65	14.8
Temp_Flag5	2	49.95	42.08	7.87
5000 cps/rxn	Extracted DNA	36.85	37.6	0.75
Temp_Flag1	3	42.83	35.17	7.66
Temp_Flag2	3	30.07	33.47	3.4
Temp_Flag3	3	36.02	40.58	4.56
Temp_Flag4	3	49.95	39.1	10.85
Temp_Flag5	3	43.58	37.6	5.98

FIGURES

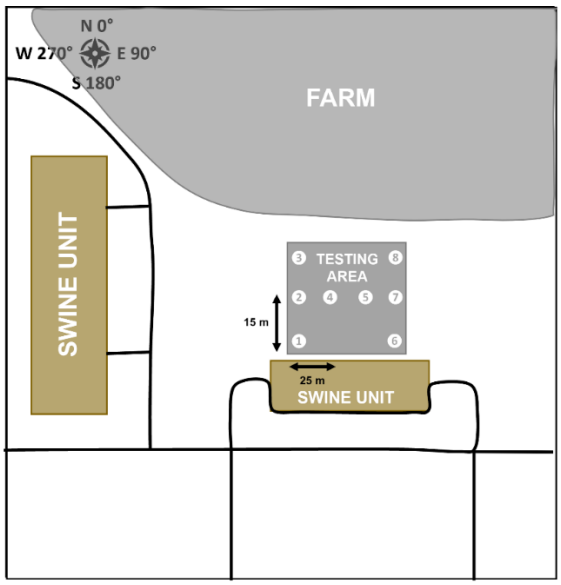

Figure 1: Simplified map of the Animal Sciences Research and Education Center (ASREC). The numbers within white circles represent the weather stations. Flags were arranged in three rows, with 15 meters between rows and 25 meters between flags within each row.

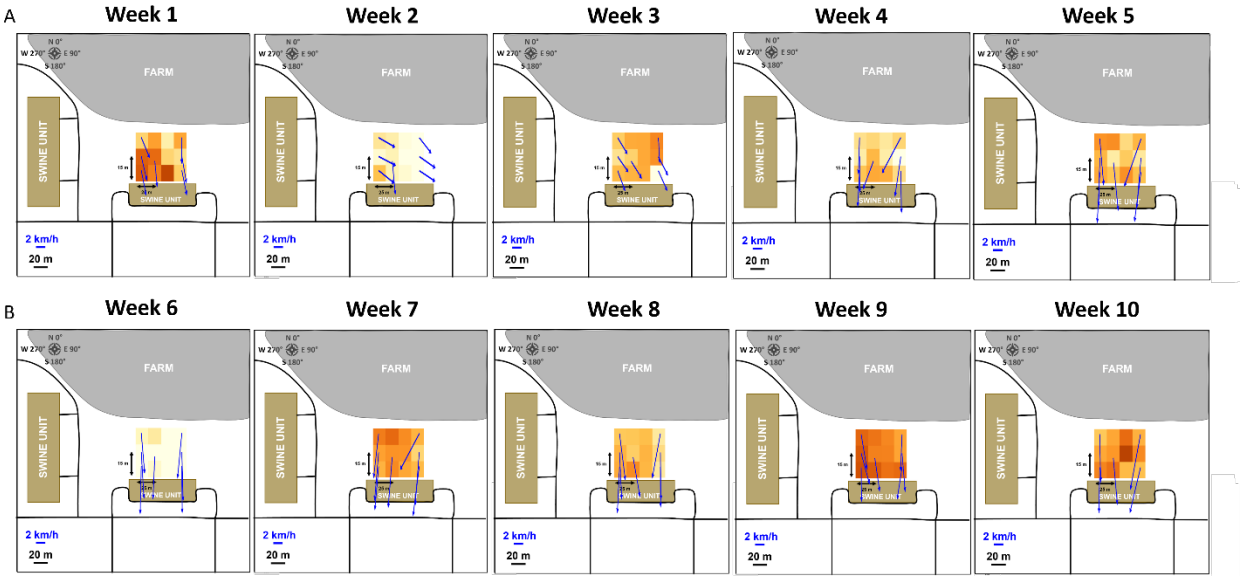

Figure 2: Weekly concentration and wind data for ASREC 10-week study. For each week, the concentration of *Bacteroidales* from each of the 12 collection flags are plotted in a heatmap with a scalebar from 0 to 3 $\log_{10}$(copies/cm^2). Wind speed and direction are displayed by the blue errors for each of the seven weather stations (weather stations 1-4,6,7,8). A) Weeks 1-5 data. B) Weeks 5-10 data.

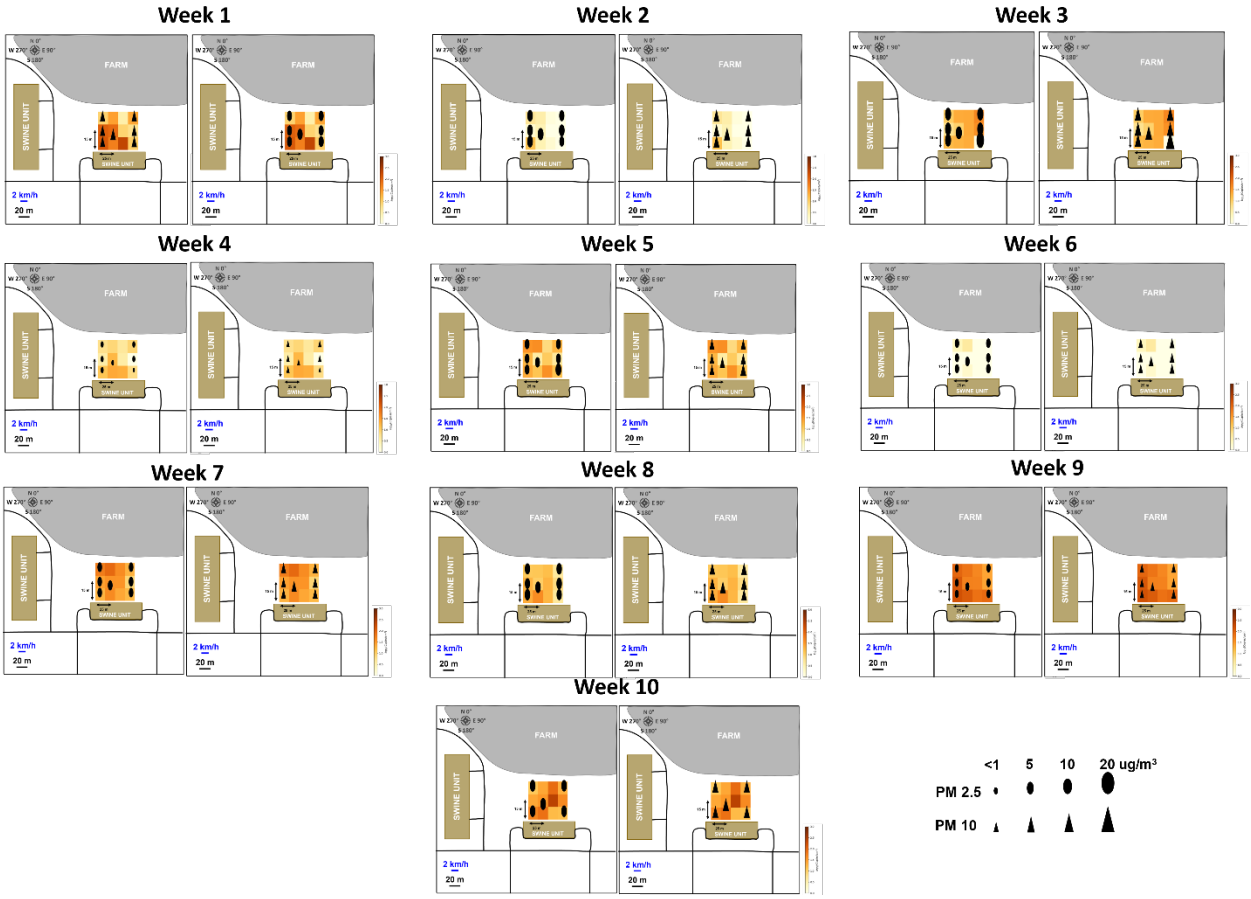

Figure 3: ASREC simplified maps with *Bacteroidales* concentration heatmaps and particulate matter (PM) data. Each week's map displays the size of the PM 2.5 data in black ovals and the PM 10 data in black triangles collected from each of the weather stations.

