DOI: 10.1111/risa.13491

A Predictive Model for Survival of *Escherichia coli* O157:H7 and Generic *E. coli* in Soil Amended with Untreated Animal Manure

Hao Pang ^(D),^{1,2,†} Amir Mokhtari,^{1,5,†} Yuhuan Chen,¹ David Oryang,¹ David T. Ingram,³ Manan Sharma,⁴ Patricia D. Millner,⁴ and Jane M. Van Doren^{1,*}

ABSTRACT: This study aimed at developing a predictive model that captures the influences of a variety of agricultural and environmental variables and is able to predict the concentrations of enteric bacteria in soil amended with untreated Biological Soil Amendments of Animal Origin (BSAAO) under dynamic conditions. We developed and validated a Random Forest model using data from a longitudinal field study conducted in mid-Atlantic United States investigating the survival of Escherichia coli O157:H7 and generic E. coli in soils amended with untreated dairy manure, horse manure, or poultry litter. Amendment type, days of rain since the previous sampling day, and soil moisture content were identified as the most influential agricultural and environmental variables impacting concentrations of viable E. coli O157:H7 and generic E. coli recovered from amended soils. Our model results also indicated that E. coli O157:H7 and generic E. coli declined at similar rates in amended soils under dynamic field conditions. The Random Forest model accurately predicted changes in viable E. coli concentrations over time under different agricultural and environmental conditions. Our model also accurately characterized the variability of E. coli concentration in amended soil over time by providing upper and lower prediction bound estimates. Crossvalidation results indicated that our model can be potentially generalized to other geographic regions and incorporated into a risk assessment for evaluating the risks associated with application of untreated BSAAO. Our model can be validated for other regions and predictive performance also can be enhanced when data sets from additional geographic regions become available.

KEY WORDS: Biological soil amendments of animal origin; enteric bacteria; predictive model; random forest

1. INTRODUCTION

Untreated biological soil amendments of animal origin (BSAAO) is a potential source of produce contamination, as it is a known reservoir

- ¹Center for Food Safety and Applied Nutrition, Food and Drug Administration, Office of Analytics and Outreach, College Park, MD, USA.
- ²Joint Institute for Food Safety and Applied Nutrition, University of Maryland, College Park, MD, USA.

³Center for Food Safety and Applied Nutrition, Food and Drug Administration, Office of Food Safety, College Park, MD, USA.

⁴U.S. Department of Agriculture, Agricultural Research Service, Northeast Area, Beltsville Agricultural Research Center, Environmental Microbial and Food Safety Laboratory, Beltsville, MD, USA. ⁵Current address: Booz Allen Hamilton, 4747 Bethesda Ave, Bethesda, MD 20814, USA.

*Address correspondence to Jane M. Van Doren, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, 5001 Campus Drive, College Park, MD 20740, USA; tel: 240-402-2927; jane.VanDoren@fda.hhs.gov.

[†]These authors contributed equally to this work.

for enteric bacteria including foodborne pathogens such as Escherichia coli O157:H7 and Salmonella (FAO/WHO, 2008; FDA, 2015). Once introduced into the soil through application of BSAAO, these pathogens have been shown to survive (Jiang, Morgan, & Doyle, 2002) and can lead to contamination of produce growing on the field (FDA, 2015), via various mechanisms; for example, direct contact (FAO/WHO, 2008) or splash from irrigation or rainfall (Cevallos-Cevallos, Danyluk, Gu, Vallad, & van Bruggen, 2012; Cevallos-Cevallos, Gu, Danyluk, & van Bruggen, 2012). Subsequently, pathogens may persist through the produce production chain and consumption of produce contaminated with pathogens poses a food safety risk to consumers (Atwill et al., 2015; Chase, Partyka, Bond, & Atwill, 2019). Understanding the impact of agricultural and environmental variables on enteric pathogen concentrations in amended soil is essential for assessing risk of foodborne illness associated with consumption of fresh produce grown in amended soils, as published studies typically indicated such variables can significantly impact enteric pathogen concentrations in amended soil (Fremaux et al., 2007; Jiang et al., 2002; Nicholson, Groves, & Chambers, 2005; van Elsas, Semenov, Costa, & Trevors, 2011). For example, enteric pathogen concentrations in amended soil can be affected by environmental variables, such as soil moisture, temperature, and precipitation (Cools, Merckx, Vlassak, & Verhaegen, 2001; Jiang et al., 2002; Moynihan, Richards, Ritz, Tyrrel, & Brennan, 2013; Park, Pachepsky, Shelton, Jeong, & Whelan, 2016; van Elsas et al., 2011; Vidovic, Block, & Korber, 2007; Williams, LeJeune, & McSpadden Gardener, 2015). In addition, E. coli O157:H7 concentrations in amended soil are affected by agricultural variables, such as manure type (Genereux, Breton, Fairbrother, Fravalo, & Cote, 2015; Rogers et al., 2011; Sharma et al., 2016), soil type (Franz et al., 2008; Nicholson et al., 2005), soil physical and chemical properties (Erickson et al., 2014; Ma et al., 2011; Ma, Ibekwe, Crowley, & Yang, 2012; Mubiru, Coyne, & Grove, 2000), and manure application methods (Avery, Hill, Killham, & Jones, 2004).

