CENTER for PRODUCE SAFETY

CPS 2017 RFP
FINAL PROJECT REPORT

Project Title
FSMA agricultural-water die-off compliance provisions benefit from condition-specific modifiers

Project Period
January 1, 2018 – December 31, 2019

Principal Investigator
Renata Ivanek
Cornell University
Department of Population Medicine and Diagnostic Sciences
S1-072 Schurman Hall
Ithaca, NY 14853
T: 607-253-4383
E: ri25@cornell.edu

Co-Principal Investigators
Martin Wiedmann
Cornell University
Department of Food Science
Ithaca, NY 14853
T: 607-254-2838
E: mw16@cornell.edu

Trevor Suslow
University California, Davis
Department of Plant Sciences
Davis, CA 95616
T: 530-754-8313
E: tvsuslow@ucdavis.edu

Ana Allende
CEBAS-CSIC Campus de Espinardo
Murcia, E-30100, Spain
T: +34-968-396-377
E: aallende@cebas.csic.es

Daniel Munther
Cleveland State University
Department of Mathematics
Cleveland, OH 44115
T: 216-523-7184
E: D.MUNTHER@csuohio.edu

Objectives

1. *Estimate die-off rates of indicators and attenuated pathogens on baby spinach and baby lettuce in a replicated controlled trial under field conditions in 3 different climatic regions.*
2. *Develop a predictive model of pathogen die-off under relevant environmental conditions and industry practices and use the model to evaluate the FSMA agricultural water matrix.*

Funding for this project provided by the Center for Produce Safety through:
CPS Campaign for Research

FINAL REPORT

Abstract

According to the Food Safety Modernization Act, when non-compliant irrigation water is applied to preharvest produce, a wait period (4-day maximum) can be used between application and harvest to allow for bacterial pathogen die-off, assuming a 0.5 $\log_{10}$ die-off/day. However, there is limited scientific evidence to support this die-off rate. This study aimed to determine the impact of weather on the die-off rate of *E. coli* and *Salmonella* on spinach and lettuce under field conditions. Standardized, replicated field trials were conducted in New York, California, and Spain over two years. Baby spinach and lettuce were grown and inoculated with a ~10^4 CFU/mL cocktail of *E. coli* and attenuated *Salmonella*. Leaf samples were collected at 0, 4, 8, 24, 48, 72, and 96 h following inoculation; *E. coli* and *Salmonella* were enumerated. The associations between study design factors (location, produce type, and bacteria), weather, and die-off were assessed using log-linear and biphasic segmented log-linear regression. A segmented log-linear model best described the net survival on inoculated leaves in most cases, with a greater variation in the segment 1 die-off rate compared to the segment 2 die-off rate; weather was a better predictor of die-off during segment 1 compared to segment 2. Relative humidity appears to be the factor driving segment 1 die-off and the breakpoint, which denotes the time when a typically more rapid die-off rate in segment 1 changes into segment 2. Relative humidity was also identified as a predictor of whether die-off would align with the FSMA die-off rate. Using the results of statistical analysis, we then developed a mathematical model to predict die-off as a function of bacteria, produce type, dew point, and relative humidity. The model has been validated against two independent experimental studies in the Salinas Valley, and the predicted die-off rates appear to show a substantial improvement over those from the current FSMA regulation.

Background

A common route for transfer of pathogenic *E. coli* and *Salmonella* in the preharvest produce environment is surface water, as indicated by several studies (32); surface water has also been identified as a potential cause of several outbreaks (5, 6, 7). Surface water can be applied to preharvest produce as irrigation water, through use in agrichemical applications, and frost protection, among others. As such, the FDA proposed an agricultural water standard to define compliant use of surface water applied to the harvestable portion of preharvest produce as part of the Food Safety Modernization Act (FSMA; 35). The standard states that, for any water source applied to the harvestable portions of produce, 20 samples must be collected over a two to four-year period prior to application and tested for generic *E. coli* level. The geometric mean *E. coli* level must be <126 CFU/100 mL and the statistical threshold value (i.e. the 90th percentile) must be <410 CFU/100 mL for those samples (8). If the water does not meet this standard, growers can choose to not use the water source, treat the water prior to use with allowed methods, or wait up to 4 days from water application to harvest to achieve sufficiently protective microbial die-off to a level compliant with the agricultural water standard, which assumes a 0.5 $\log_{10}$ die-off per day (9). However, the die-off rate within the agricultural water standard has been challenged due to lack of agreement between scientific studies (39).

Many studies have investigated the survival and die-off of attenuated Shiga toxin–producing *E. coli* (STEC) and non-pathogenic *E. coli* (2, 4, 10, 11, 14, 15, 16, 17, 19, 20, 22, 25, 26, 27, 29, 31, 40, 41) and *Salmonella* (19, 23, 25) on in-field leafy greens. For instance, Moyne et al. (27) irrigated in-field lettuce with irrigation water contaminated with attenuated *E. coli* O157:H7 in three replicated trials in Salinas, California, and observed die-off of between 2 and 3 $\log_{10}$ within 2 hours of inoculation. On the other hand, Chase et al. (11), inoculated in-field lettuce in two replicated trials in Salinas, California, with the same strains of attenuated *E. coli* O157:H7 and observed a net die-off of 2.6 and 3.2 $\log_{10}$ over 10 days, indicating substantial variation in die-off observed from study to study. As such, additional research is needed to better understand the drivers (e.g., weather) of this variability in foodborne pathogen population dynamics on produce in the preharvest environment. Additionally, only a small number of studies have investigated die-off in different climatic regions using a standardized protocol (4, 16). Without multiple climatic regions there is less weather variability between trials, making it difficult to identify true associations between weather and die-off. Furthermore, several studies indicate a log-linear die-off, as is suggested in the FSMA agricultural water rule, may not be appropriate to model in-field pathogen die-off, which can lead to an over- or under-estimation of the actual net die-off (26, 40). While producers may benefit from reduced preharvest waiting time if over-estimated for local conditions, the more serious consequence, understandably, is under-estimation. Lastly, as the majority of studies are conducted on *E. coli* die-off on lettuce, more information is needed on the difference between die-off of foodborne pathogens, indicator organisms, and surrogate organisms used to conduct such field studies; additional information on the variations in die-off on different leafy greens varieties is also needed.

As such, the current study aims to contribute to filling these knowledge gaps and develop a better understanding of the population dynamics of surrogate organisms across three climatic regions to investigate the impact of weather on net viable microbial recovery over the proposed die-off interval. In particular, the objectives of this study were to (i) quantify and compare the die-off rates and die-off patterns of *E. coli* and attenuated *Salmonella* on in-field baby spinach and lettuce in replicated controlled trials in Davis (California), Freeville (New York), and Murcia, Spain; (ii) identify weather factors associated with the die-off rate and die-off pattern; and (iii) use the identified associations between study design factors, weather, and die-off to develop predictive models of microbial die-off under varying weather conditions. These results can be used by industry to inform food safety programs, by academia to develop risk assessment models and prioritize future research, and by government to update regulations.

Research Methods and Results

Objective 1 Methods:

Field set-up: Replicated controlled field trials were conducted in three locations: Davis, California (University of CA, Plant Sciences Field Research Facility); Freeville, New York (Homer C. Thompson Research Farm); and Murcia, Spain (La Matanza Research Farm). In each trial, 6 rows of lettuce or spinach seed were sown into plots approx. 1.5 m wide by 4 m long; the seeding rate was between 1.25 cm and 4 cm; lettuce and spinach were sown in separate beds. Due to uncontrollable factors (e.g., animal intrusion, poor stand germination, and adverse weather), the number of plots, spinach and lettuce varieties, and time between planting and harvest varied between trials (Table 1). For instance, in trial NY3 two varieties of spinach were compared: one variety used in the other two location trials and a variety well suited for New York conditions. The new variety was used in the remaining trials in New York (i.e., NY1, NY2, and NY4) due to the poor stand germination of the variety used in CA and Spain.

Testing for naturally occurring rifampicin-resistant *E. coli* and attenuated *Salmonella*: The following control samples were collected to test for naturally occurring rifampicin-resistant *E. coli* and *Salmonella* prior to each trial: soil samples from each experimental plot, one spinach sample, one lettuce sample, and 1 L of water used for irrigation. For the produce and soil samples, 25 g was weighed out in separate sterile bags (Nasco, Fort Atkinson, WI). The samples were then diluted 1:10 with tryptic soy broth (TSB; Becton Dickson, Franklin Lakes, NJ) supplemented with 0.1 g/L rifampicin (MilliporeSigma, Burlington, MA). The 1-L water sample was run through 0.45-µm NEO-GRID filters (Neogen Corp., Lansing, MI). The filters were then transferred to a sterile bag and 90 mL of TSB supplemented with 0.1 g/L rifampicin was added. All control samples were incubated at 37°C for 18 to 24 h. After incubation, 50 µL of each enriched sample was streaked onto an *E. coli* ChromAgar (DRG International, Springfield, NJ) plate supplemented with 0.1 g/L rifampicin (ECC+R) and onto a *Salmonella* ChromAgar (DRG International) plate supplemented with 0.1 g/L rifampicin (SC+R). The plates were incubated at 37°C for 18 to 24 h. The presence of blue colonies on ECC+R plates indicates a positive result for rifampicin-resistant *E. coli*, and the presence of mauve colonies on SC+R plates indicates a positive result for rifampicin-resistant *Salmonella*.

Inoculum preparation and inoculation: Spinach and lettuce plots were inoculated with three strains of rifampicin-resistant *E. coli* (TVS 353, TVS 354, and TVS 355), provided by the Suslow lab at the University of California, Davis (34), and two strains of rifampicin-resistant attenuated *Salmonella* [*Salmonella enterica* sv. Typhimurium strain MHM112 (13) and *Salmonella enterica* sv. Typhimurium UK-χ3985 (12)]. The rifampicin resistance in the two *Salmonella* strains was developed by the Suslow lab using the procedure described by Lopez-Velasco et al. (24); the rifampicin-resistant strain derived from MHM112 is named attPTVS 355 and the rifampicin-resistant strain derived from UK-χ3985 is named attPTVS 337. Each strain was streaked onto separate tryptic soy agar plates (Becton Dickson) supplemented with 0.1 g/L rifampicin to form a confluent lawn. The plates were incubated at 37°C for 18 to 24 h. After incubation, each plate was flooded with 5 mL of phosphate buffered saline (PBS), and the cells were suspended using a sterile loop or L-spreader. Each strain's cell suspension was then transferred to a separate sterile bottle. If there were visible cell masses remaining on the plates, the washing step was repeated. Once there were no longer visible cell masses on the plates, enough PBS was added to each sterile bottle containing the cell suspensions to reach final volumes of 100 mL. Each bottle was vortexed, and 10 mL was transferred to a separate sterile 15-mL tube. The cell suspensions for each strain were spun down at 2,000 x g for 10 min in New York, or 2,500 x g for 5 min in Spain and California; the differences in centrifuge conditions were due to differences in the equipment available in each lab. Following centrifugation, the supernatant was pipetted off and the cells were washed twice in PBS using the same centrifuge conditions listed above. The cells were then re-suspended in 6 mL of PBS, and the optical density at 600 nm was measured to confirm the suspensions were at approx. 9 $\log_{10}$ CFU/mL. The cell suspensions were stored at 4°C overnight. The next morning, the suspensions were mixed and diluted to a final concentration of approx. 5 $\log_{10}$ CFU/mL of each bacterium in sanitized 2-L bottles. A 5-mL aliquot from each bottle was transferred to separate sterile 15-mL tubes and stored at 4°C to be used to confirm the concentration of the inoculum.

Confirmation of the inoculum concentration was performed within 24 h of inoculum preparation. The aliquot from each bottle was diluted by 10^{-1} and 10^{-2}, and 100 µL of each dilution was plated on ECC+R. The plates were incubated at 37°C for 18 to 24 h. After incubation, the number of blue colonies (*E. coli*) and white colonies (*Salmonella*) were counted and recorded.

Inoculation was targeted to be performed when the lettuce plants developed 6 true leaves; spinach plants were expected to have 10 to 12 true leaves at this time (approx. 30–40

days after seeding). However, due to weather conditions, some trials could not be conducted at this targeted time (Table 1). Inoculation was performed using a CO_2 powered backpack sprayer with the pressure between 27 and 30 psi and two Turbo TeeJet (tip #8) nozzles spaced 38 inches apart. The inoculum was applied to the plots at 2 L per approx. 45 m^2.

Sample collection: After inoculation, samples were to be collected at the following time points (in hours, h): 0, 4, 8, 24, 48, 72, and 96 h. For trials CAp (preliminary trial in California), SP2 spinach, and SP4 spinach, samples were collected at a reduced number of time points due to crop loss (Table 1); these trials or produce/trial combinations were excluded from all data analyses. At each time point, 5 samples were collected per plot, with the exception of trials CAp and SP4 (Table 1). Each sample consisted of 6 adjacent plants from a single row; locations of these samples were randomly selected. No samples were harvested from the outer rows of plots.