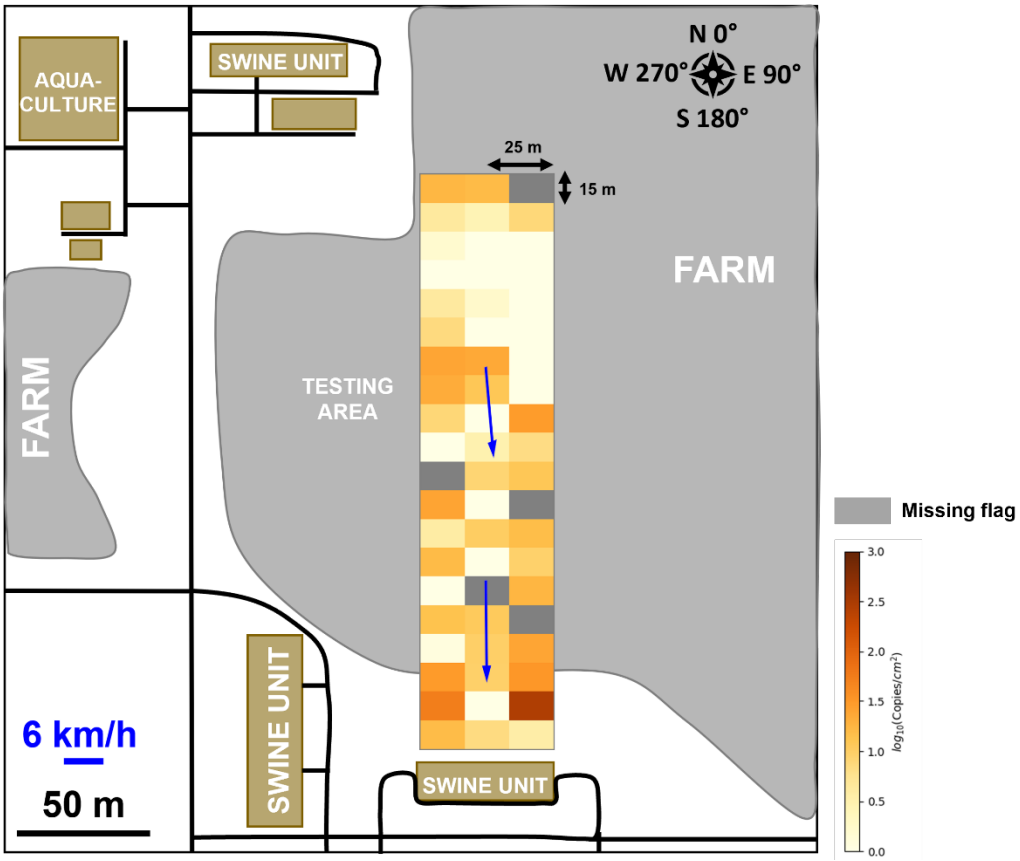

Figure 4: ASREC 300 meters distance study simplified map with *Bacteroidales* concentration heatmap and wind data. Missing collection flags are displayed in gray on the heatmap. Fecal contamination seems to decrease as distance from the high-risk area (swine unit) increases as seen with the decrease in *Bacteroidales* concentration on the collection flags further from the swine unit.

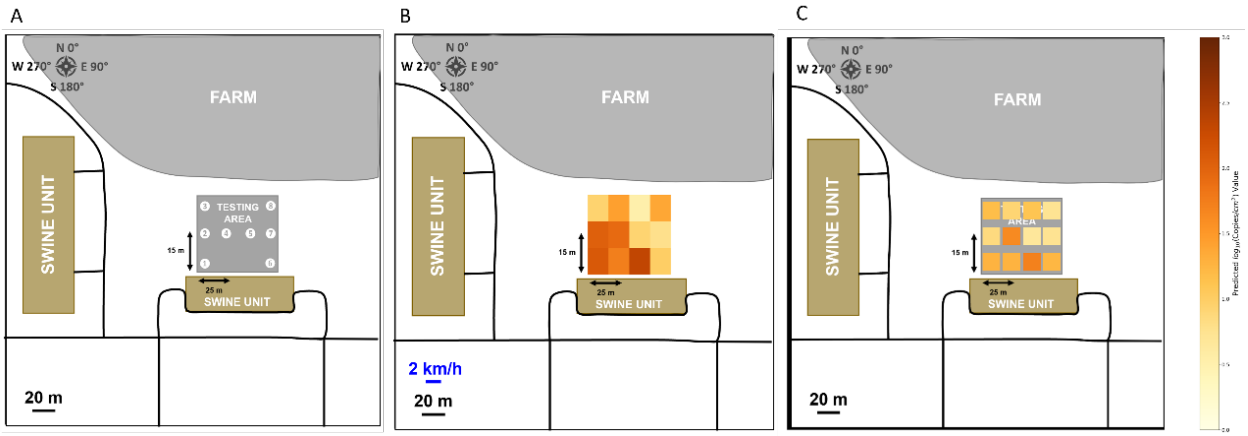

Figure 5: Comparison of actual versus predicted *Bacteroidales* concentrations for ASREC week-1 experiment. A) Original simple map of ASREC study outlining the location of each weather station (1-8). B) Actual field sample concentration with each square representing a flag. C) Predicted field sample concentration using the random forest regression model.

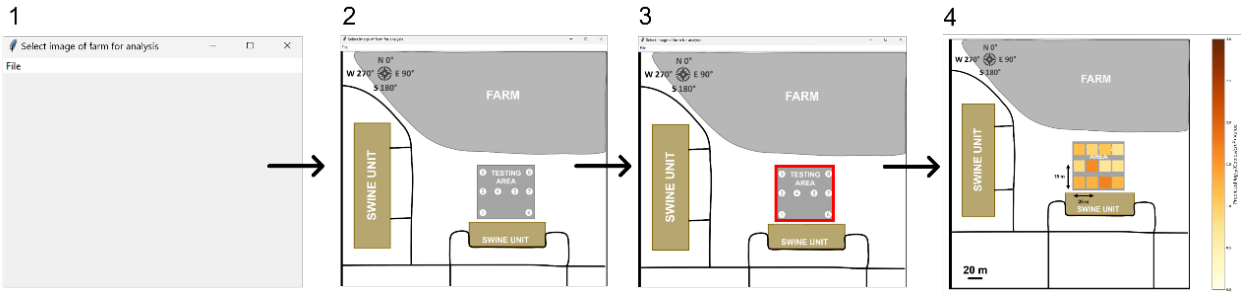

Figure 6: Procedural steps in the user-interface to predict the concentration of *Bacteroidales* in a specified area on the farm. 1) Initiate the program and define weather data, number of flags, and data output format (rows and columns). 2) Import image of the farm for analysis. The program will display a pop-up of the output of the selected image. 3) Draw a box around the area of evaluation. 4) The program will output predicted *Bacteroidales* concentration heatmap for the boxed area.