Predictive modeling is a valuable tool for providing estimates of microbial population dynamics (Esser, Leveau, & Meyer, 2015; McMeekin, Olley, Ratkowsky, & Ross, 2002; Peleg & Corradini, 2011). Models with various levels of complexity have been proposed to describe changes in concentrations of enteric bacteria in amended soil (Bezanson et al., 2012; Franz et al., 2008; Ma et al., 2012; Ma, Mark Ibekwe, Crowley, & Yang, 2014; Ongeng et al., 2014; Ongeng, Muyanja, Geeraerd, Springael, & Ryckeboer, 2011). Primary survival models, which describe changes in microbial numbers over time under particular environmental conditions, have been widely reported in the literature. Examples include the classical log-linear model, which has been used to fit survival curves of E. coli O157:H7 and/or Salmonella enterica in manure and/or manure-amended soil (Bolton, Byrne, Sheridan, McDowell, & Blair, 1999; Ongeng et al., 2011; Ongeng et al., 2011; You et al., 2006), as well as the Weibull model, which can be fitted to survival data that exhibit simple nonlinear deviations, such as concave and convex shapes (Bezanson et al., 2012; Franz et al., 2008; Ma et al., 2012; Ma et al., 2014; Naganandhini, Kennedy, Uyttendaele, & Balachandar, 2015; Ongeng et al., 2011; Wang et al., 2014; Yao et al., 2013; Zhang et al., 2013). Secondary models also have been proposed to describe the response of one or more parameters of a primary model to changes in variables of interest (Coroller, Jeuge, Couvert, Christieans, & Ellouze, 2015; Farakos, Frank, & Schaffner, 2013; Ratkowsky, Ollev, Mcmeekin, & Ball, 1982). Reviews and discussions of the limitations of selected primary and secondary models for fate of enteric pathogens in the agricultural environment are provided elsewhere (Ongeng et al., 2014).

There are important limitations with regard to the application of currently used survival models for enteric bacteria. Specifically, the underlying primary models typically assume a particular functional form (e.g., log-linear, convex, or concave curves) for survival patterns of enteric pathogens in amended soil over time. However, due to the complexity of the survival patterns, such models cannot always establish a statistically good fit as measured by the coefficient of determinations (R^2) or the root mean squared error (RMSE). Furthermore, the underlying primary models typically do not consider oscillations of enteric-pathogen concentrations in manureamended soil often observed in survival data sets (Ongeng et al., 2011; Semenov, Franz, van Overbeek, Termorshuizen, & van Bruggen, 2008; Vidovic et al., 2007), and, hence, fail to accurately characterize the survival behavior of pathogens under agricultural field environments (Ongeng et al., 2014).