Microbial testing: For samples collected at 0, 4, or 8 h past inoculation, the samples were diluted 1:5 with PBS. For samples collected at 24, 48, 72, or 96 h past inoculation, the samples were diluted 1:10 with PBS. All samples were massaged by hand for 1 min. All samples were enumerated for rifampicin-resistant *E. coli* and *Salmonella*. Based on the discretion of each location, between 10 µL and 100 mL were plated for each bacterium per sample to increase the likelihood of being within the countable range. For plating volumes less than or equal to 250 µL, the samples were spread plated on ECC+R and incubated at 37°C for 18 to 24 h. The blue colonies and white colonies were counted and recorded as *E. coli* and *Salmonella*, respectively. For plating volumes of 1 mL or greater, the samples were filtered through 0.45-µm pore-size NEO-GRID units. When using the filters, *Salmonella* can no longer be reliably counted on ECC+R plates, so filtering of each sample was performed in duplicate; the first filter was aseptically transferred to an ECC+R plate and the second filter was aseptically transferred to a SC+R plate. The ECC+R plates were incubated at 37°C for 18 to 24 h and the SC+R plates were incubated at 37°C for 42 to 48 h. The number of characteristic colonies per plate was converted to CFU/100 g of produce, which is referred to as the "population level" of *Salmonella* or *E. coli* present on the sample; all counts were reported in Log_{10} CFU/100 g of produce to avoid negative Log_{10} counts.

Enrichment: For any sample negative for either *E. coli* or *Salmonella* by enumeration, enrichment was performed. To do so, the sample was diluted 1:2 with 2X TSB supplemented with 200 g/L of rifampicin (2X TSB+R), based on the volume of PBS remaining in the sample (e.g., if there was 50 mL of PBS, 50 mL of 2X TSB+R was added to the sample bag). The enrichments were then incubated at 37°C for 18 to 24 h. Following incubation, 50 µL of the enrichment was streaked onto an ECC+R plate to test for the presence of *E. coli* and/or onto a SC+R plate to test for the presence of *Salmonella*. All plates were incubated at 37°C for 18 to 24 h.

Strain identification: A PCR protocol was developed to differentiate the *E. coli* strains to determine if there was a difference in strain survival. The protocol utilized 4 primer sets: 1 specific for each strain and 1 that amplified all 3 strains (Table 2). Fifty µL reactions were performed, with the reagents at the following final concentrations: 1X Green GoTaq Flexi Reaction Buffer (Promega, Madison, WI), 1.5 mM $MgCl_2$ (Promega), 0.2 mM of each dNTP (Thermo Scientific, Waltham, MA), 0.2 µM 353F, 0.2 µM 353R, 0.2 µM 354F, 0.2 µM 354R, 0.2 µM 355F, 0.2 µM 355R, 0.2 µM 35XF, 0.2 µM 35XR (Integrate DNA Technologies, Coralville, IA), 1.25 U GoTaq Polymerase (Promega), <0.5 µg/50 mL DNA template. The thermocycler conditions were as follows: 94°C for 2 minutes; 30 cycles of 94°C for 30 seconds, 57°C for 1 minute, 72°C for 1.5 minutes; 72°C for 7 minutes, 4°C for ∞. The primers amplified a 240 bp fragment from TVS 353, a 521 bp fragment from TVS 354, and a 960 bp fragment from TVS 355,

and 1835 bp fragment with the universal primers. In New York, PCR was performed on up to six isolates for each *E. coli*–positive sample (by either enumeration or enrichment) for each trial.

To determine differential survival of the two attenuated *Salmonella* inoculum strains, *Salmonella* colonies were streaked on Xylose Lysine Deoxycholate agar supplemented with 0.1 g/L rifampicin (XLD+R). The XLD+R plates were incubated at 35°C for 18 to 24 h and the resulting color of the colonies was recorded. PTVS 337 does not produce H_2S and forms pink colonies on XLD+R, while PTVS 355 does produce H_2S and produces black colonies on XLD+R. In trial NY4, 40 characteristic *Salmonella* colonies per plot were tested from two time points (24 and 96 h).

E. coli and *Salmonella* strain identification were also performed in California and Spain. The analysis of this data is still in progress, and as such, is not included here.

Monitoring environmental conditions: The following weather conditions were recorded in all locations: temperature (°C), relative humidity (%), solar radiation (kilowatts per meter square, kW/m^2), precipitation (mm), and wind speed (m/s). The weather station is located at 38.53, -121.79 (elevation: 18.3 m) in California; 42.52, -76.22 (elevation: 335.3 m) in New York; and 38.11, -1.03 (elevation: 135 m) in Spain. All weather data was cleaned in R version 3.5.3 (R Core Team, Vienna, Austria). Hourly dew point was also calculated using the humidity.to.dewpoint function in the "weathermetrics" package (1). The final weather variables are listed in Table 3. All weather variables were calculated over the 96h following inoculation.

Statistical analyses: All data cleaning and analyses were performed in R version 3.5.3. The population levels of some samples were out of the countable range. As such, multiple imputations were performed to select population levels for these samples (18). To do so, the lower and upper limits of detection were calculated for each of these samples and a population level between those two values was randomly selected (i.e., imputed) using a uniform distribution; ten imputation rounds were performed per sample. For samples above the countable range, the upper limit was set at 10^7 CFU/100 g of produce (i.e., ~1 $\log_{10}$ above the highest counted population level with filter plating). The raw and imputed data were compiled into a "raw imputed" dataset.

The "raw imputed" sample-level data was subset by each combination of plot and bacteria (Figure 1). For instance, an individual data set was created for *E. coli* on the first spinach plot in trial NY3. A biphasic segmented log-linear and a log-linear regression model were fit using the lm function for each of the ten imputed data sets for each subset (i.e., for each plot-bacteria combination, where plot could be spinach or lettuce). These are referred to as plots. The breakpoint in each of the segmented models was identified by selecting the breakpoint that minimized the deviance of the model. Additional biphasic models (e.g., Weibull) were not evaluated due to the difficulty with practical interpretation. The ten models for the same plot were then combined using the pool function in the "mice" package (36) into a single model per plot. For each plot, the following statistics were recorded and were subsequently used as outcome variables in analysis of predictors: (i) segment 1 die-off rate (seg1), (ii) segment 1 die-off rate standard error (se1), (iii) segment 2 die-off rate (seg2), (iv) segment 2 die-off rate standard error (se2), (v) breakpoint between segment 1 and segment 2 (bp), (vi) if a segmented log-linear or log-linear model had a superior fit for the subset (die-off distribution), and (vii) if the observed die-off in the subset is compliant with FSMA (FSMA compliance). The 7 outcomes and predictors [i.e., study design factors (i.e., bacteria, produce type, and location) and weather factors (Table 3)] from an individual plot were compiled into a "processed" plot-level dataset for further analyses. This approach was taken because we hypothesize different weather factors are associated with each of the die-off segments; it is more practical for produce growers, public health agencies, and academia to utilize the relationship of weather and study design factors with die-off rates

compared to the relationship with microbial concentration on individual samples; this analysis method also readily takes data with multiple imputations. For the “die-off distribution” outcome, to assess if a segmented log-linear or log-linear model had a superior fit for the plot, the Bayesian information criterion (BIC) was estimated for each model type. For the segmented fit to be considered preferred, its BIC value must be 10 or more than the BIC value of the log-linear model. For the “FSMA compliance” outcome, to assess if the observed die-off in a plot is compliant with FSMA (i.e., at least a 0.5 $\log_{10}$ reduction/day), it was determined if the segmented die-off calculated for each plot would achieve at least a 2 $\log_{10}$ reduction in 4 days. This assumes, starting at a population level of 4 $\log_{10}$ CFU/100 g of produce and 100 mL is applied to each 100 g of produce, at least a 2 $\log_{10}$ reduction would need to be achieved in 4 days for the produce to be considered compliant (i.e., <126 CFU/100 mL) with the FSMA agricultural water standard.

Mixed effects linear regression was performed on the “processed” dataset using the lmer function in the “lme4” package (3) to determine univariable associations between each of the 7 continuous outcomes [(i) seg1, (ii) se1, (iii) cov1, (iv) seg2, (v) se2, (vi) cov2, and (vii) bp] and explanatory variables (study design factors and weather factors). Trial was included in the models as a random effect. Each study design factor and weather factor were tested separately as fixed effects. A F-test was used to compare the model fit with the fixed predictor to the model fit with only random effects. Weather predictors with a p-value of less than 0.1 in univariable analysis of an association with the particular outcome were subjected to principle component analysis for variable reduction.

Similarly, mixed effects logistic regression was performed on the “processed” dataset using the glmer function in the “lme4” package (3) to determine the univariable associations between each of the binary outcomes [(i) die-off distribution, and (ii) FSMA compliance] and explanatory variables (study design factors and weather factors). Other steps in the analyses followed the same plan as for the continuous outcomes. Namely, trial was included in the models as a random effect. Each study design factor and weather factor were tested separately as fixed effects. Weather predictors with a p-value of less than 0.1 in univariable analysis were subjected to principle component analysis for variable reduction.

Principle component analysis (PCA) was performed using the prcomp function, such that the number of components retained must explain greater than or equal to 90% of the variation in the data and each retained variable can only have a major loading on 1 principle component. PCA was performed separately for each outcome (because each outcome has different associated predictors based on univariable analysis). If an outcome had 4 or fewer predictors with $p<0.1$, PCA was not performed. One representative variable at a time for each principle component was included in multivariable analysis.

Mixed effects linear regression was performed using the lmer function in the “lme4” package (3) for multivariable analysis for each of the 7 continuous outcomes: seg1, se1, seg2, se2, and bp. Trial was included in the models as a random effect. All study design factors significant at 10% level in univariable analysis were first included in the multivariable models. Backwards selection was performed by removing the variable with the lowest t-value. The final base model was selected by the simplest model that was significantly different ($p<0.05$) from the next simplest nested model using a F-test. Once the base model was developed, a representative for each principle component was included in the base model and backwards selection was performed as described above to determine the final model.

Similarly, mixed effects logistic regression was performed using the glmer function in the “lme4” package (3) for multivariable analysis of each of the following binary outcomes: (i) die-off distribution and (ii) FSMA compliance. Trial was included in the models as a random effect. All study design factors significant at 10% level in univariable analysis were first included in the multivariable models. Backwards selection was performed by removing variables with the highest

p-value. The final base model was the simplest model significantly different ($p<0.05$) from the next simplest nested model using a likelihood ratio test. Once the base model was developed, representatives for each principle component were included in the base model and backwards selection was performed as described above to determine the final model.

Classification and regression trees were fit for each of the nine outcomes described above using the rpart function in the "rpart" package (33) to visualize possible interactions and aid interpretation of regression analysis. Tree pruning was performed such that the complexity parameter was set to minimize the cross validation relative error to prevent overfitting in the trees; a 10-fold cross-validation was used. Classification and regression trees cannot control for clustering, however, subsetting as part of the performed 10-fold cross-validation is expected to have reduced the effect of pseudo-replication at the trial level. As an internal validation step, the predictions from the classification trees are compared with the data and results are described in terms of sensitivity, specificity and negative and positive predictive values. Only weather and study design factors associated with the outcome in the univariable regression analysis were tested for inclusion in the classification and regression trees.

To determine the effect of trial, time, and produce variety on the survival of the three *E. coli* inoculum strains, multinomial regression was performed using the multinom function in the "nnet" package (37). To determine the effect of time and produce variety on the survival of the two *Salmonella* inoculum strains, mixed effects logistic regression was performed using the glmer function (3); sample id (i.e., the sample the isolate was taken from) was included in the model as a random effect.

Objective 1 Results:

Descriptions of samples and weather: In total, 5,252 data points were collected. However, only 4,900 data points were used in analyses due to crop loss in some trials (Table 1). Of the 4,900 data points used in analyses, 1,260 were *E. coli* on lettuce, 1,260 were *Salmonella* on lettuce, 1,190 were *E. coli* on spinach, and 1,190 were *Salmonella* on spinach. Of the 4,900 data points, 1,680, 1,680, and 1,540 data points were from the California, New York, and Spain trials, respectively. Summary statistics of the weather variables can be found in Table 4.

Descriptions of microbial counts from field data: The levels of *Salmonella* and *E. coli* in the inoculum ranged from 3.68–5.84 $\log_{10}$ CFU/mL and 3.77–5.84 $\log_{10}$ CFU/mL in all trials, respectively. The levels of *Salmonella* on produce at 0 h following inoculation ranged from 4.43–5.21 $\log_{10}$ CFU/100 g produce in California trials, 3.76–4.73 $\log_{10}$ CFU/100 g produce in New York trials, and 3.04–4.51 $\log_{10}$ CFU/100 g produce in Spain trials. The levels of *E. coli* on produce at 0 h following inoculation ranged from 5.51–6.67 $\log_{10}$ CFU/100 g produce in California trials, 4.42–5.68 $\log_{10}$ CFU/100 g produce in New York trials, and 3.40–6.56 $\log_{10}$ CFU/100 g produce in Spain trials. The mean level of *Salmonella* on produce at 96 h following inoculation ranged from 1.61–2.69 $\log_{10}$ CFU/100 g produce in California trials, 0.68–4.33 $\log_{10}$ CFU/100 g produce in New York trials, and 0.02–1.07 $\log_{10}$ CFU/100 g produce in Spain trials. As such, the range in total *Salmonella* reduction over 4 days was 1.94–3.14 $\log_{10}$ CFU/100 g produce in California trials, −0.12 (i.e., growth) to 3.28 $\log_{10}$ CFU/100 g produce in New York trials, and 2.56–4.26 $\log_{10}$ CFU/100 g produce in Spain trials. The mean concentration of *E. coli* on produce at 96 h following inoculation ranged from 1.62–2.74 $\log_{10}$ CFU/100 g produce in California trials, 0.55–2.89 $\log_{10}$ CFU/100 g produce in New York trials, and 0.36–3.34 $\log_{10}$ CFU/100 g produce in Spain trials. As such, the range in total *E. coli* reduction over 4 days was 3.48–4.40 $\log_{10}$ CFU/100 g produce in California trials, 2.29–4.21 $\log_{10}$ CFU/100 g produce in New York trials, and 2.63–4.97 $\log_{10}$ CFU/100 g produce in Spain trials.