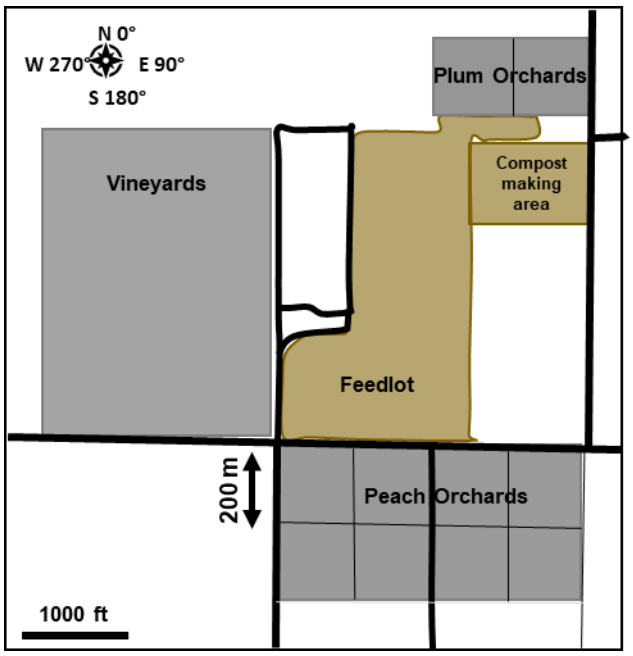

Figure 7: Simplified map for areas tested in stone fruit operations in California (testbed 2). Plum orchards and peach orchards were used for data collection.

Figure 8: Wooden stakes with attached collection flags. Each stake was hammered into the ground, and a collection flag was stapled to it while wearing gloves.

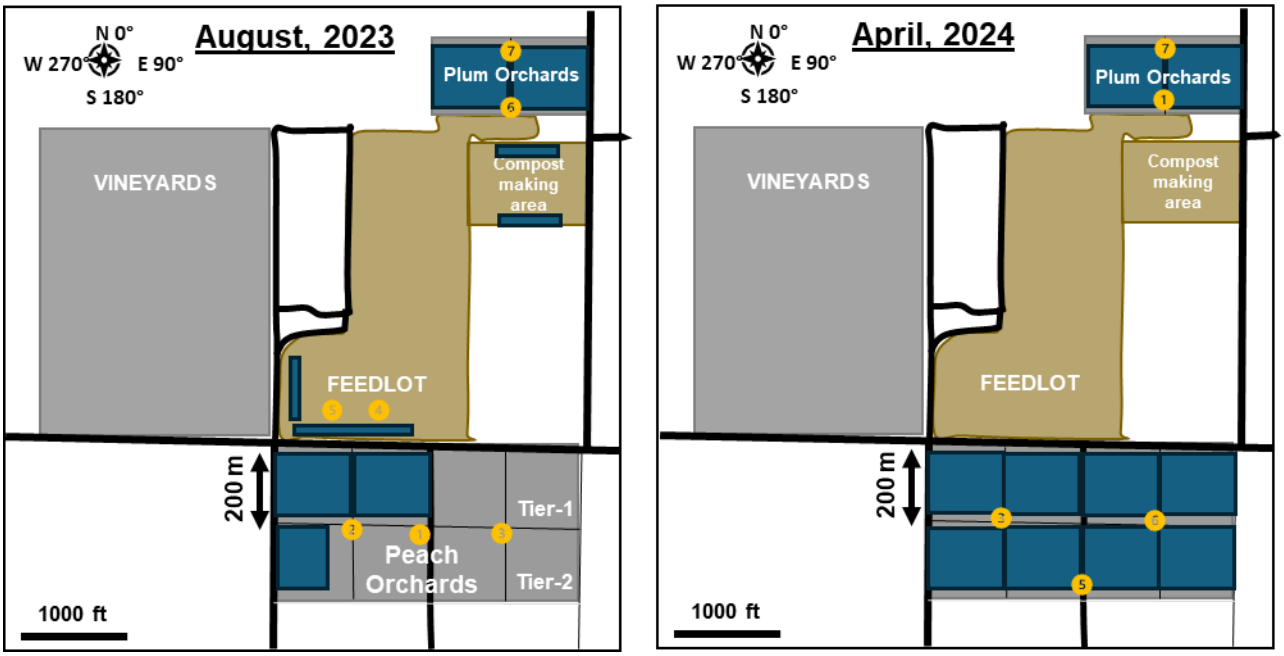

Figure 9: Simplified map of the tested location in California (testbed 2), showing the deployment of field equipment and monitoring stations. The blue areas indicate the locations where collection flags were deployed, while the yellow circles represent the positions of weather stations used for environmental data collection. During the August 2023 study, 390 collection flags were deployed across the site. Eight Ambient WS-5000 weather stations and eight Ambient AQIN air quality monitors were strategically positioned among the plum orchards (north), peach orchards (south), feedlot (central area), and vineyard (west). Similarly, in the April 2024 study, 812 collection flags were deployed along with five weather stations and air quality monitors across the sites.