The purpose of this work was to develop and validate a nonparametric predictive model that can provide accurate quantitative estimates of enteric bacteria concentrations (log₁₀cfu/g) in amended soil with BSAAO under different agricultural and

environmental conditions. Data were assembled for our study using a longitudinal field experiment conducted between 2011 and 2014 in mid-Atlantic United States (Sharma et al., 2019). We used the Random Forest approach (Breiman, 2001) to capture the underlying relationships among a variety of dynamic environmental variables (e.g., ambient temperature and soil moisture content), agricultural variables (e.g., amendment type and depth of amendment application), and observed enteric bacteria concentrations in amended soil over time. We used sensitivity analysis to identify influential variables and characterize their relationship with bacteria concentrations in amended soil. We evaluated the proposed model's performance in capturing distinct observed bacteria survival patterns over time and in predicting pathogen survival behavior in locations other than those for which data were available.

2. MATERIALS AND METHODS

2.1. Data Description

A detailed description of the experimental design and microbial analysis is available in Sharma et al. (2019). Briefly, data were assembled for this study using a longitudinal field experiment conducted between 2011 and 2014 investigating the survival of E. coli in manure-amended soils at three different experiment sites over 12 different seasonal trials: Southeast Agricultural Research & Extension Center, in Manheim, PA; U.S. Department of Agriculture Beltsville Area Research Center in Beltsville, MD (BARC); and University of Maryland Eastern Shore, in Princess Anne, MD (UMES). Manure was further collected from the following sources: dairy manure solids and liquids, USDA-Agricultural Research Service, Beltsville, MD; horse manure, University of Maryland; and poultry litter, University of Maryland Eastern Shore poultry houses. For each of the 12 seasonal trials conducted, rifampicin-resistant (Rif^R) attenuated E. coli O157:H7 strains (attPTVS 154 and attPTVS 155) and Rif^R nonpathogenic E. coli strains (TVS 353, 354, and 355) cultures prepared in poultry litter extract were applied to soils at either a high inoculum (10⁶ CFU/ml) or low inoculum (10⁴ CFU/ml) level. Following a randomized complete block experimental design (n = 4), manure was applied at an application rate of 5 tons per acre (4.54 metric tons per 0.4 ha) and evenly distributed over the experimental field plots, with ad-

ditional unamended plots as controls. Experimental plots containing manure were spray-inoculated with either low or high inocula; then plots were either left surface-amended or mechanically tilled to a depth of 10 cm. Soil samples were taken periodically from the experiment plots and populations of E. coli were enumerated using both direct plating and most probable number (MPN) methods (Sharma et al., 2019). The detection limit for MPN assay was $-0.24 \log_{10} MPN/g$. When E. coli O157:H7 and generic E. coli concentrations fell below the MPN detection limit, bag enrichment was performed to determine the presence or absence of E. coli O157:H7 and generic E. coli. The detection limit for the bag enrichment was -0.52log₁₀ CFU/g. Additional data representing potential influential agricultural and experiment variables were also collected, including experiment location, season (determined by the start date of each trial), soil management, amendment type, depth of amendment application, bacteria type, and inoculation levels. On each sampling day, in addition to levels of E. coli in amended soils, other environmental data variables were also collected, including soil moisture content, ambient temperature, and amount of precipitation. If weather data were not reported for a specific experiment location, we identified the nearest weather station and retrieved the ambient temperature as well as precipitation during the experiment timeframe using the National Climatic Data Center (http://www.ncdc.noaa.gov/).

2.2. Data Preparation

As described in Sharma et al. (2019), a total of 324 survival profiles (agricultural and environmental field conditions) were observed, determined by unique combinations of the experimental variables, including site, year, season, inoculum level, E. coli type, amendment type, soil management, and depth of manure application. For each of the 324 survival profiles, E. coli concentrations were measured from samples taken from four replicate field plots under the same agricultural and environmental conditions, yielding a total of 1,296 replicate field plots with 13,318 data observations. For each replicate field plot, soil samples were collected periodically until concentrations of E. coli O157:H7 or generic E. coli declined below the bag enrichment detection limit of $-0.52 \log_{10}$ CFU/g. When concentrations below the detection limit were observed in two successive scheduled sampling events, no additional samples for those plots were collected and analyzed, assuming concentrations of *E. coli* O157:H7 or generic *E. coli* would remain below limit of detection afterward. To address the uncertainty associated with the nondetect values, we assigned a random value between the detection limit ($-0.52 \log_{10} \text{ CFU/g}$) and $1 \log_{10}$ below the detection limit ($-1.52 \log_{10} \text{ CFU/g}$) to each of the nondetect observations.