In trial NY3, microbial die-off on two varieties of spinach were compared: Harris Seeds Seaside F1 (new variety) and Enza Zaden Acadia F1 (original variety); the original variety, which

was used in California and Spain trials, showed poor stand germination under New York conditions. No significant differences in the *E. coli* and *Salmonella* counts via a Wilcoxon rank sum test (p=0.801) or log-linear microbial die-off (p=0.988) were observed between the two spinach varieties. As such, the new variety of spinach was utilized in the remainder of New York trials and data from the new variety of spinach was combined with that of the original spinach variety for all analyses.

Several control samples (i.e., water used for irrigation, soil, or produce prior to inoculation) were presumptively positive (i.e., blue colonies on ECC+R or mauve colonies on SC+R for *E. coli* and *Salmonella*, respectively) for the presence of naturally occurring rifampicin-resistant *E. coli* and *Salmonella* in New York trials. In trial NY3, naturally occurring rifampicin-resistant *E. coli* and *Salmonella* were isolated from 1/1 irrigation water samples used in spinach growth. However, no further testing was performed, as no colonies indicative of *E. coli* or *Salmonella* were observed from soil or produce samples. In trial NY1, naturally occurring rifampicin-resistant *E. coli* was identified in 3/6 soil samples and naturally occurring rifampicin-resistant *Salmonella* was found in 2/6 soil samples and in 1/1 irrigation water samples. No further testing was performed, as *Salmonella* was not isolated from the 1 spinach or 1 lettuce sample. However, 1/1 spinach sample was positive for rifampicin-resistant *E. coli*. To determine if the contamination of spinach samples altered the population levels of *E. coli* on experimental spinach samples, the *E. coli* strain identification PCR protocol and PCR and subsequent sequencing of the *clpX* gene was performed on 24 randomly selected (i.e., using a random number generator) *E. coli* isolates from experimental samples. PCR and subsequent sequencing on the *clpX* gene was performed as previously described by Walk et al. (38) and Weller et al. (40). If the *E. coli* strain identification PCR banding pattern and *clpX* allelic type of the 24 isolates did not match one of the inoculum strains, there was 95% confidence the percentage of colonies of naturally occurring rifampicin-resistant *E. coli* was below 12%. This threshold was selected as there is indication that the uncertainty threshold of a typical plate count is approx. 12% (21). None of the 24 isolates tested from spinach samples in trial NY1 matched the *E. coli* strain identification PCR banding pattern or *clpX* allelic type of one of the inoculum strains, indicating the presence of naturally occurring rifampicin-resistant *E. coli* on spinach did not impact the results of the plate counts by more than would be expected in the typical uncertainty of the method. In trial NY2, naturally occurring rifampicin-resistant *E. coli* and *Salmonella* were isolated from the 1 irrigation water sample. Naturally occurring rifampicin-resistant *Salmonella* was also isolated from 6/6 soil samples. To account for the presence of rifampicin-resistant *E. coli* on the spinach sample collected in trial NY1, enumeration of naturally occurring *E. coli* and *Salmonella* in 1 spinach and 1 lettuce sample was also performed in trial NY3 and NY4. Both produce samples were negative for naturally occurring rifampicin-resistant *E. coli* and *Salmonella* by enumeration, with a limit of detection of 0.04 CFU per gram of produce. No control samples were positive for rifampicin-resistant *E. coli* or *Salmonella* in trial NY4, California, or Spain.

Variation in microbial die-off across trials: While the FSMA die-off rate (i.e., 0.5 $\log_{10}$ die-off/day) assumes a log-linear die-off pattern of foodborne pathogens and indicator organisms on in-field fresh produce, visual examination of the data and results from previous studies (26, 40) indicate a biphasic pattern may be more appropriate in most cases. As such, both segmented log-linear and log-linear models were fit to the data from each trial (Figures 2 and 3). The overall log-linear die-off rate across all data was −0.60 (95% CI: −0.63, −0.58) $\log_{10}$/day. However, when comparing the log-linear die-off on the trial level, there is substantial variation in the die-off rate, which ranged from −0.20 (95% CI: −0.27, −0.13) $\log_{10}$/day in trial NY1 to −1.01 (95% CI: −1.06, −0.95) $\log_{10}$/day in trial SP1 (Figure 2). The overall segmented log-linear die-off rate across all trials was −4.41 (95% CI: −4.69, −4.12) $\log_{10}$/day for 0 to 10 h, and −0.3 (95% CI: −0.33, −0.26) $\log_{10}$/day for 10 to 96 h. When comparing die-off across trials, the segment 1 die-off rate ranged

from −0.46 (95% CI: −0.52, −0.41) log_{10}/day in trial CA2 to −6.99 (95% CI: −7.38, −6.59) log_{10}/day in trial CA1. The breakpoints ranged from 2.5 h in trial NY1 to 3.48 d in trial NY2, with 9 out of 11 trials having a breakpoint at 12 h or earlier. The segment 2 die-off rate was less variable than the segment 1 die-off rate, in that it ranged from 0.28 (95% CI: −0.20, 0.77) log_{10}/day in trial NY2 to −1.00 (95% CI: −1.16, −0.85) log_{10}/day in trial CA2 (Figure 3), where a rate >0 indicates growth while rate <0 indicates die-off. The die-off rates on the plot level followed the same overall trend as those on the trial level, however, there was substantial variation in die-off between plots from a single trial in some cases (Table 5). As such, the remainder of the analyses were conducted on the plot level to account for this variation; trial was included as a random affect.

Associations between the die-off distribution outcome and explanatory variables – study design factors (i.e., produce type, bacteria, and location) and weather: Log-linear and segmented log-linear regression models were fit to the data from each plot (e.g., *Salmonella* population levels from the second lettuce plot in the first trial in Spain or *E. coli* population levels from the fourth spinach plot in trial NY4). Model fits were compared for each plot to determine superior fit. Based on the threshold of ≥10 larger BIC score for a segmented model to be considered superior, there was log-linear die-off on 33 plots (24%) and a segmented log-linear die-off on 107 plots (76%). For the multivariable regression, the only predictive factors retained in the model were average dew point (°C) and relative humidity range (%), such that a 1°C increase in average dew point is associated with a −0.35 (95% CI: −0.37, −0.32) change in log odds, and a 1% increase in relative humidity range is associated with a 0.09 (95% CI: 0.08, 0.09) change in log odds of following a segmented die-off pattern. No study design factors were retained in the model (Table 6). Only the maximum change in dew point from one hour to the next (i.e., maximum dew point change rate, °C/h) was retained in the classification tree for die-off pattern (Figure 4). Based on internal validation of this classification tree, the sensitivity, the specificity, positive predictive value, and negative predictive value were 0.64, 0.97, 0.88, and 0.90, respectively.

Associations between outcomes describing segmented die-off and study design factors and weather: Associations between segmented model outcomes and predictors describing study design and weather were assessed. We hypothesized the effects of study design and weather predictors on individual segmented die-off outcomes would differ, so each of the following outcomes were modeled separately: (i) segment 1 die-off rate (seg1), (ii) segment 1 die-off rate standard error (se1), (iii) segment 2 die-off rate (seg2), (iv) segment 2 die-off rate standard error (se2), and (v) breakpoint between segment 1 and segment 2 (bp). For each of these statistics, we report the outcomes as interquartile ranges (IQR; i.e., the middle 50% of observations), as the interquartile ranges are less impacted by outliers compared to the standard deviation; the mean and standard deviation are reported in Table 5.

The IQR of seg1 (log_{10} change/day) was from −7.32 to −3.45 for *E. coli* on spinach, −9.04 to −5.52 for *E. coli* on lettuce, −3.88 to −0.18 for *Salmonella* on spinach, and −5.32 to −0.93 for *Salmonella* on lettuce (Table 5). While there was seg1 die-off in the majority of plots, there was seg1 growth in 7.9% (11/140) of plots. For multivariable regression of seg1, produce variety, bacteria, and relative humidity range were retained in the model, such that *Salmonella* has slower seg1 die-off rates compared to *E. coli*, seg1 die-off on spinach is slower than on lettuce, and as the relative humidity range increases, seg1 die-off is faster (Table 7). Additionally, bacteria and minimum relative humidity were retained in the seg1 regression tree (Figure 5). Therefore, a similar pattern is seen between the results of multivariable regression and the regression tree for seg1 die-off rate. We hypothesize the more rapid die-off under lower relative humidity values is caused by desiccation of the leaf surface, which reduces the moisture available to the bacteria.

The IQR of se1 (log_{10} change/day) was from 0.48 to 0.89 for *E. coli* on spinach, 0.48 to 1.57 for *E. coli* on lettuce, 0.25 to 0.54 for *Salmonella* on spinach, and 0.24 to 0.72 for *Salmonella* on lettuce (Table 5). For multivariable regression of the se1, produce variety, bacteria, and maximum temperature (°C) were retained in the model (Table 7). The model showed there was a greater se1 for *E. coli* compared to *Salmonella*, there was a greater se1 for lettuce compared to spinach, and as the maximum temperature increased the se1 increased. For the se1 regression tree, maximum dew point (°C) and produce variety were retained in the model (Figure 6). However, maximum dew point and maximum temperature were strongly correlated (Spearman's rank coefficient = 0.80), again providing support for similarity between results of the two modeling approaches. We hypothesize the increase in variation in die-off is due to more stressful conditions that occur at higher temperatures or dew points. As such, it causes more variation in the die-off rates according to the resistance of the sub-populations of *E. coli* or *Salmonella* present on the spinach or lettuce.

The IQR of seg2 (log_{10} change/day) was from −0.56 to −0.17 for *E. coli* on spinach, −0.55 to −0.13 for *E. coli* on lettuce, −0.40 to −0.07 for *Salmonella* on spinach, −0.43 to −0.08 for *Salmonella* on lettuce (Table 5). While there was seg2 die-off in the majority of plots, there was seg2 growth in 16.4% (23/140) of plots. For multivariable regression for seg2, maximum relative humidity (%) and maximum relative humidity change rate (%/h) were retained in the model (Table 7). However, due to the small variation in seg2 die-off rates from plot to plot, the effect size of these variables was small. Additionally, no variables were retained in the seg2 regression tree.

The IQR of se2 (log_{10} change/day) was from 0.06 to 0.12 for *E. coli* on spinach, 0.06 to 0.13 for *E. coli* on lettuce, 0.08 to 0.13 for *Salmonella* on spinach, and 0.06 to 0.11 for *Salmonella* on lettuce (Table 5). For the multivariable se2 regression model, relative humidity range (%) was retained in the model (Table 7). However, the effect sizes of the weather variables on se2 are substantially smaller than those for se1. Additionally, no variables were retained in the se2 regression tree.

The IQR of bp (days) was from 0.38 to 0.54 day for *E. coli* on spinach, 0.25 to 0.48 day for *E. coli* on lettuce, 0.50 to 1.20 day for *Salmonella* on spinach, and 0.45 to 1.00 day for *Salmonella* on lettuce (Table 5). For the bp multivariable regression model, bacteria and relative humidity range were retained in the model, such that *Salmonella* is associated with a later bp compared to *E. coli*, and an increase in relative humidity range is associated with an earlier bp, which are the same relationships as with seg1 (Table 7). This may indicate that, as conditions become more stressful, the more sensitive sub-population die-off is more rapid and the underlying slow die-off of the more resistant sub-population becomes apparent at an earlier time following inoculation. However, the fit for the bp regression model is poor due to a non-linear relationship (i.e., at low relative humidity ranges the breakpoint appears to follow no pattern, and at higher relative humidity ranges the breakpoint tends to occur earlier). This indicates there are likely additional variables impacting bp. For the bp regression tree, minimum relative humidity, average relative humidity, and bacteria are retained (Figure 7).

Associations between the FSMA compliance outcome, study design factors, and weather: FSMA compliance was designated if the segmented die-off rate calculated for a plot would achieve at least a 2 log_{10} reduction in 4 days (i.e., this is the maximum allowable die-off, assuming a 0.5 log_{10}/day reduction for a maximum of 4 days as specified in FSMA). In total, 75% (105/140) of plots were compliant with FSMA. Additionally, 79% (27/34) of *E. coli* on spinach plots, 97% (35/36) of *E. coli* on lettuce plots, 56% of *Salmonella* on spinach plots, and 67% of *Salmonella* on lettuce plots were compliant with FSMA. According to multivariable logistic regression, produce type, bacteria, and average relative humidity were retained in the model (Table 6). Spinach is associated with a decrease in log odds of being compliant compared to lettuce (P<0.001), *Salmonella* is associated with a decrease in log odds of being compliant compared to *E. coli*

(*P*=0.017), and a decrease in average relative humidity is associated with an increased log odds of being compliant (*P*=0.002). Minimum relative humidity and bacteria were retained in the FSMA compliance classification tree (Figure 8). Internal validation indicated the sensitivity, specificity, positive predictive value, and negative predictive value for this classification tree were 0.69, 0.96, 0.86, and 0.90, respectively.

Comparison in the survival of *E. coli* and *Salmonella* inoculum strains: PCR of *E. coli* isolates was performed in all New York trials to determine if there was a difference in survival between the strains. According to multinomial regression, there was a significantly higher odds of an isolate being TVS 354 compared to TVS 353 as time increased (*P*=0.001), however, the effect size is small [OR=1.004 (95% CI: 1.002, 1.006); Table 8]. There was also a significantly higher odds of an isolate being strain TVS 355 compared to TVS 353 in trial NY2 (*P*<0.001), with an odds ratio of 1.858 (95% CI: 1.417, 2.435). There were no other significant differences between strains. Produce variety was not retained in the model.