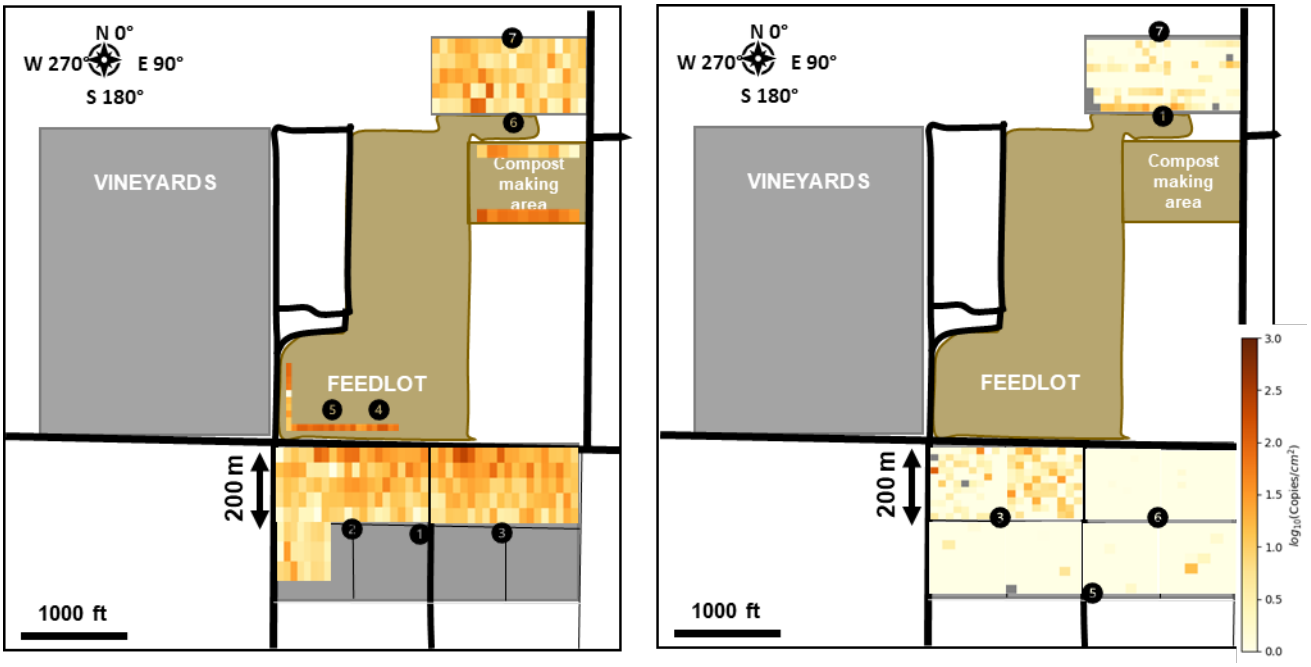

Figure 10: A simplified map along with heatmaps of *Bacteriodales* concentration

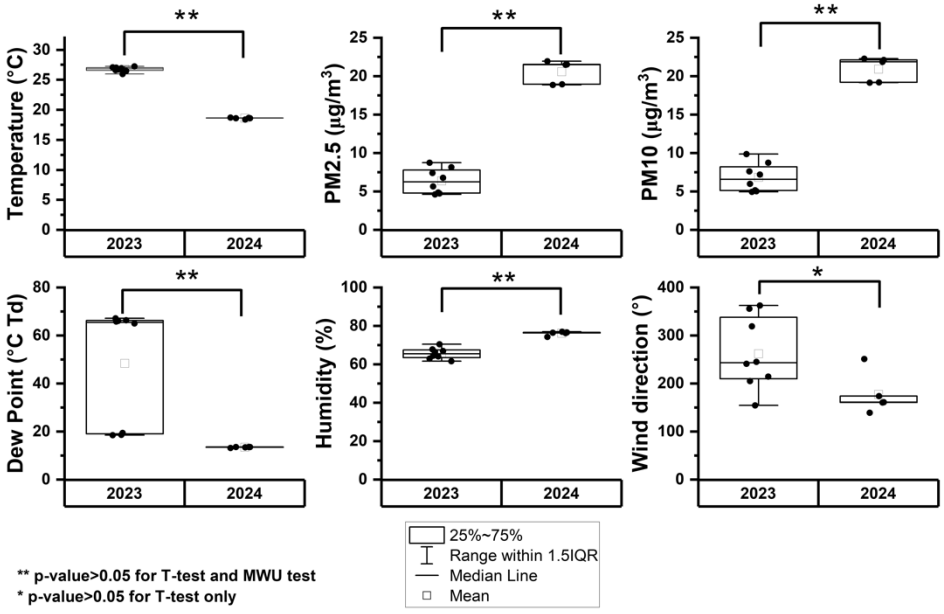

Figure 11: Weather Parameters Showing Significant Differences: Temperature, Humidity, Dew Point, PM2.5, PM10, and Wind Direction. The average values of weather parameters for each weather station were calculated, and statistical analysis was performed to compare these averages. A t-test was used to compare means, and the Mann-Whitney U (MWU) test was applied to account for the small, non-parametric dataset.

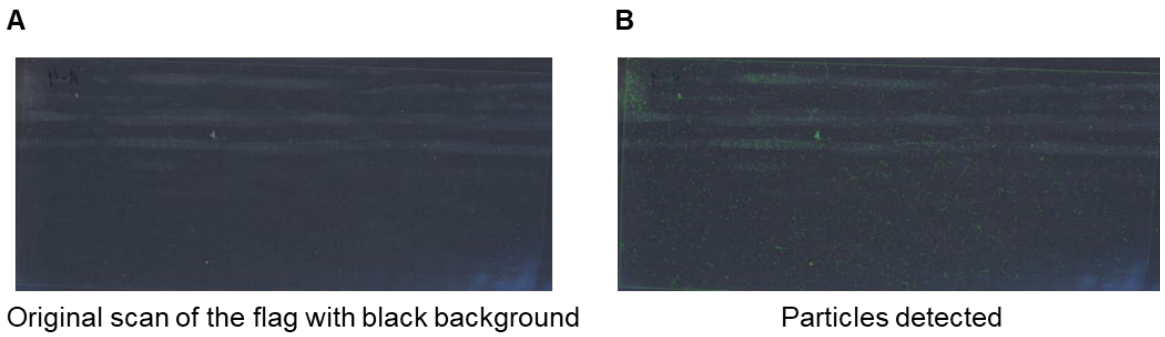

Figure 12: Particle Analysis on 90 Collection Flags from the 2024 Study. Flags with a black background were used to enhance contrast for particle detection. A Python-based program was utilized to count and analyze particles. (A) Scanned flag images with a black sheet as the background. (B) Detected particles highlighted in green.

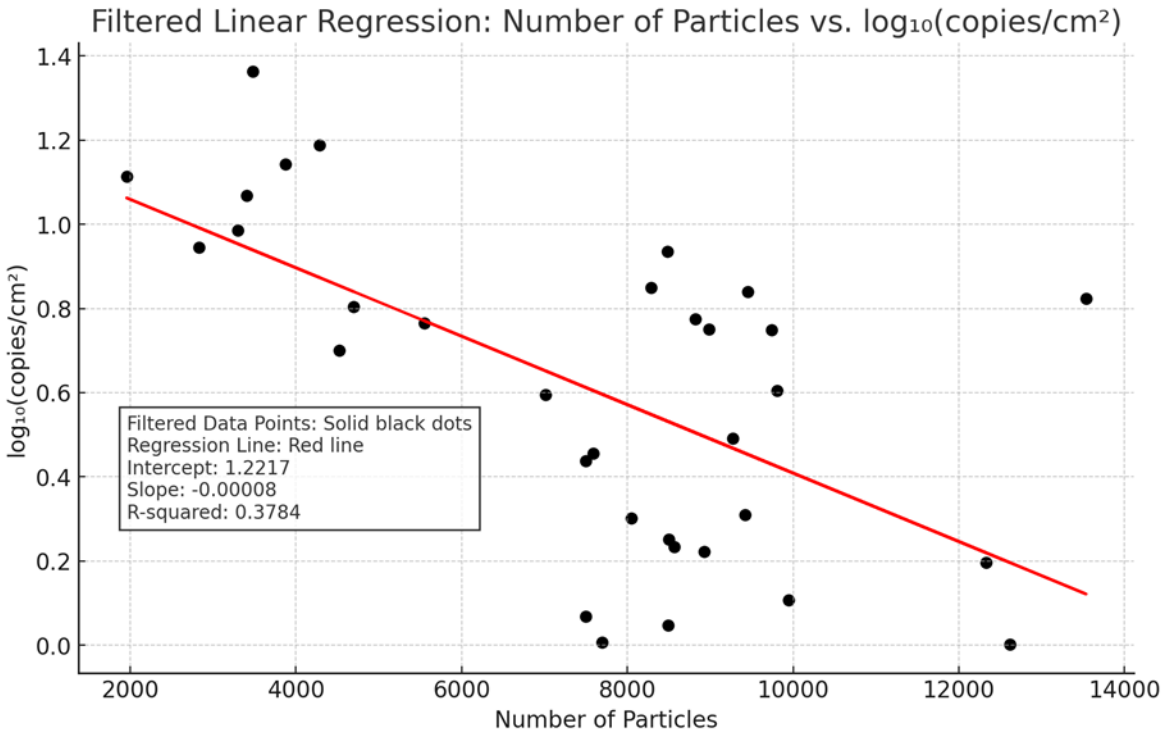

Figure 13: Linear Regression Analysis of the relationship between particle counts and *Bacteroidales* concentrations (log_{10}(copies/cm²). The regression line showed a slight negative trend, with a R^2 value of 0.3784.