2.3. Random Forest Model Development

We developed a predictive model using Random Forest to describe the response of both E. coli O157:H7 and generic E. coli concentrations (hereafter referred to as E. coli concentrations) in amended soil (log₁₀ CFU/g) to changes observed in the agricultural and environmental conditions over time described by the explanatory variables listed in Table I. Random Forest, first proposed by Breiman (2001), is an ensemble method consisting of thousands of classification and regression trees (CART) combined to obtain a prediction of the outcome of interest (Jones & Linder, 2015). During the construction of a Random Forest, data are randomly selected using bootstrap method, and then nonpruned classification and regression trees are built with a number of variables randomly selected at each node to determine the best split. By utilizing these two randomizations during model construction, Random Forest predictions are robust to overfitting when new data are introduced, which makes it preferable compared to the CART method that involves a single tree (Jones & Linder, 2015; Liaw & Wiener, 2002; Philibert et al., 2011; Prasad, Iverson, & Liaw, 2006). Furthermore, as a nonparametric approach, Random Forest does not require distributional assumptions (e.g., normality) and has the ability to deal with nonlinearity and interactions in complex data sets (Jones & Linder, 2015), making it a powerful alternative to traditional approaches, such as multivariate regressions.

Observed *E. coli* concentrations in collected amended soil samples demonstrated two types of variability: (i) between-profile variability representing observed variability in *E. coli* concentrations in amended soil under different agricultural and environmental conditions; and (ii) within-profile variability representing observed variability in *E. coli* concentrations in amended soil between replicate field plots associated with each particular survival profile under the same agricultural and environmental conditions. To quantify the overall survival pattern as well as between- and within-profiles variability in *E. coli* concentrations in amended soil, we developed three Random Forest models to predict average, lower bound, and upper bound of *E. coli* concentrations on different days post soil amendment application. The former model was built using the average *E. coli* concentrations $(\log_{10}cfu/g)$ in amended soil on each sampling day across all different replicate field plots associated with each of the 324 observed survival profiles. The latter two Random Forest models were built using the 2.5th and 97.5th percentiles of the observed *E. coli* concentrations $(\log_{10}cfu/g)$ in amended soil on each sampling day across all different replicate field plots. These two models represent within-profile variability in *E. coli* concentrations in amended soil under different agricultural and environmental conditions.

We developed the Random Forest models in R using the "randomForest" package (Liaw & Wiener, 2002) each with a forest size of 100,000 trees. To train the forest, we randomly selected 70% of the 324 survival profiles, including all values associated with the explanatory variables on different sampling days as well as the corresponding concentration values. The remaining part of the profiles was later used for validating the Random Forest model. Recursive feature elimination method was used to select the optimal set of variables to train the final model.

2.4. Model Sensitivity Analysis

The impact of the explanatory variables on the Random Forest model response (e.g., average *E. coli* concentrations in amended soil) was evaluated using the permutation importance scores (Liaw & Wiener, 2002). The permutation importance scores reflect the decrease in prediction accuracy resulting from permutation of an explanatory variable. If permuting the value of an explanatory variable does not affect (or increase) the prediction accuracy, then the variable is not related to the model response. However, if permuting the value decreases the prediction accuracy of the model, the variable is related to the model response, and the larger the decrease, the stronger the relationship (Genuer, Poggi, & Tuleau-Malot, 2010; Jones & Linder, 2015).