Identification of *Salmonella* strains was performed in trial NY4 only. According to mixed effects logistic regression, there was an interaction between produce variety and time (*P*<0.001); as such, separate models were fit for spinach and lettuce for interpretation of results. On spinach, the odds of an isolate being attPTVS 355 compared to attPTVS 337 significantly decreased with time [OR=0.973 (95% CI: 0.957, 0.987); *P*<0.001; Table 9]. On lettuce, the odds of an isolate being attPTVS 355 compared to attPTVS 337 appeared to be stable with time [OR=0.997 (95% CI: 0.972, 1.019); *P*=0.774; Table 9].

Objective 2 Methods:

Internal validation of segmented log-linear model: For each trial, regression coefficients were determined for segment 1 and segment 2 die-off rates as well as for breakpoint from plot level data (relative to bacteria and produce type). Combining these coefficients into sets indexed relative to plot number, normal distributions were fit to capture the field level variation in segment 1 and segment 2 die-off and breakpoint. Monte Carlo simulations were then performed, sampling from the respective distributions for die-off rates and breakpoint.

Predictive model development and application to published field studies: A segmented log-linear model was built to extrapolate from the current study, predictions of bacteria die-off as a function of weather factors, bacteria (*E. coli* or *Salmonella*), and produce type (lettuce or spinach):

$$P(t) = \begin{cases} P_0 + \theta_1 t, & 0 \le t \le T^*, \quad (1a) \\ P_0 + \theta_1 T^* + \theta_2 (t - T^*), & t > T^*, \quad (1b) \end{cases}$$

where $P(t)$ (Log_{10} CFU) is the bacteria population at time t (h), P_0 (Log_{10} CFU) is the initial bacteria population level, θ_1 (Log_{10} CFU/h) is the segment 1 die-off rate, T^* (h) is the breakpoint, θ_2 (Log_{10} CFU/h) is the segment 2 die-off rate. Monte Carlo simulations were used to sample from distributions for P_0 , θ_1, T^*, and θ_2 to create a stochastic model.

Given the normality of the values from the plot level regression analysis for θ_2, a normal distribution was fit to all trials (except 1 & 2, which were found to follow a log-linear die-off pattern). The regression coefficients for the breakpoint T^* were normally distributed relative to trials using *E. coli* and lognormally distributed for trials using *Salmonella*. The respective distributions for T^* were therefore utilized relative to pathogen type.

While the distributions for T^* and θ_2 implicitly depend on weather, among all the weather variables recorded, regression analysis for segment 1 die-off at the plot level showed the

strongest association to relative humidity. Using the standard error from this analysis, the upper and lower bounds for segment 1 die-off for each trial (and produce/bacteria type) were correlated to the minimum relative humidity. Segment 1 die-off was defined to be a function of produce and bacteria type as well as minimum relative humidity:

$$\theta_1 = I_{\theta_1} + S_{\theta_1} \times (\min RH)$$

where I_{θ_1} is the distribution for die-off given 0 minimum relative humidity and S_{θ_1} is the distribution defining the increase in die-off (i.e. becoming less negative) relative to an incremental increase in the minimum relative humidity.

Mechanistic model and application to published field studies: As observed in this project, *E. coli* and *Salmonella* populations on lettuce and spinach in the field generally followed a bi-phasic decay. One mechanistic explanation for this pattern involves the hypothesis that persister cells, known as dormant cells originating through phenotypic variation, make up a significant fraction of the subpopulation(s) that survive over relatively longer periods of time. Using *E. coli* O157:H7 (EcO157), Munther et al. (30) showed the presence of persister cells on lettuce plants under various environmental conditions in the laboratory; at low relative humidity, the declining EcO157 populations showed the largest increase in the persister fraction. To develop a mechanistic understanding of the persisters' formation on plants, we developed a differential equation model that described the average dynamics of EcO157 normal and persister cell populations post inoculation (see ref 30) and applied it to data from published studies.

Objective 2 Results:

Internal validation of segmented log-linear model: The mathematical models parameterized based on the results of statistical analyses (Objective 1) showed a relatively good internal validity. The predicted microbial die-off was typically highly variable, which reflects the plot level variation among the regression coefficients for segment 1 and 2 die-off and breakpoint and also suggests that additional, yet unidentified, predictors may be able to further explain the variation observed at the plot level. At the same time the predictive model was able to capture the die-off trend in most cases. More specifically, there was a good agreement between the model predictions and the middle 50% (IQR) of the data for most trials. Figures 9 and 10 are provided for illustration. In each of these figures, population level data are shown as boxplots, whereas the model predictions are illustrated by the red (middle 50% of the model prediction) and blue (upper and lower 25%) regions. Features observed in model predictions were large variations in segment 1 die-off or breakpoint and almost mean zero segment 2 die-off with sufficient variation to allow the possibility of segment 2 growth.

Predictive model development and application to published field studies: The model (1a)–(1b) has been tested by application to published field studies on *E. coli* O157:H7 (EcO157:H7) die-off on lettuce conducted in the Salinas Valley in California (27, 28, 29). Figures 11 and 12 show the model predictions versus the collected field data in those independent studies. Since the field data were presented as mean $\pm$ standard deviation, the model outputs include the mean population (solid black curve) and the standard deviation about the mean, given by the red region. While the model predictions tend to overestimate the population level at 2 h past inoculation, they generally show good agreement with the observed data at later time points (Figures 11 and 12). Recall that for prediction the model only uses population level information from the observed data at time zero.

It is also important to mention that the predictions are sensitive to the segment 1 die-off distribution and are therefore sensitive to the relative humidity (Figures 11 and 12). This observation along with the relative success of the model predictions indicate relative humidity

information contains predictive merit. However, Figures 11 and 12 also further support that there are likely other factors responsible for some of the variability in the microbial die-off.

Testing of the mechanistic model of persisters formation: Applying the model to the low relative humidity experiment in the lab, the switching rate from normal to persister state, α_d, was found to be between $[7.7 \times 10^{-6}, 2.8 \times 10^{-5}]$ (1/h) (see Figure 13 for more information). This was very similar to the results from applying the same model to four EcO157 field studies: Bezanson et al. (4); Moyne et al. (27); Moyne et al. (28); and Erickson et al. (14), where α_d was mostly within the range $[1.1 \times 10^{-6}, 4.8 \times 10^{-5}]$ (1/h). These results support that the persisters' formation on produce may be the mechanism contributing to the observed biphasic die-off and demonstrate how laboratory and field level data could be used together to further elucidate the existence and characteristics of the persisters' formation mechanism.

Outcomes and Accomplishments

- Developed standardized field and laboratory protocols, which were used by collaborators in three study locations. This allowed researchers to gather data under a wider variety of open field conditions and develop a better understanding of the factors impacting die-off. This outcome was necessary in the completion of objective 1.
- Identified the pattern (i.e., log-linear or segmented log-linear) of die-off *E. coli* and *Salmonella* followed on baby spinach and lettuce under field conditions. Additionally, we identified weather factors associated with the die-off pattern (i.e., when die-off followed a log-linear vs. segmented log-linear pattern). This outcome was necessary for the completion of objective 1.
- Identified study design factors (i.e., bacteria, produce type, location) and weather factors associated with segmented die-off and the standard error of the segmented die-off. This outcome was necessary for the completion of objective 1.
- Identified study design factors and weather factors associated with whether in-field die-off of *Salmonella* and *E. coli* was compliant with the FSMA die-off rate (i.e., 0.5 $\log_{10}$ die-off per day for up to 4 days). *This outcome was an additional achievement outside the scope of the original objectives.*
- Identified the strain of the *E. coli* isolates from samples at different time points following inoculation to determine if there was a difference in survival among the strains. To do so, a novel PCR protocol (including primers) was developed to differentiate the 3 inoculum strains. The strains that were used are also commonly used in other preharvest produce studies. As such, other researchers can use the PCR protocol in the future to characterize survival of the inoculum strains. *This outcome was an additional achievement outside the scope of the original objectives.*
- Identified the strain of the *Salmonella* isolates from the samples at different time points following inoculation to determine if there was a difference in survival among the strains. *This outcome was an additional achievement outside the scope of the original objectives.*
- Developed a model to predict the die-off rate of *Salmonella* and *E. coli* on baby spinach and lettuce based on the weather conditions. This outcome was necessary for the completion of objective 2.
- Developed a mechanistic model of persisters' formation. This outcome was necessary for the completion of objective 2.

Summary of Findings and Recommendations

- The dew point was found to be associated with the die-off pattern of *Salmonella* and *E. coli* on baby spinach and lettuce.
- The results indicate die-off of *Salmonella* and *E. coli* on baby spinach and lettuce follows a segmented log-linear pattern the majority of the time. The die-off rate in the first segment is variable and appears to be associated with relative humidity, produce type, and bacteria.
- After the breakpoint, the die-off rate is less variable, however, there is still a large variation in the counts at each time point in this segment across experimental plots and trials. This indicates the produce industry should be aware that a time-to-harvest intervention does not completely eliminate the associated food safety risks.
- The experimental die-off for the plots investigated in the current study were compliant with FSMA the majority of time, and weather can be used to help explain when this is the case.
- The development of a predictive model, which takes into account weather, bacteria, and produce type to describe microbial die-off on produce appears to perform better than using a fixed die-off rate, as is included in FSMA. As such, public health officials should consider updating the regulation to take into account weather, produce variety, and bacteria.

APPENDICES

Publications and Presentations

Publications:

- A manuscript discussing the findings from objective one of this project, titled "Effect of weather on the die-off of *Escherichia coli* and attenuated *Salmonella* Typhimurium on preharvest baby lettuce and spinach following a standardized simulated irrigation event with contaminated water in New York, California, and Spain," was submitted and is under review by Applied and Environmental Microbiology.
- A manuscript discussing the findings from objective two of this project, with a tentative title "Modeling foodborne *Escherichia coli* and attenuated *Salmonella* Typhimurium die-off on preharvest lettuce and spinach as a function of local weather" will also be submitted.
- The work describing a mechanistic model of persisters' formation as part of objective 2 has been published: "Formation of *E. coli* O157:H7 persister cells in the lettuce phyllosphere and application of differential equation models to predict their prevalence on lettuce plants in the field", Applied and Environmental Microbiology, 2019, 01602-19. (ref 30)

Presentations:

- *"*Die-off of *E. coli* and attenuated *Salmonella* Typhimurium on baby lettuce and spinach under field conditions following a standardized simulated irrigation event with contaminated water in New York, California, and Spain*.*" International Association of Food Protection Annual Meeting. Poster Presentation. July 2019.
- "Microbial safety of fresh produce – the role of surface waters and weather". Global Water and Food Safety Summit (Nov 19-21, 2019), organized by the Joint Institute for Food Safety and Applied Nutrition – Center for Food Safety and Security Systems (JIFSAN-CFS3), College Park, MD, USA.

Budget Summary

The total funds awarded to this project were $399,204. All funds will be expended by the end of the budget term.

Tables and Figures (see below)

Tables 1–9 and Figures 1–13

Table 1. Description of the experimental setup for each trial.

Location	Trial	Produce Type	Produce Variety[a,b]	Number of Plots[c]	Date of Inoculation	Time from Planting to Inoculation (days)	Sample Collection Times (h)[d]	No. Samples Collected per Plot per Timepoint	Data included in analysis[e]
California	CAp	Lettuce	Tamarindo	4	7/19/2018	44	0, 24, 96	3	No
		Spinach	Acadia F1	4	7/12/2018	37	0, 24, 96	3	No
	CA1	Lettuce	Tamarindo	4	11/12/2018	38	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	11/12/2018	38	0, 4, 8, 24, 48, 72, 96	5	Yes
	CA2	Lettuce	Tamarindo	4	12/18/2018	57	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	12/18/2018	57	0, 4, 8, 24, 48, 72, 96	5	Yes
	CA3	Lettuce	Tamarindo	4	7/1/2019	59	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	7/1/2019	59	0, 4, 8, 24, 48, 72, 96	5	Yes
New York	NY1	Lettuce	Tamarindo	3	8/27/2018	28	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Seaside F1	3	8/27/2018	48	0, 4, 8, 24, 48, 72, 96	5	Yes
	NY2	Lettuce	Tamarindo	3	10/1/2018	38	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Seaside F1	3	10/1/2018	38	0, 4, 8, 24, 48, 72, 96	5	Yes
	NY3	Spinach	Acadia F1	2	7/16/2018	31	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Seaside F1	2	7/16/2018	31	0, 4, 8, 24, 48, 72, 96	5	Yes
	NY4	Lettuce	Tamarindo	4	7/1/2019	40	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Seaside F1	4	7/1/2019	40	0, 4, 8, 24, 48, 72, 96	5	Yes
Spain	SP1	Lettuce	Tamarindo	4	5/29/2018	47	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	5/29/2018	47	0, 4, 8, 24, 48, 72, 96	5	Yes
	SP2	Lettuce	Tamarindo	4	1/8/2019	91	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	1/8/2019	91	0, 4, 8, 24	5	No
	SP3	Lettuce	Tamarindo	4	4/29/2019	77	0, 4, 8, 24, 48, 72, 96	5	Yes
		Spinach	Acadia F1	4	4/29/2019	77	0, 4, 8, 24, 48, 72, 96	5	Yes
	SP4	Lettuce	Tamarindo	2	7/1/2019	26	0, 4, 8, 24, 48, 72, 96	3	Yes
		Spinach	Acadia F1	4	7/1/2019	26	0, 4, 8, 24	3	No

[a]Tamarindo lettuce and Acadia F1 spinach were supplied by Enza Zaden (Enkhuizen, Netherlands). Tamarindo is red leaf lettuce. Tamarindo and Acadia F1 are ideal for baby leaf.