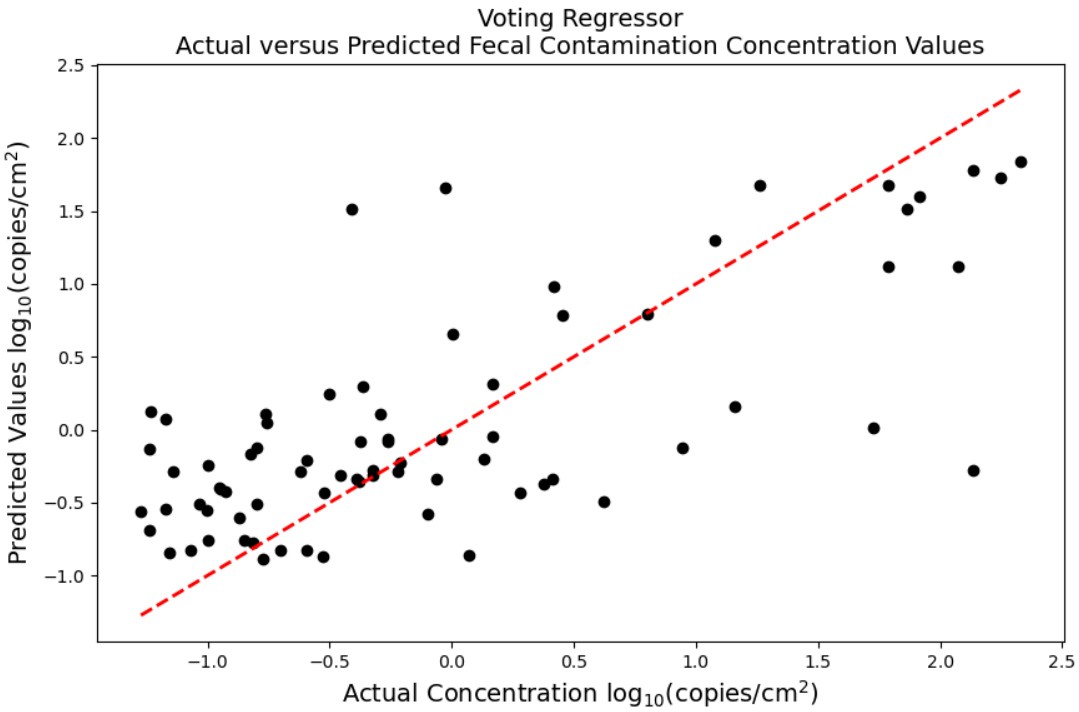

Figure 14: The figure depicts the relationship between predicted and actual $\log_{10}$(copies/cm^2) from the voting regression-based prediction model developed using experiment site 2 seasonal data.

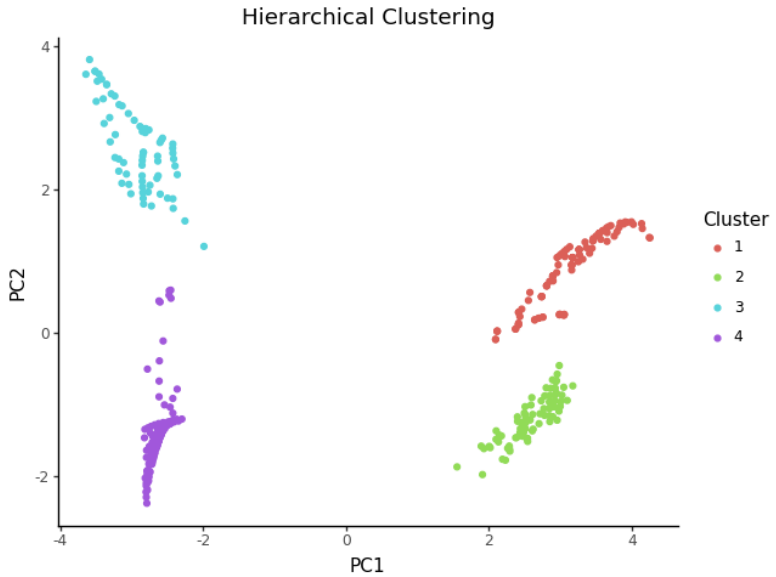

Figure 15: Hierarchical clustering for the preprocessed weather and *Bacteroidales* concentration data captured at experiment site 2

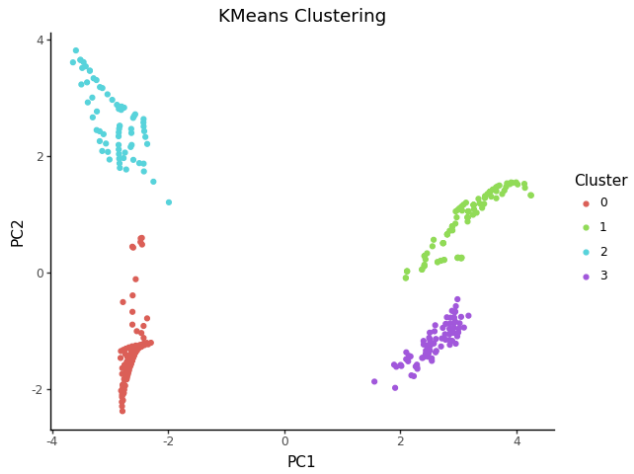

Figure 16: K-means clustering for the preprocessed weather and *Bacteroidales* concentration data captured at experiment site 2

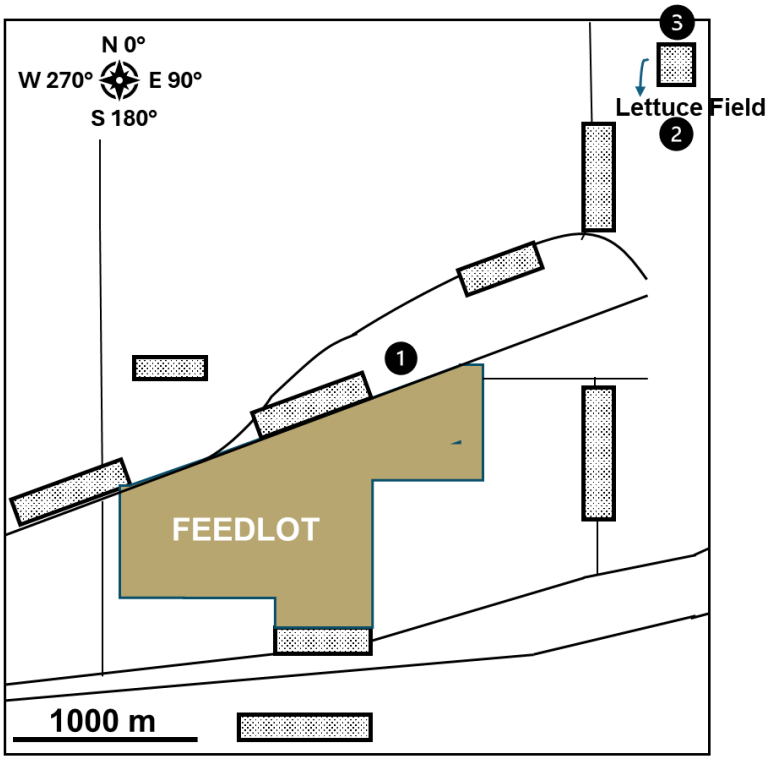

Figure 17: Simplified map for Arizona testing site. The black dots represent the location of weather stations and dotted rectangles are the areas tested for *Bacteriodales* levels using collection flags

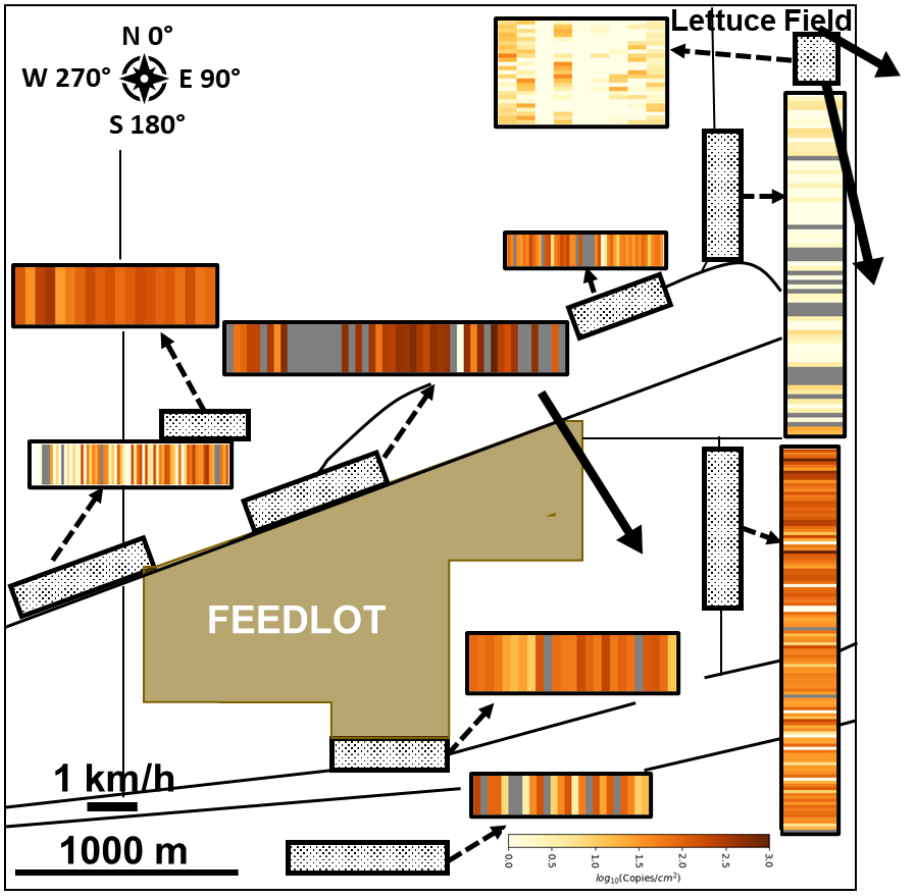

Figure 18: Simplified map for Arizona testing site with heatmaps overlayed on the testing sites with solid arrows indicating wind speed and direction

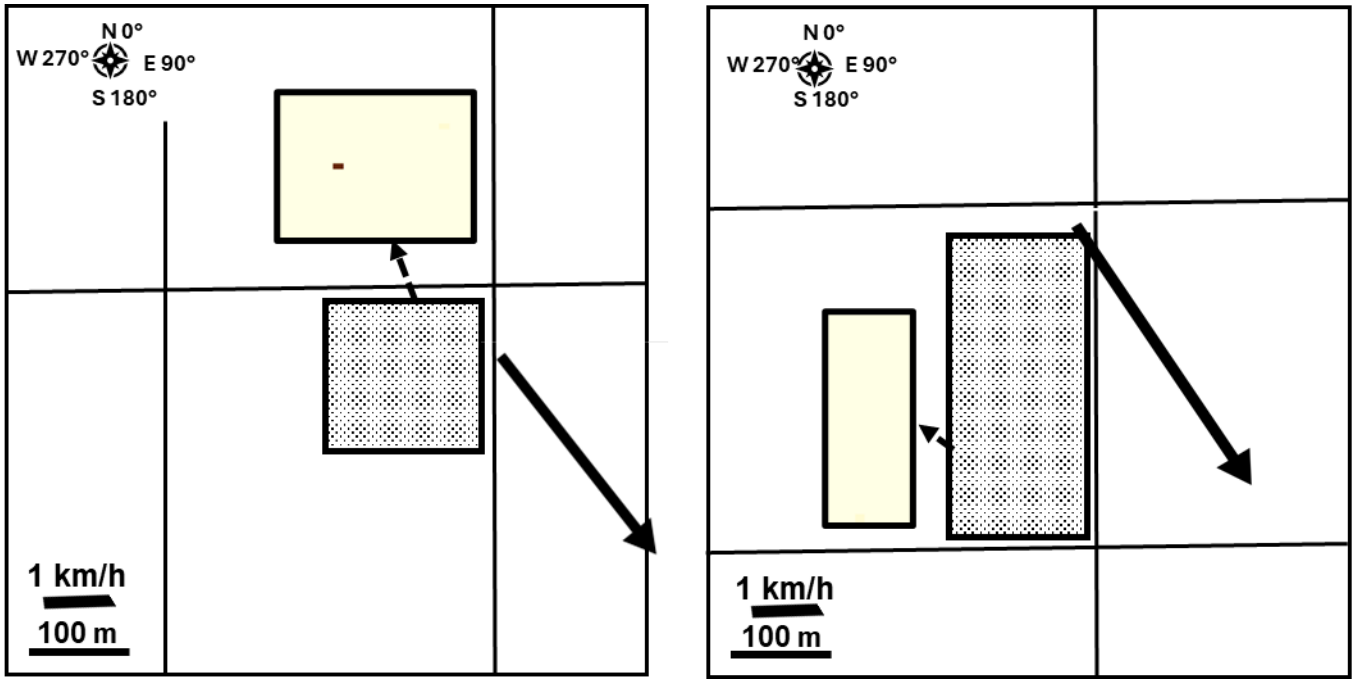

Figure 19: Simplified map for industrial farms, with heatmaps overlayed on the testing sites with solid arrows indicating wind speed and direction

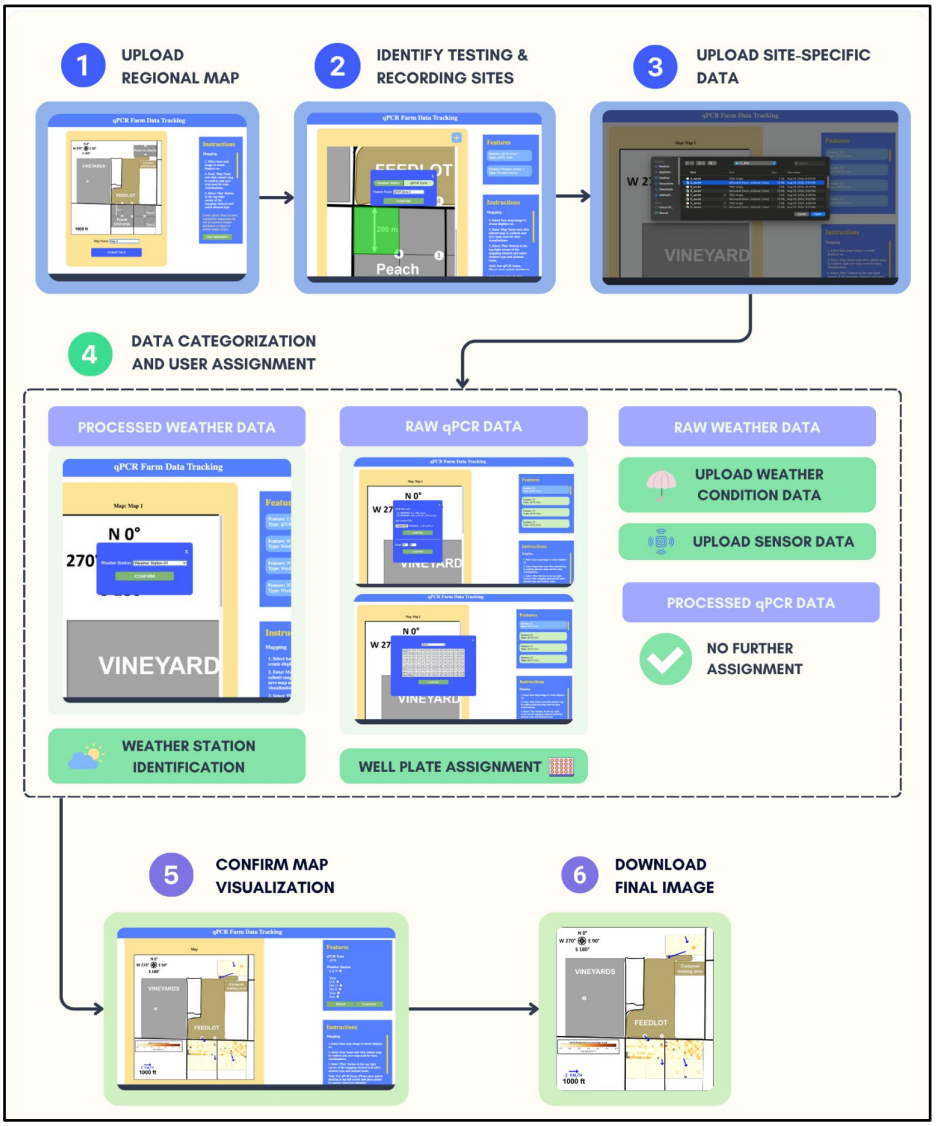

Figure 20: Web platform image annotation workflow. (1) Uploading maps to be annotated. (2) Creation of regions of interest where annotations will appear when data is assigned. (3) Assigns data to each region of interest. (4) Separates data into raw and processed weather and qPCR data to collect additional user inputs as needed. (5) Map Visualization. (6) Final downloadable map.