[b]Seaside F1 spinach was supplied by Harris Seeds (Rochester, NY) and is ideal for baby leaf.

[c]The number of plots planted per trial varied to balance available resources and the need for additional trials in each location.

[d]Sample collection times are reported in hours past inoculation. Trials where samples were not collected at all seven time points was caused by crop loss.

[e]Trials or produce varieties in a trial where samples were not collected at all time points were excluded from data analysis. This was due to crop loss.

Table 2. Primer sequences used to differentiate between the 3 *E. coli* inoculum strains.

Primer Name	Target Strain	Sequence
353F	TVS353	TGACGGACAGGGACTCTATCTG
353R	TVS353	CAGCGTTCGCTCACTGAGAG
354F	TVS354	TAGGTTTGTTCACATTAGGTGATGTCG
354R	TVS354	AAATGTGGGTATGGCATATGGCAG
355F	TVS355	GTGACACCAATGACATCTGATGTTATCC
355R	TVS355	CGTCCTTATCCTGTTGGCTTGTG
35XF	All 3 strains	TTCGACAACGGTATTATTCTCTGCC
35XR	All 3 strains	TATCAATGACCCGAATCTGATCCTCG

Table 3. Weather variables created for data analysis.

Weather Variables
Minimum temperature (°C)
Maximum temperature (°C)
Average temperature (°C)
Temperature Range (°C)
Maximum Temperature Change Rate (°C/h)[a]
Minimum relative humidity (%)
Maximum relative humidity (%)
Average relative humidity (%)
Relative humidity range (%)
Maximum relative humidity change rate (%/h)[a]
Maximum solar radiation (kW/m^2)[b]
Average solar radiation (kW/m^2)[b]
Maximum solar radiation change rate ($kW/m^2 \cdot h$)[a,b]
Total precipitation (mm)
Minimum wind speed (m/s)
Maximum wind speed (m/s)
Average wind speed (m/s)
Wind speed range (m/s)
Maximum wind speed change rage (m/s·h)[a]
Minimum dew point (°C)
Maximum dew point (°C)
Average dew point (°C)
Dew point range (°C)
Maximum dew point change rate (°C/h)[a]

[a]The maximum change rate refers to the maximum change in the weather variable from one hour to the next.
[b]Solar radiation variables were not created for the 8h weather variables due to missing data.

Table 4. Summary statistics for weather variables across 11 trials over the whole length of the experiment of 96 h.

Category	Variable	Minimum	Q1	Median	Q3	Maximum	Mean	SD
Temperature (°C)	Minimum	-0.10	5.60	11.85	13.17	20.80	9.81	5.43
	Maximum	17.40	19.60	28.27	31.70	32.22	25.74	5.67
	Average	7.79	10.77	19.99	21.61	26.46	17.33	5.64
	Range	11.02	11.80	16.33	19.00	24.33	15.93	3.77
	Rate[a]	2.45	2.56	3.50	4.72	6.06	3.79	1.10
Relative Humidity (%)	Minimum	13.00	29.25	35.75	54.00	67.00	40.21	17.15
	Maximum	76.25	91.00	94.75	97.00	100.00	93.33	6.02
	Average	44.30	59.62	66.55	77.30	91.39	68.40	13.45
	Range	16.00	43.00	61.00	62.25	81.00	53.11	17.63
	Rate[a]	9.00	16.75	20.00	26.00	27.00	19.96	5.11
Solar Radiation (kW/m^2)	Maximum	0.25	0.44	0.78	0.86	0.99	0.68	0.23
	Average	0.05	0.09	0.23	0.25	0.35	0.19	0.10
	Rate[a]	0.13	0.18	0.27	0.28	0.39	0.25	0.08
Precipitation (mm)	Total	0.00	0.00	0.10	3.00	53.34	7.26	14.79
Wind Speed (m/s)	Minimum	0.00	0.00	0.00	0.50	0.60	0.18	0.26
	Maximum	2.50	3.89	4.29	4.97	5.23	4.19	0.86
	Average	0.42	0.87	1.35	1.67	2.22	1.40	0.51
	Range	2.00	3.50	4.10	4.97	5.23	4.01	0.99
	Rate[a]	1.20	1.56	2.10	2.26	3.74	2.14	0.83
Dew Point (°C)	Minimum	-9.20	1.00	7.22	10.90	14.44	5.14	7.52
	Maximum	2.60	12.20	15.80	20.56	22.78	14.93	6.15
	Average	-3.06	9.31	12.55	13.96	20.18	10.64	-3.06
	Range	4.90	8.36	8.89	11.80	14.44	9.79	4.90
	Rate[a]	2.22	2.50	3.23	3.80	5.10	3.26	2.22

[a]Rate refers to the maximum change in the weather variable from one hour to the next.

Table 5. Summary statistics for die-off factors separated by produce type (i.e., spinach and lettuce) and bacteria (i.e., *Salmonella* and *E. coli*). Die-off rates were calculated on the plot level.

Category	Variable	Minimum	Q1	Median	Q3	Maximum	Mean	SD
E. coli, spinach	Linear Die-off	-1.16	-0.93	-0.80	-0.62	-0.07	-0.72	0.29
	Linear SE	0.04	0.06	0.09	0.12	0.14	0.09	0.03
	Segment 1 Die-off	-10.42	-7.32	-4.93	-3.45	-0.14	-5.07	2.80
	Segment 1 SE	0.07	0.48	0.64	0.89	2.48	0.69	0.46
	Segment 2 Die-off	-0.75	-0.56	-0.44	-0.17	15.80	0.13	2.79
	Segment 2 SE	0.03	0.06	0.09	0.12	3.45	0.21	0.58
	Breakpoint	0.17	0.38	0.45	0.54	3.92	0.77	0.95
E. coli, lettuce	Linear Die-off	-1.04	-0.94	-0.79	-0.67	-0.33	-0.77	0.21
	Linear SE	0.04	0.09	0.11	0.13	0.20	0.11	0.04
	Segment 1 Die-off	-16.52	-9.04	-6.66	-5.52	-0.47	-7.07	3.41
	Segment 1 SE	0.06	0.48	0.66	1.57	4.12	1.07	0.94
	Segment 2 Die-off	-1.94	-0.55	-0.22	-0.13	3.04	-0.24	0.70
	Segment 2 SE	0.03	0.06	0.09	0.13	0.95	0.14	0.18
	Breakpoint	0.11	0.25	0.38	0.48	3.71	0.68	0.98
Salmonella, spinach	Linear Die-off	-1.04	-0.72	-0.56	-0.15	0.40	-0.45	0.38
	Linear SE	0.04	0.07	0.08	0.09	0.14	0.08	0.02
	Segment 1 Die-off	-7.52	-3.88	-2.77	-0.18	0.97	-2.37	2.10
	Segment 1 SE	0.04	0.25	0.36	0.54	1.41	0.41	0.29
	Segment 2 Die-off	-1.16	-0.40	-0.28	-0.07	1.81	-0.20	0.55
	Segment 2 SE	0.05	0.08	0.10	0.13	0.62	0.12	0.09
	Breakpoint	0.21	0.50	0.87	1.20	3.56	1.09	0.86
Salmonella, lettuce	Linear Die-off	-1.00	-0.70	-0.56	-0.42	0.07	-0.55	0.24
	Linear SE	0.03	0.06	0.09	0.11	0.13	0.08	0.03
	Segment 1 Die-off	-9.70	-5.32	-4.63	-0.93	0.33	-3.71	2.62
	Segment 1 SE	0.05	0.24	0.43	0.72	4.32	0.64	0.78
	Segment 2 Die-off	-7.30	-0.43	-0.25	-0.08	0.45	-0.48	1.21
	Segment 2 SE	0.04	0.06	0.08	0.11	2.56	0.16	0.42
	Breakpoint	0.11	0.45	0.52	1.00	3.89	0.86	0.84

Table 6. Final mixed effects multivariable logistic regression model parameters for experimental die-off pattern and FSMA compliance. Trial was included as a random effect in both models[a].

Outcome	Factor	Log Odds	95% CI	p-value
Die-off Pattern[b]	Intercept	2.02	(1.63, 2.41)	0.393
	Average Dew Point (°C)	-0.35	(-0.37, -0.32)	0.012
	Relative Humidity Range (%)	0.09	(0.08, 0.09)	0.008
FSMA Compliance[c]	Intercept	17.88	(17.06, 18.70)	0.000
	Produce Type (Spinach)[d]	-1.63	(-1.74, -1.51)	0.000
	Bacteria (*Salmonella*)[d]	-3.02	(-3.14, -2.89)	0.017
	Average Relative Humidity (%/10)	-1.88	(-1.98, -1.77)	0.002

[a]For the die-off pattern model, the variance and standard deviation for the trial random effect were 1.291 and 1.136, respectively. For the FSMA compliance model, the variance and standard deviation for the trial random effect were 2.343 and 1.531, respectively.

[b]Die-off pattern indicates if a biphasic segmented log-linear fit is superior to a log-linear fit for each plot and bacteria combination. The superior model fit for each plot and bacteria subset was determined such that for the segmented fit to be superior, its BIC value must be 10 or more than the BIC value of the log-linear model.

[c]FSMA compliance indicates if there was at least a 2 $\log_{10}$/ day die-off for each plot and bacteria combination.

[d]The baseline produce type is lettuce and the baseline bacteria is *E. coli*.

Table 7. Final mixed effects multivariable linear regression model parameters for the associations between study design factors (i.e., produce type, location, and bacteria) and 96h weather variables and the following outcomes: segment 1 die-off rate, segment 1 die-off rate standard error, segment 2 die-off rate, segment 2 die-off rate standard error, and the breakpoint between segment 1 and segment 2. Trial is included as a random effect in all models[e].

Outcome	Factor	Coefficient[b]	95% CI[c]
Segment 1 Die-off Rate (Log_{10} CFU/day)	Intercept	-1.25	(-1.43, -1.06)
	Produce Type (Spinach)[d]	1.77	(1.71, 1.82)
	Bacteria (*Salmonella*)[d]	3.04	(2.99, 3.09)
	Relative Humidity Range (%)	-0.11	(-0.11, -0.11)
Segment 1 Die-off Rate SE (Log_{10} CFU/day)	Intercept	-0.28	(-0.36, -0.19)
	Produce Type (Spinach)[d]	-0.37	(-0.39, -0.36)
	Bacteria (*Salmonella*)[d]	-0.35	(-0.37, -0.34)
	Maximum Temperature (°C)	0.05	(0.05, 0.06)
Segment 2 Die-off Rate (Log_{10} CFU/day)	Intercept	6.22	(5.88, 6.56)
	Maximum Relative Humidity (%)	-0.06	(-0.06, -0.05)
	Maximum Relative Humidity Change Rate (%/h)[a]	-0.06	(-0.06, -0.05)
Segment 2 Die-off Rate SE (Log_{10} CFU/day)	Intercept	0.43	(0.41, 0.45)
	Relative Humidity Range (%)	-0.01	(-0.01, 0.00)
Breakpoint (days)	Intercept	2.22	(2.16, 2.28)
	Bacteria (*Salmonella*)[d]	0.25	(0.23, 0.27)
	Relative Humidity Range (%)	-0.03	(-0.03, -0.03)

[a]The maximum change rate refers to the maximum change in the weather variable from one hour to the next.
[b]Coefficients were estimated using multivariable mixed effects linear regression via the lmer() function in R. Cohort was included as a random effect in all models.
[c]95% CI indicates a 95% confidence interval.
[d]The baseline produce type is lettuce and the baseline bacteria is *E. coli*.
[e]For the segment 1 die-off rate model, the residual variance and intercept for the random effects are 3.713 and 0.899, respectively. For the segment 1 die-off rate standard error model, the residual variance and intercept for the random effects are 0.331 and 0.093, respectively. For the segment 2 die-off rate model, the residual variance and intercept for the random effects are 2.245 and 0.015, respectively. For the segment 2 die-off rate standard error model, the residual variance and intercept for the random effects are 0.123 and 0.006, respectively. For the breakpoint model, the residual variance and intercept for the random effects are 0.543 and 0.078, respectively.

Table 8. Multinomial regression model parameters for the associations between time and trial and the odds of an isolate being *E. coli* inoculum strain TVS 354 or TVS 355 compared to strain TVS 353 in New York trials.

Factor[a]	TVS 354		TVS 355	
	Odds Ratio (95% CI)	P-value	Odds Ratio (95% CI)	P-value
Intercept	1.121 (0.895, 1.404)	0.321	1.150 (0.917, 1.442)	0.226
Time	1.004 (1.002, 1.006)	0.001	1.000 (0.998, 1.003)	0.682
Trial (NY2)	1.275 (0.970, 1.676)	0.082	1.858 (1.417, 2.435)	<0.001
Trial (NY3)	0.952 (0.729, 1.241)	0.714	1.095 (0.839, 1.428)	0.504
Trial (NY4)	0.989 (0.766, 1.277)	0.933	0.941 (0.726, 1.219)	0.645

[a]Trial NY1 was the baseline in the model.