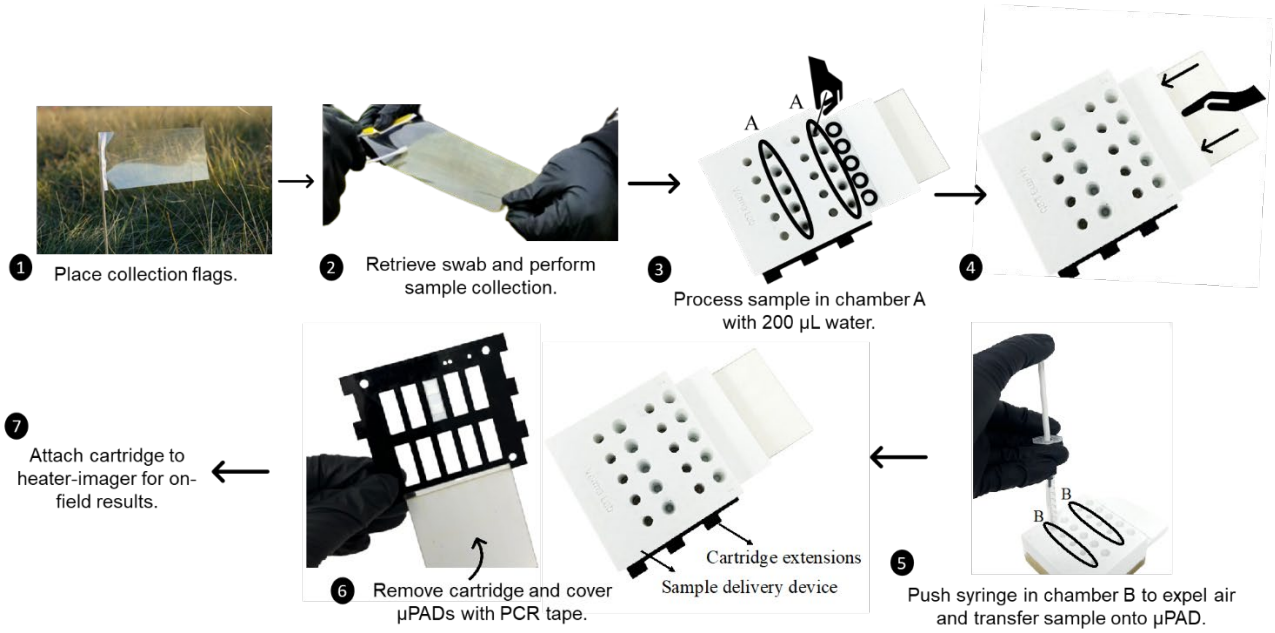

Figure 21: Sample processing steps using Field-Applicable Sampling Tool (FAST). Device dimensions: L = 67.7mm ~ 2.7”, W = 65mm ~ 2.6”, D = 30.8mm ~ 1.2”.

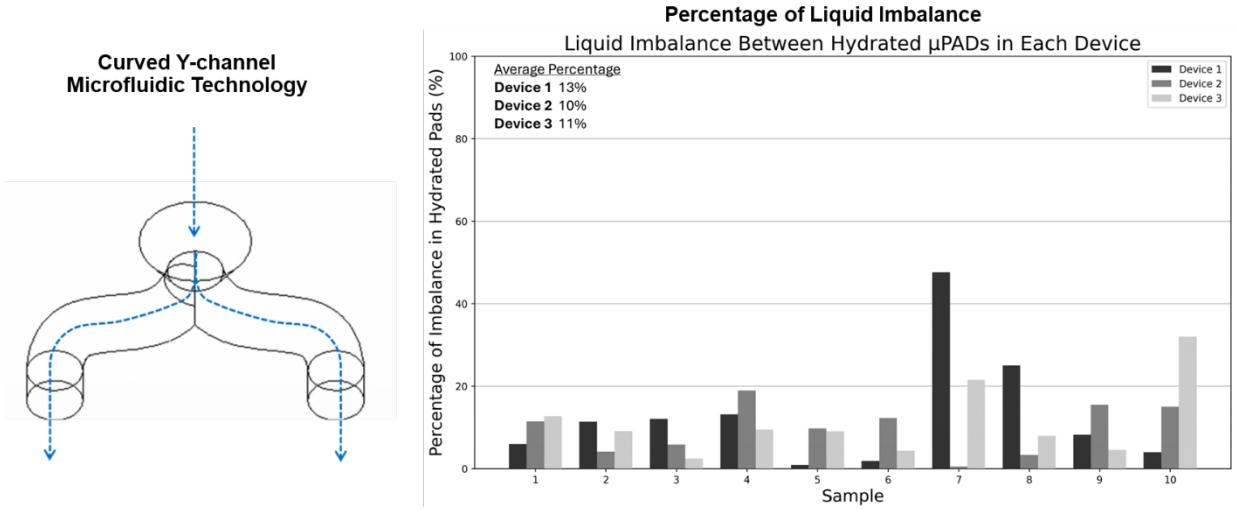

Figure 22: Curved Y-channel microfluidic technology (Left). FAST follows a curved Y-channel formation to allow for equal flow-splitting from the air chamber to the µPADs in the sample cartridge. The blue arrows designate the direction of flow. Liquid imbalance between hydrated µPADs in each device (Right). This graph shows the percentage of imbalance in the set of hydrated pads for each sample, or µPAD, across three different devices. The average percentages of imbalance are low for all three devices (13% for device 1, 10% for device 2, and 11% for device 3).

Figure 23: Our LAMP platform used by three different users in a room in a packaging facility in CA. Users performed different steps of µPAD LAMP using FAST or laboratory micropipettes and our heater-imager platform and compared both the processes and gave their feedback.

Figure 24: Implementation of Field-Applicable Sampling Tool (FAST) a fully integrated paper-based LAMP testing platform on fresh produce farms in CA (April 2024).

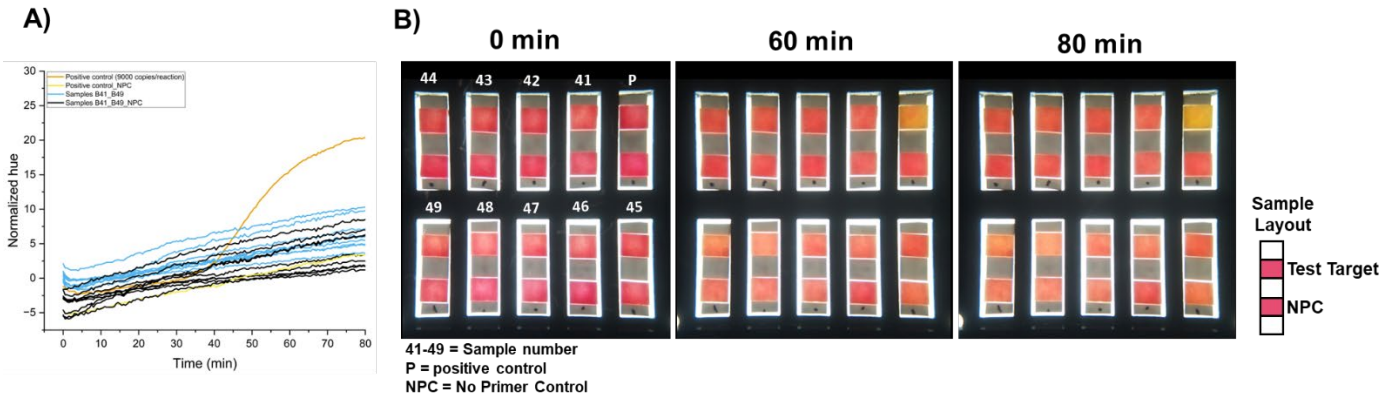

Figure 25: A) Normalized hue plotted against time for the LAMP assay performed on the stone fruit farm. The orange plot is for the positive control, blue for the samples collected on the field and black are the no primer controls. B) Camera pictures from the heater at different times (0, 60 and 80 mins) during the reaction. The sample labels are mentioned over each strip at 0 mins. P denotes the positive controls.

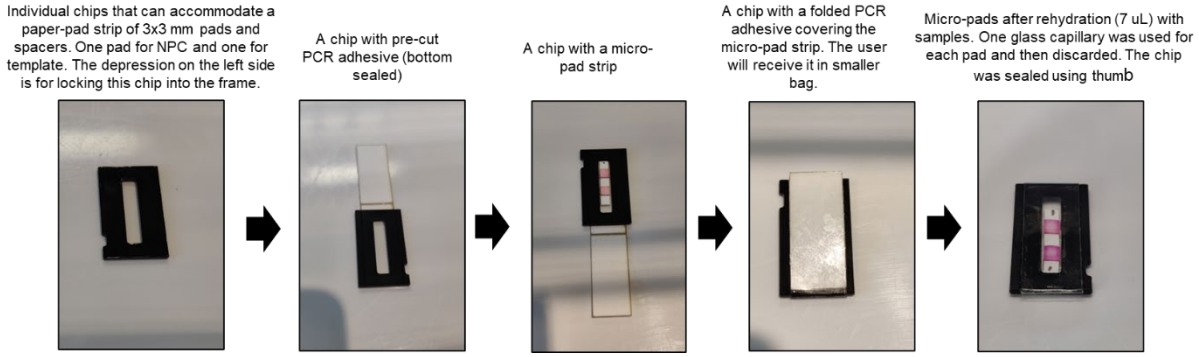

Figure 26: Steps for fabricating LAMP assay chips

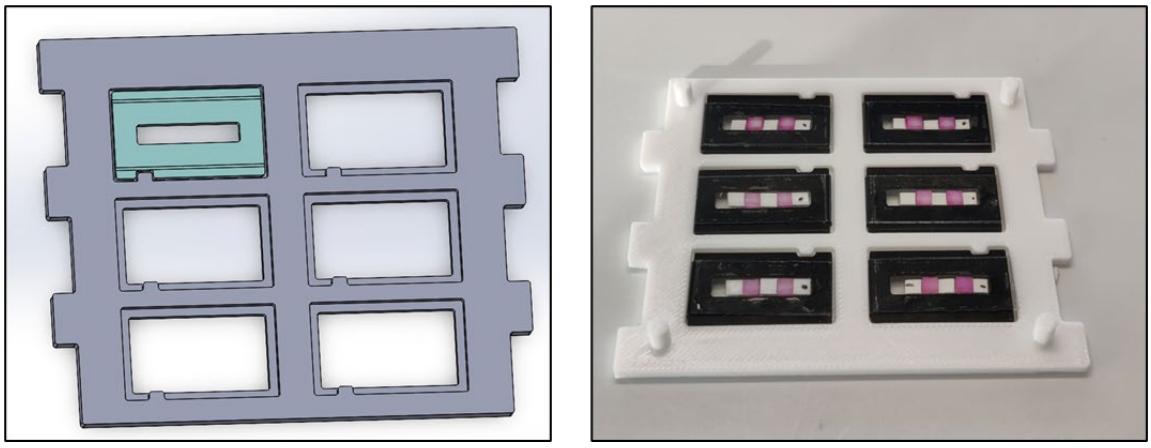

Figure 27: Cartridges for the LAMP assay chips that goes into the holder for heater

Figure 28: Pictures of the users trying the glass capillaries to provide feedback at ASREC, swine unit

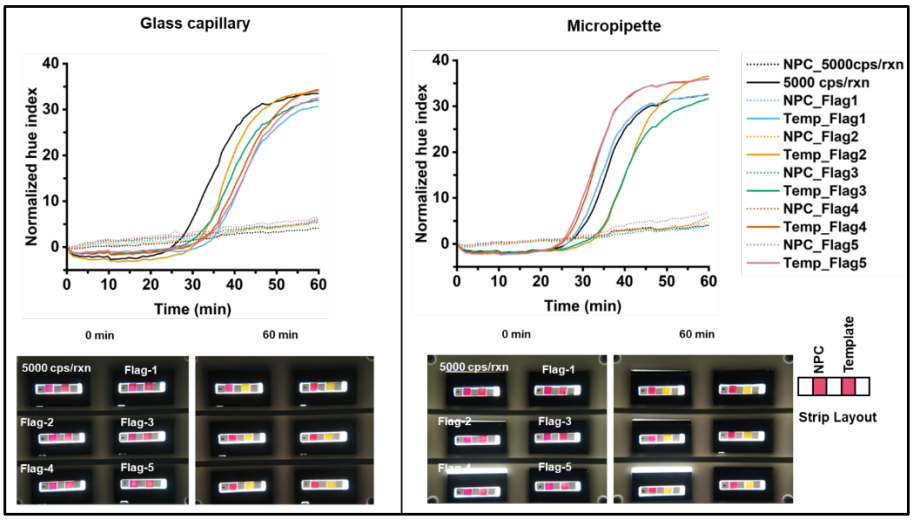

Figure 29: Results are shown for samples collected at ASREC using the glass capillary system (left) and micropipette (right). Normalized hue plotted against time for the LAMP assay, with corresponding images captured by the heater camera at 0 and 60 minutes.

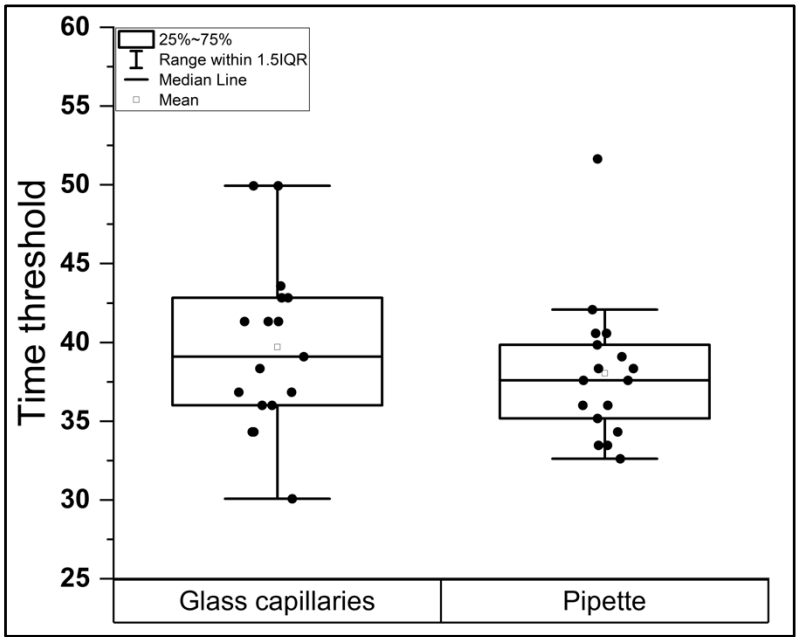

Figure 30: Time threshold values for samples run using glass capillaries and pipette