Table 9. Mixed effects logistic regression model parameters for the associations between time and the odds of an isolate being *Salmonella* inoculum strain PTVS 355 compared to strain PTVS 337 when isolated from spinach and lettuce in trial New York 4. Sample ID was included in each the model as a random effect[b].

Produce Type	Factor	Odds Ratio (95% CI)[a]	P-value
Spinach	Intercept	2.595 (0.997, 7.425)	0.051
	Time	0.973 (0.957, 0.987)	<0.001
Lettuce	Intercept	0.713 (0.196, 2.667)	0.585
	Time	0.997 (0.972, 1.019)	0.774

[a]95% CI: 95% confidence interval

[b]The variance and standard deviation of the random effect sample id for the spinach model were 1.75 and 1.323, respectively. The variance and standard deviation of the random effect sample id for the lettuce model were 3.1 and 1.761, respectively.

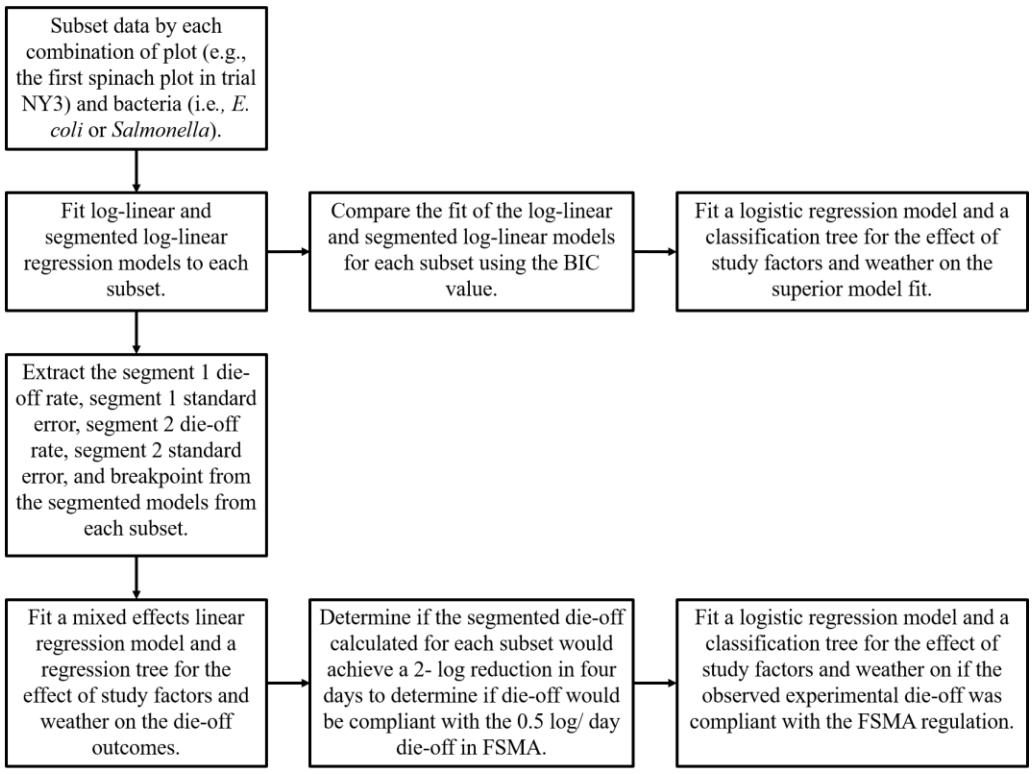


Figure 1. Overview of the statistical analysis plan on the plot level.

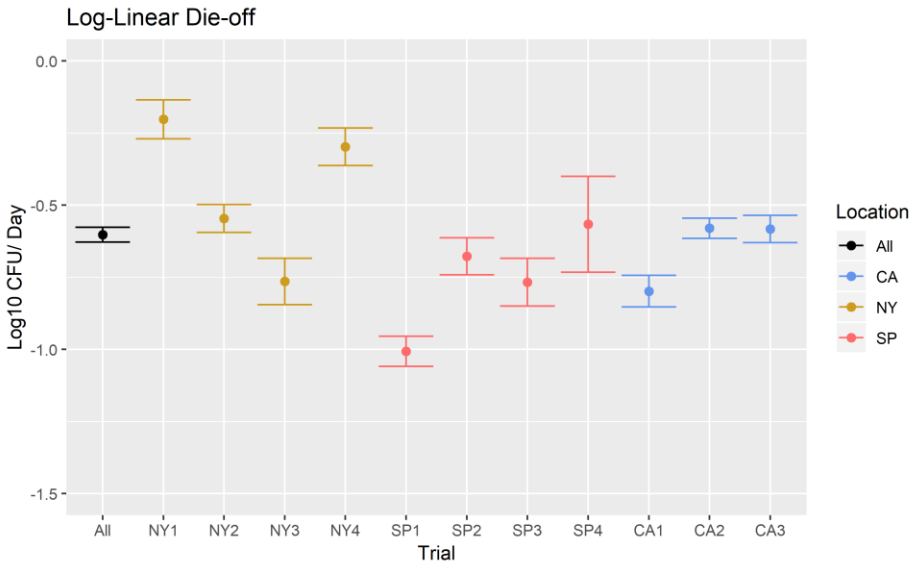


Figure 2. Log-linear die-off by trial as Log_{10} CFU die-off per day. Black points indicate the mean die-off for data across all trials, blue points indicate die-off for California trials, yellow points indicate die-off for New York trials, and pink points indicate die-off for Spain trials. Error bars represent the 95% confidence intervals for the mean die-off rates from each respective trial.

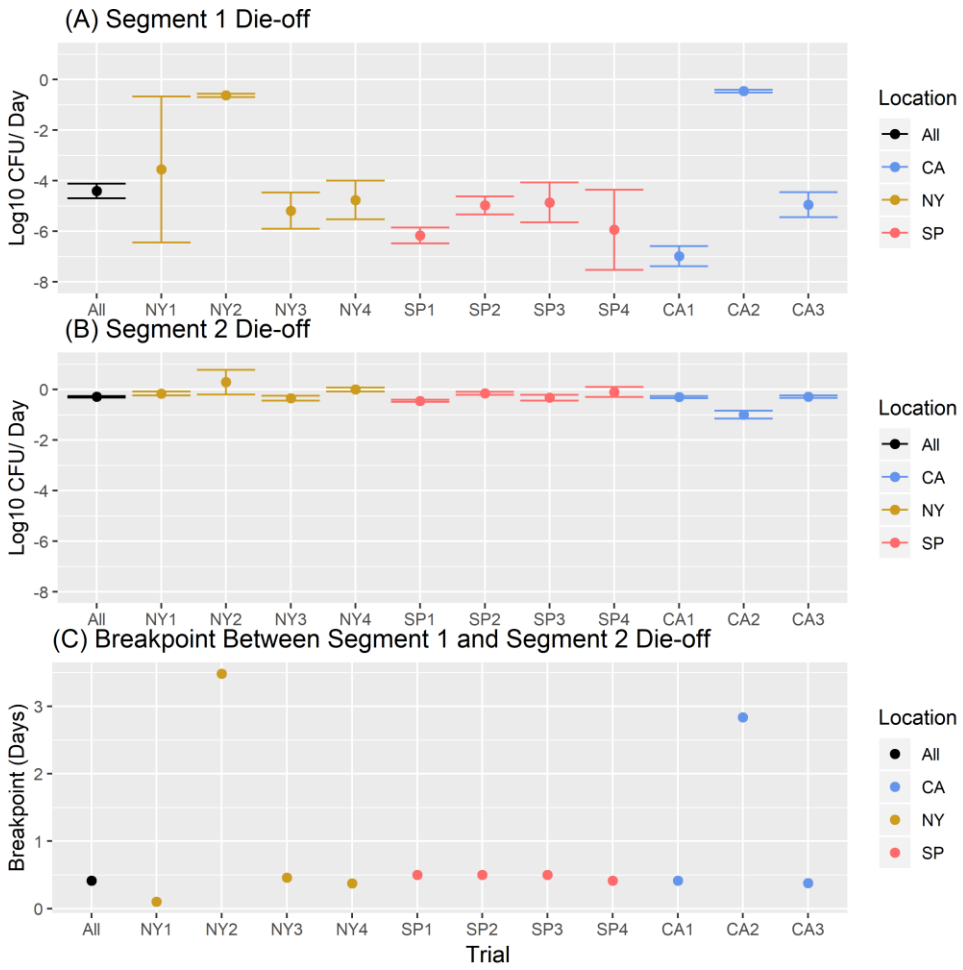


Figure 3. Segmented log-linear die-off by trial. (A) Segment 1 die-off rate in Log_{10} CFU die-off per day, (B) Segment 2 die-off rate in Log10 CFU die-off per day, and (C) breakpoint between segment 1 and segment 2 die-off. In plots A and B, black points indicate the mean die-off rate for data across all trials, blue points indicate die-off for California trials, yellow points indicate die-off for New York trials, and pink points indicate die-off for Spain trials. Error bars for segment 1 and segment 2 die-off rate represent the 95% confidence intervals for the mean die-off rates from each respective trial.

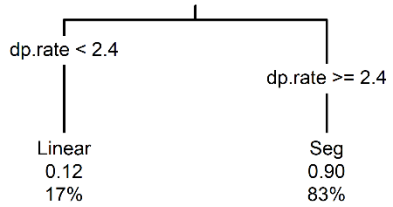


Figure 4. Classification tree displaying the relationship between the superior die-off pattern (i.e., log-linear or segmented log-linear) and maximum dew point change rate (°C/h) for the experimental plots (N=140). The superior model fit for each plot and bacteria subset was determined such that for the segmented fit to be superior, its BIC value must be 10 or more than the BIC value of the log-linear model. The classification tree was fit using the rpart function in R; tree pruning was performed to avoid overfitting. At the end of each terminal node, the superior die-off pattern is designated as either Linear (i.e., log-linear die-off is superior) or Seg (i.e., segmented log-linear die-off is superior). The first number below the designated die-off pattern indicates the probability the segmented log-linear model is superior, and the second number indicates the percentage of observations that fall in that node.

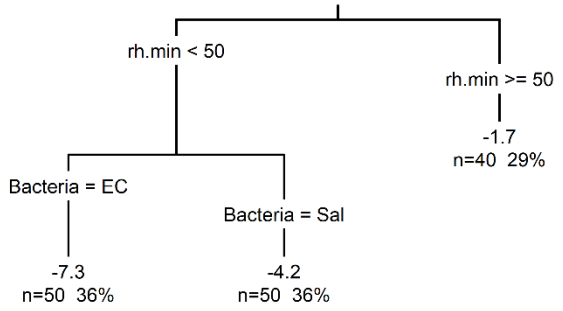


Figure 5. Regression tree displaying the relationship between segment 1 die-off rate (Log_{10} CFU/day), bacteria (EC= *E. coli*, Sal= *Salmonella*) and minimum relative humidity (%) for the experimental plots. The regression tree was fit using the rpart function in R; tree pruning was performed to avoid overfitting. The first number listed at each terminal node is the mean segment 1 die-off rate (Log_{10} CFU/day) for that node, the next number (i.e., n=) designates the number of observations that fall in that node, and the final number designates the percentage of observations that fall in that node.

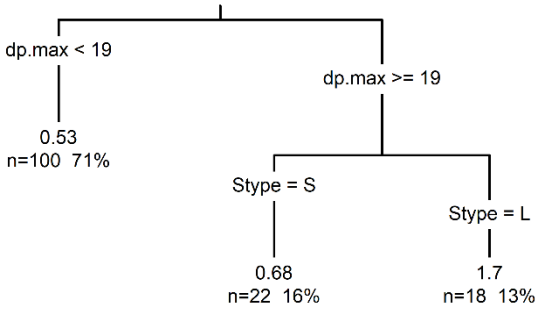


Figure 6. Regression tree displaying the relationship between segment 1 die-off rate standard error (Log_{10} CFU/day), maximum dew point (°C), and produce type (Stype, where S=spinach and L=lettuce) for the experimental plots. The regression tree was fit using the rpart function in R; tree pruning was performed to avoid overfitting. The first number listed at each terminal node is the mean segment 1 die-off rate standard error (Log_{10} CFU/day) for that node, the next number (i.e., n=) designates the number of observations that fall in that node, and the final number designates the percentage of observations that fall in that node.

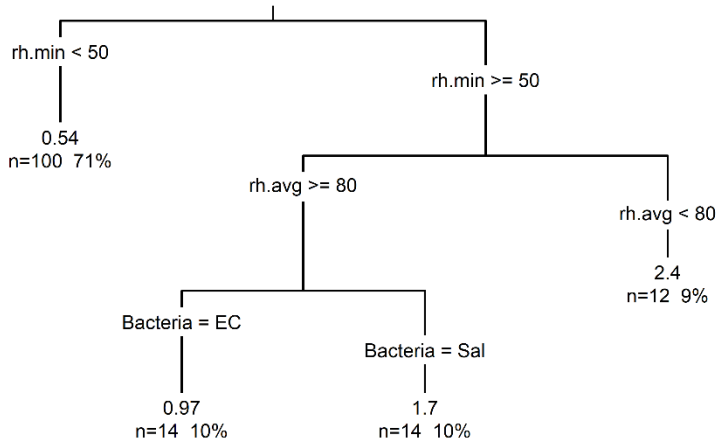


Figure 7. Regression tree displaying the relationship between the breakpoint (days) between segment 1 and segment 2, minimum relative humidity (%), average relative humidity, and bacteria (EC= *E. coli*, Sal= *Salmonella*) for the experimental plots. The regression tree was fit using the rpart function in R; tree pruning was performed to avoid overfitting. The first number listed at each terminal node is the mean breakpoint (days) for that node, the next number (i.e., n=) designates the number of observations that fall in that node, and the final number designates the percentage of observations that fall in that node.

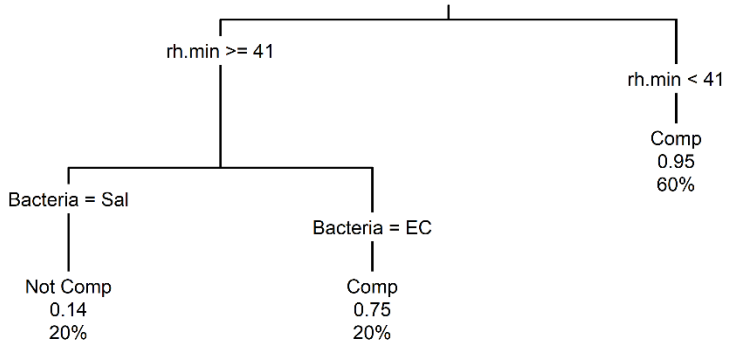


Figure 8. Classification tree displaying the relationship between compliance with FSMA, minimum relative humidity (%), and bacteria (EC= *E. coli*, Sal= *Salmonella*) for the experimental plots (N=140). Compliance was designated if the segmented die-off calculated for each experimental plot would achieve at least a 2- log reduction in 4 days (i.e., assumes a 0.5 log_{10}/day die-off as specified is FSMA). The classification tree was using the rpart function in R; tree pruning was performed to avoid overfitting. At the end of each terminal node, whether the experimental die-off was compliant with FSMA (i.e., Comp) or was not compliant with FSMA (i.e., Not Comp) is designated. The first number below the FSMA compliance designation is the probability of being compliant and the second number indicates the percentage of observations that fall in that node.

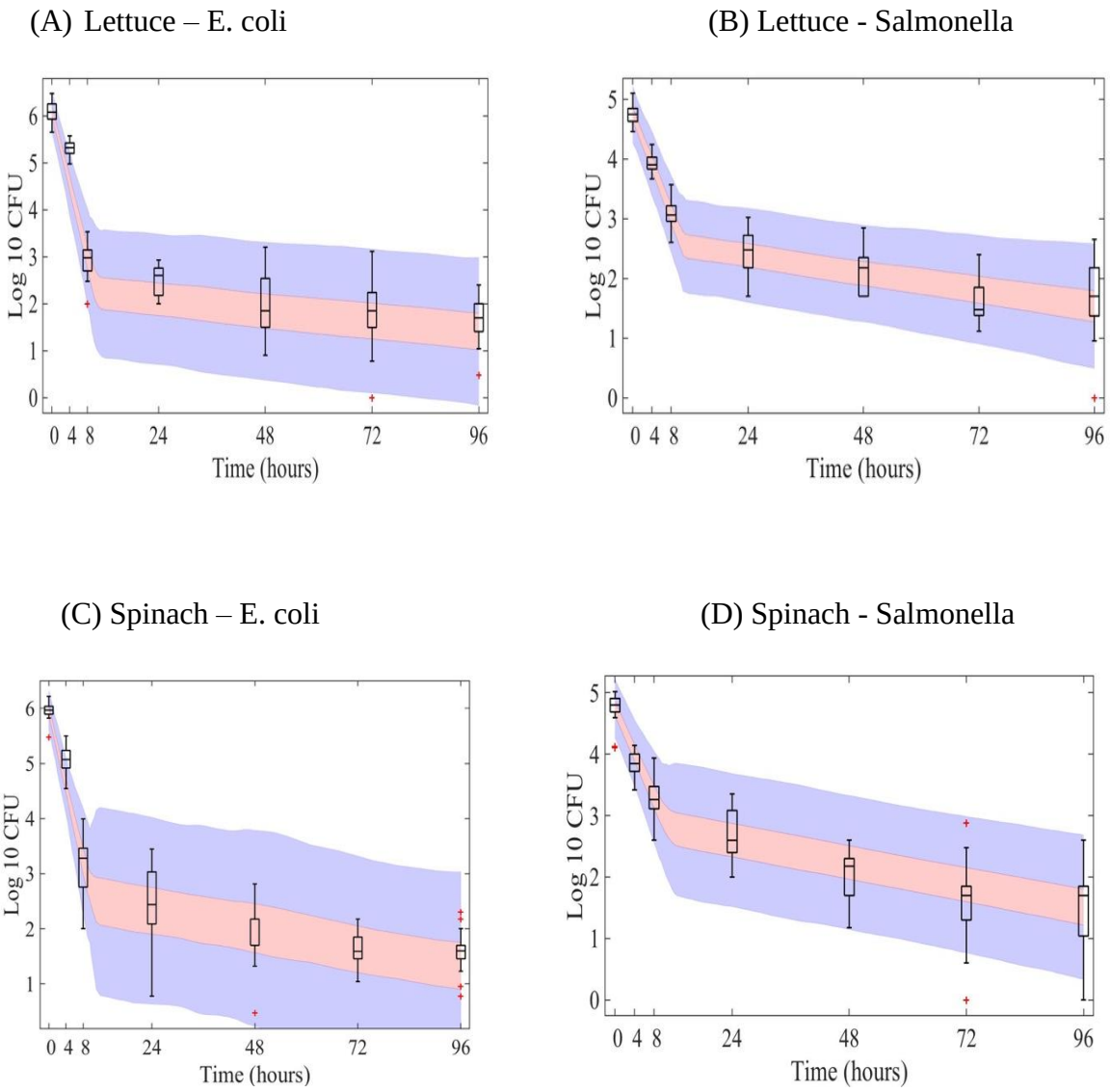
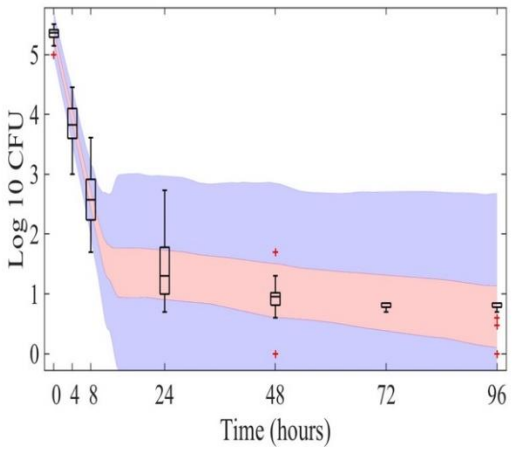
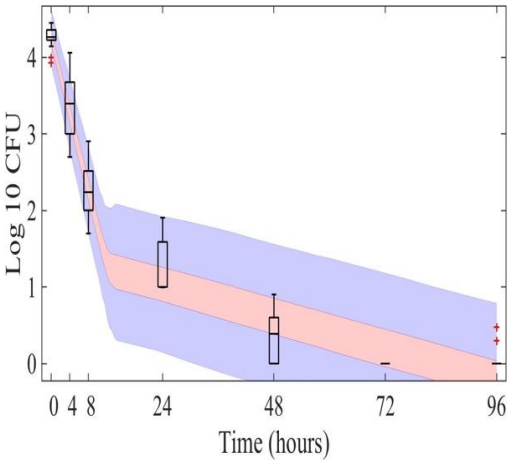


Figure 9. California Trial 7. Internal validation for segmented log-linear model. Observed time series data for the respective bacteria population type are illustrated via boxplots. Corresponding model predictions are given via the red region (middle 50%) and blue regions (upper and lower 25%).

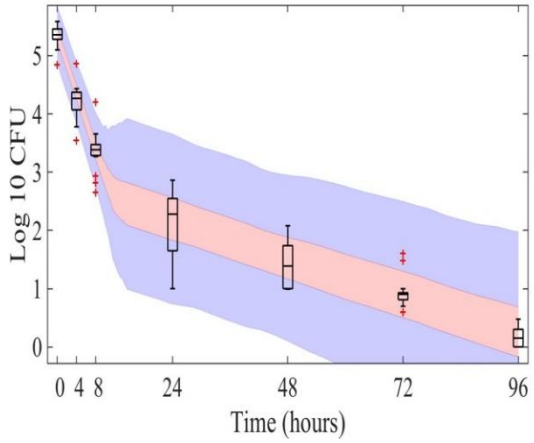
(A) Lettuce – E. coli



(B) Lettuce - Salmonella



(C) Spinach – E. coli



(D) Spinach - Salmonella

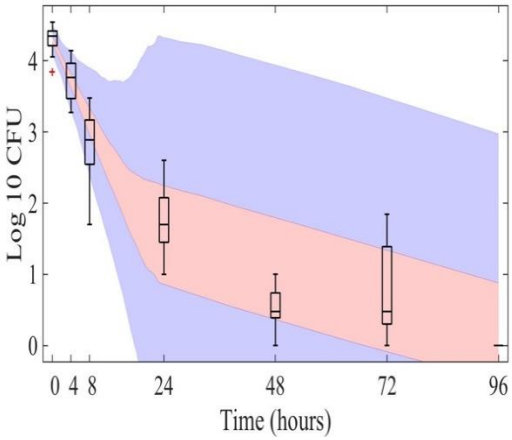


Figure 10. Spain Trial 4. Internal validation for segmented log-linear model. Observed time series data for the respective bacteria population type are illustrated via boxplots. Corresponding model predictions are given via the red region (middle 50%) and blue regions (upper and lower 25%).

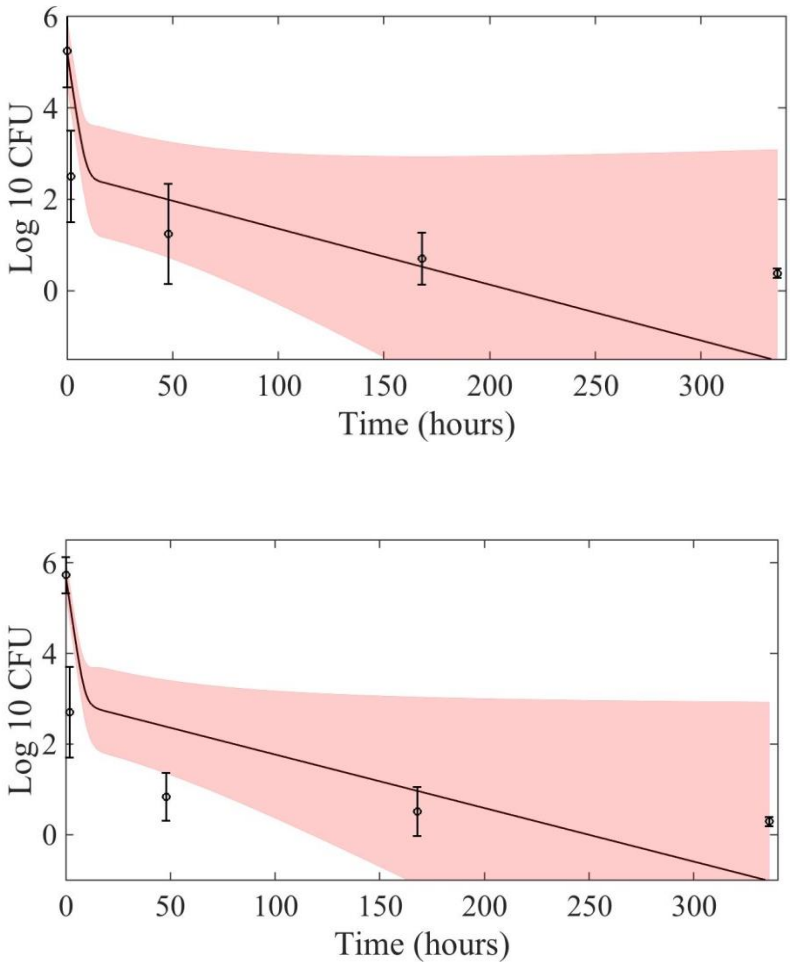


Figure 11. Model prediction of EcO157 die-off from Moyne et al., 2011. EcO157 field data are presented as mean (circles) ± standard deviation (error bars) against the model prediction of the mean (black curve) and the standard deviation about the mean (red region). The top panel is from the Spring 2009 trial and the bottom panel is from the Fall 2009 trial.

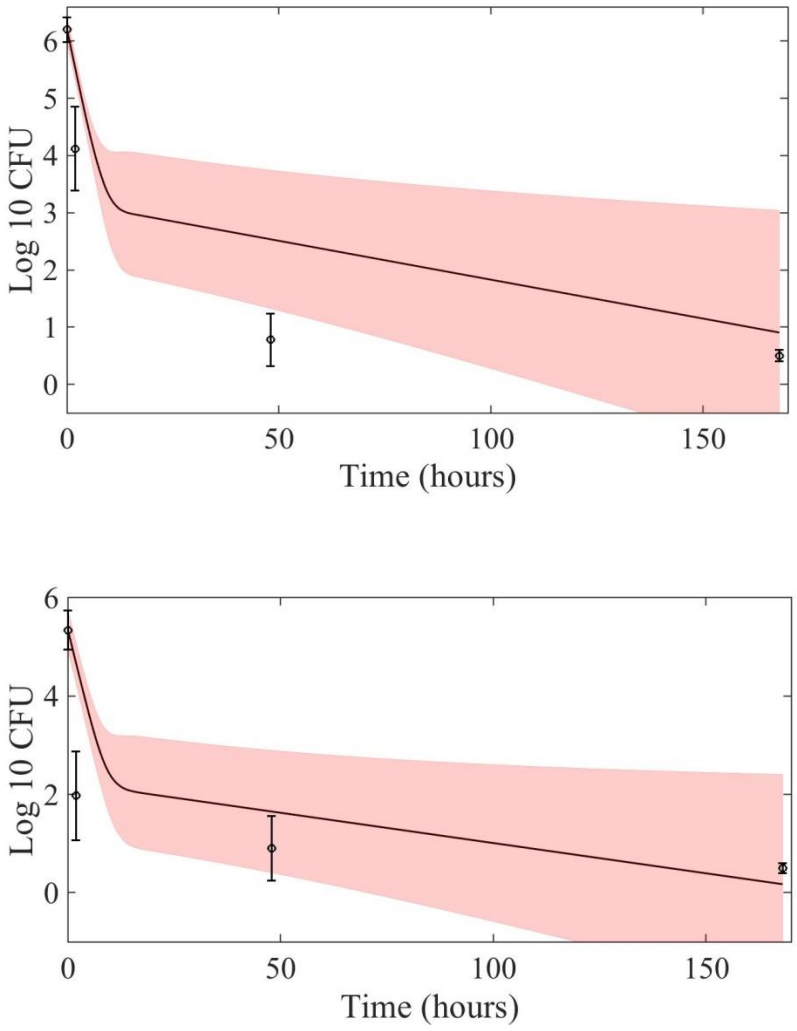


Figure 12. Model prediction of EcO157 die-off from Moyne et al., 2013. EcO157 field data are presented as mean (circles) ± standard deviation (error bars) against the model prediction of the mean (black curve) and the standard deviation about the mean (red region). The top panel is from the Spring 2010 trial and the bottom panel is from the Summer 2010 trial.

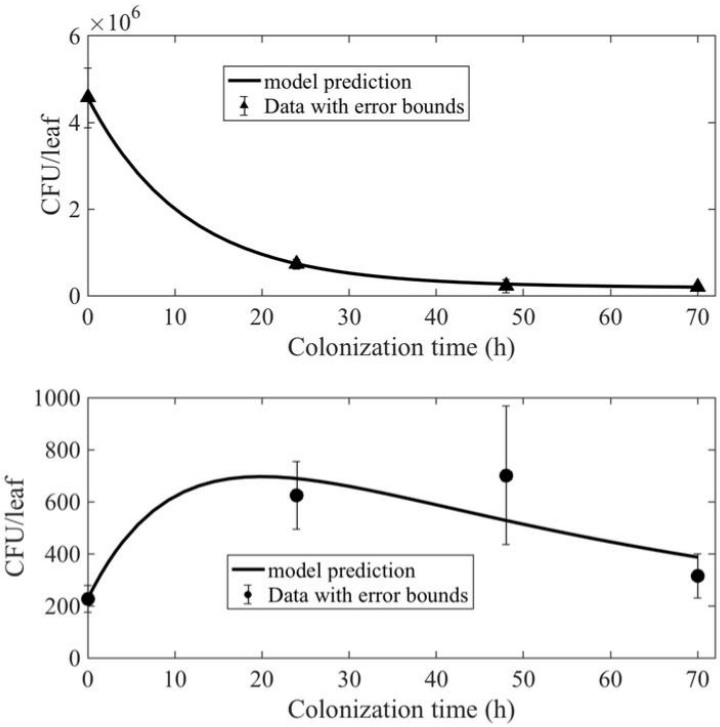


Figure 13. Differential equation model prediction for low relative humidity regime. Model predictions of EcO157 total (top panel) and persister (bottom panel) populations under experimental decay conditions for inoculated lettuce leaves. Requiring the persister population to remain within the experimental error bounds for the persister population at the respective time points, $\alpha_d \in [7.7 \times 10^{-6}, 2.8 \times 10^{-5}]$ (1/h). Within this region of parameter space, the best model fit corresponds to $\alpha_d = 1.6 \times 10^{-5}$ (1/h), with R^2 = 0.75 (Figure from Munther et al., 2019; ref 30).

References Cited

1. Anderson GB, Bell ML, Peng RD. 2013. Methods to calculate the heat index as an exposure metric in environmental health research. Environ Health Perspect 121(10):1111-1119.
2. Barker-Reid F, Harapas D, Engleitner S, Kreidl S, Holmes R, Faggian R. 2009. Persistence of *Escherichia coli* on injured iceberg lettuce in the field, overhead irrigated with contaminated water. J Food Prot 72(3):458-464.
3. Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1-48.
4. Bezanson G, Delaquis P, Bach S, McKellar R, Topp E, Gill A, Blais B, Gilmour M. 2012. Comparative examination of *Escherichia coli* O157:H7 survival on romaine lettuce and in soula and two independent experimental sites. J Food Prot 75(3):480-487.
5. Centers for Disease Control and Prevention. 2006. Multistate outbreak of *E. coli* O157:H7 infections linked to fresh spinach (final update). Accessed on December 15, 2019. https://www.cdc.gov/ecoli/2006/spinach-10-2006.html
6. Centers for Disease Control and Prevention. 2018. Multistate outbreak of *E. coli* O157:H7 infections linked to romaine lettuce (final update). Accessed on December 15, 2019. https://www.cdc.gov/ecoli/2018/o157h7-04-18/index.html
7. Centers for Disease Control and Prevention. 2019. Outbreak of *E. coli* infections linked to romaine lettuce (final update). Accessed on December 15, 2019. https://www.cdc.gov/ecoli/2018/o157h7-11-18/index.html
8. 21 CFR § 112.44. 2019.
9. 21 CFR § 112.45. 2019.
10. Chase JA, Atwill ER, Partyka ML, Bond RF, Oryang D. 2017. Inactivation of *Escherichia coli* O157:H7 on romaine lettuce when inoculated in a fecal slurry mix. J Food Prot 80(5):792-798.
11. Chase JA, Partyka ML, Bond RF, Atwill ER. 2019. Environmental inactivation and irrigation-mediated regrowth of *Escherichia coli* O157:H7 on romaine lettuce when inoculated in a fecal slurry matrix. PeerJ 7:e6591.
12. Curtiss R III, Kelly SM. 1987. *Salmonella* Typhimurium deletion mutants lacking adenylate cyclase and cyclic AMP receptor protein are avirulent and immunogenic. Infect Immun 55(12):3035-3043.
13. de Moraes MH, Chapin TK, Ginn A, Wright AC, Parker K, Hoffman C, Pascual DW, Danyluk MD, Teplitski M. 2016. Development of an avirulent *Salmonella* surrogate for modeling pathogen behavior in pre- and postharvest environments. Appl Environ Microbiol 82(14):4100-4111.
14. Erickson MC, Webb CC, Diaz-Perez JC, Phatak SC, Silvoy JJ, Davey L, Payton AS, Liao J, Ma L, Doyle MP. 2010. Surface and internalized *Escherichia coli* O157:H7 on field-grown spinach and lettuce treated with spray-contaminated irrigation water. J Food Prot 73(6):1023-1029.
15. Fonesca JM, Fallon SD, Sanchez CA, Nolte KD. 2011. *Escherichia coli* survival in lettuce fields following its introduction through different irrigation systems. J Appl Microbiol 110:893-902.
16. Gutierrez-Rodriguez E, Gunderson A, Sbodio AO, Suslow TV. 2011. Variable agronomic practices, cultivar, strain source and initial contamination dose differentially affect survival of *Escherichia coli* on spinach. J Appl Microbiol 112:109-118.

17. Gutierrez-Rodriguez E, Gunderson A, Sbodio A, Koike S, Suslow TV. 2019. Evaluation of post-contamination survival and persistence of applied attenuated *E. coli* O157:H7 and naturally-contaminating *E. coli* O157:H7 on spinach under field conditions and following postharvest handling. Food Microbiol 77:173-184.
18. Harel O, Perkins N, Schisterman EF. 2014. The use of multiple imputations for data subject to limits of detection. Sri Lankan J Appl Stat 5(4):227-246.
19. Hutchison ML, Avery SM, Monaghan JM. 2008. The air-borne distribution of zoonotic agents from livestock waste spreading and microbiological risk to fresh produce from contaminated irrigation sources. J Appl Microbiol 105:848-857.
20. Islam M, Doyle MP, Phatak SC, Millner P, Jiang X. 2004. Persistence of Enterohemorrhagic *Escherichia coli* O157:H7 in soil and on leaf lettuce and parsley grown in fields treated with contaminated manure composts or irrigation water. J Food Prot 67(7):1365-1370.
21. Jarvis B, Corry JEL, Hedges AJ. 2007. Estimates of measurement uncertainty from proficiency testing schemes, internal laboratory quality monitoring and during routine enforcement examination of foods. J Appl Microbiol 103:462-467.
22. Jeamsripong S, Chase JA, Jay-Russell MT, Buchanan RL, Atwill ER. 2019. Experimental in-field transfer and survival of *Escherichia coli* from animal feces to romaine lettuce in Salinas Valley, California. Microorganisms 7(408).
23. Lee D, Tertuliano M, Harris C, Vellidis G, Levy K, Coolong T. 2019. *Salmonella* survival in soil and transfer onto produce via splash events. J Food Prot 82(12):2023-2037.
24. Lopez-Velasco G, Sbodio A, Tomas-Callejas A, Wei P, Tan KH, Suslow TV. 2012. Assessment of root uptake and systematic vine-transport of *Salmonella enterica* sv. Typhimurium by melon (*Cucumis melo*) during field production. Int J Food Microbiol 158:65-72.
25. Lopez-Velasco G, Tomas-Callejas A, Sbodio AO, Pham X, Wei P, Diribsa D, Suslow TV. 2015. Factors affecting cell population density during enrichment and subsequent molecular detection of *Salmonella enterica* and *Escherichia coli* O157:H7 on lettuce contaminated during field production. Food Control 54:165-175.
26. McKellar RC, Perez-Rodriguez F, Harris LJ, Moyne AL, Blais B, Topp E, Bezanson G, Bach S, Delaquis P. 2014. Evaluation of different approaches for modeling *Escherichia coli* O157:H7 survival on field lettuce. Int J Food Microbiol 184:74-85.
27. Moyne AL, Sudarshana MR, Blessington T, Koike ST, Cahn MD, Harris LJ. 2011. Fate of *Escherichia coli* O157:H7 in field-inoculated lettuce. Food Microbiol 28:1417-1425.
28. Moyne, A. L., Harris, L. J., & Marco, M. L. 2013. Assessments of total and viable *Escherichia coli* O157: H7 on field and laboratory grown lettuce. PloS one *8*(7).
29. Moyne AL, Blessington T, Williams TR, Koike ST, Cahn MD, Marco ML, Harris LJ. 2020. Conditions at the time of inoculation influence survival of attenuated *Escherichia coli* O157:H7 on field-inoculated lettuce. Food Microbiol 85:103274.
30. Munther DS, Carter MQ, Aldric CV, Ivanek R, Brandl MT. 2019. Formation of E. coli O157:H7 persister cells in the lettuce phyllosphere and application of differential equation models to predict their prevalence on lettuce plants in the field. Appl Environ Microbiol 01602-19.
31. Oliveira M, Vinas I, Usall J, Anguera M, Abadias M. 2012. Presence and survival of *Escherichia coli* O157:H7on lettuce leaves and in soil treated with contaminated compost and irrigation water. Int J Food Microbiol 156:133-140.

32. Park S, Szonyi B, Gautam R, Nightingale K, Aniciso J, Ivanek R. 2012. Risk factors for microbial contamination in fruits and vegetables at the preharvest level: a systematic review. J Food Prot 75(11):2055-2081.
33. Therneau T, Atkinson B. 2019. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. https://CRAN.R-project.org/package=rpart
34. Tomas-Callejas A, Lopez-Velasco G, Camacho AB, Artes F, Artes-Hernandez F, Suslow TV. 2011. Survival and distribution of *Escherichia coli* on diverse fresh-cut baby leafy greens under preharvest through postharvest conditions. Int J Food Microbiol 151:216-222.
35. U.S. Food and Drug Administration. 2015. Standards for the Growing, Harvesting, Packing, and Holding of Produce for Human Consumption. Final Rule, Docket No. FDA– 2011–N–0921, Accessed on December 15, 2019. https://www.federalregister.gov/documents/2015/11/27/2015-28159/standards-for-the-growing-harvesting-packing-and-holding-of-produce-for-human-consumption
36. van Buuren S, Groothuis-Oudshoorn K. 2011. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. URL https://www.jstatsoft.org/v45/i03/.
37. Venables WN, Ripley BD. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.
38. Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA, Tiedje JM, Whittam TS. 2009. Cryptic lineages of genus *Escherichia.* Appl Environ Microbiol 75(20):6534-6544.
39. Wall GL, Clements DP, Fisk CL, Stoeckel DM, Woods KL, Bihn EA. 2019. Meeting report: keys outcomes from a collaborative summit on agricultural water standards for fresh produce. Compr Rev Food Sci Food Saf 18:723-737.
40. Weller DL, Kovac J, Roof S, Kent DJ, Tokman JI, Kowalcyk B, Oryang D, Ivanek R, Aceituno A, Sroka C, Wiedmann M. 2017. Survival of *Escherichia coli* on lettuce under field conditions encountered in the northeastern United States. J Food Prot 80(7):1214-1221.
41. Wood JD, Bezanson GS, Gordon RJ, Jamieson R. 2010. Population dynamics of *Escherichia coli* inoculated by irrigation into the phyllosphere of spinach grown under commercial production conditions. Int J Food Microbiol 143:198-204.