We also generated partial dependence plots (PDPs), used to interpret complex algorithms to investigate how influential explanatory variables were related to the Random Forest model response (Friedman, 2001). PDPs are graphical visualizations of the marginal effect of a given explanatory variable (e.g., inoculation level) or multiple explanatory variables (e.g., inoculation level and

Explanatory Variables	Units	Observed Values	
Depth of manure application	Unitless	Surface or tillage	
Soil management	Unitless	Organic or conventional	
Amendment type	Unitless	Dairy, poultry, horse, unamended	
Experiment site	Unitless	BARC, PA, UMES	
Season	Unitless	Spring, summer, fall, winter	
E. coli type	Unitless	E. coli O157:H7 or generic E. coli	
Inoculation level	log (cfu/g)	0.1–7.3	
Sampling day (i.e., days postinoculation)	Days	0–219	
Soil moisture content on the sampling day	%	0-54.8	
Ambient temperature on the sampling day	°C	0.6–30	
Average ambient temperature over the week prior to the sampling day	°C	-0.3-28.2	
Average ambient temperature since the previous sampling day	°C	0.8-28.25	
Precipitation on the sampling day	Mm	0–96.7	
Average daily precipitation over the week prior to the sampling day	Mm	0–27.7	
Average daily precipitation since the previous sampling day	Mm	0–49.8	
Days of rain during the week prior to the sampling day	Days	0–6	
Days of rain since the previous sampling day	Days	0–37	

Table I. Agricultural and Environmental Explanatory Variables Used in the Survival Model

days postinoculation) on a Random Forest model response (e.g., average *E. coli* concentrations). Typically, PDPs are restricted to only one or two explanatory variables to foster understanding, but as a result, PDPs can be misleading when there are significant hidden higher order interactions. Despite this, PDPs can still be extremely useful for knowledge discovery in large data sets, especially when the Random Forest is dominated by lower order interactions and main effects. Statistical explanation of the PDPs is provided elsewhere (Genuer et al., 2010). PDPs were generated in R using the "partialPlot" function in the "randomForest" package (Liaw & Wiener, 2002).

2.5. Model Validation

To validate the Random Forest models, we used the portion of the survival profiles that was not initially used in the training step (remaining 30% of the 324 survival profiles). We compared the predicted model response values (e.g., average E. coli concentrations in amended soil) in our validation data set with the observed values from the survival data set, for the same sampling days. We further calculated the prediction error (PE) as the difference between observed and predicted values. The acceptable prediction zone (APZ) approach was then followed to determine the predictive performance of the developed Random Forest models (Oscar, 2005). The APZ was established to be between $-1.0 \log_{10}$ cfu/g (fail-safe) and 0.5 log₁₀cfu/g (fail-dangerous), and the percentage of PEs in the APZ was used as a measurement of predictive performance. A model with a percentage of PEs in the APZ over 70% is considered to provide acceptable and validated predictions (Oscar, 2009). Model performances were also evaluated by regressing the predicted model response values (e.g., average *E. coli* concentrations) against the observed values (in log-scale) and comparing against the 1:1 line. Additionally, the coefficient of determinations (R^2) and the normalized RMSE (NRMSE) were used to measure the model performance. NRMSE was calculated as the RMSE divided by the difference between the maximum and minimum observed *E. coli* concentration (log₁₀cfu/g). NRMSE was expressed as percentage and an NRMSE value close to 0 indicates a good predictive performance.

To test the robustness of the model predictions, we repeated the validation step 2,000 times. In each repeat, a different set of 70% of the 324 survival profiles were randomly selected as a training data set for developing all three Random Forest models discussed above, and the remaining 30% profiles were used as a testing data set for model validation. By repeating the validation 2,000 times, the variabilities in the goodness-of-fit measures (e.g., NRMSE, R^2 , and APZ) were quantified.

In order to investigate whether we could extrapolate the model predictions to other geographic regions, we conducted a cross-validation analysis based on the three experiment locations available in the survival data set. In cross-validation, we trained our Random Forest models using data from two of the three experiment locations (PA, BARC, and UMES) and using data from the other experiment location as the testing data set to evaluate the predictive performance of the corresponding models. This process was repeated three times, each time with a different experiment location as the testing data set.

3. RESULTS

3.1. Sensitivity Analysis Results

Fig. 1 shows the list of the explanatory variables, rank ordered based on the relative magnitude of their calculated permutation scores as a measure of variable importance with respect to their impacts on the predicted average E. coli concentrations in amended soil. Similar analysis was performed for the other two Random Forest models representing the 95% variability in within-profiles E. coli concentrations in amended soil (results not shown here). Our results indicated that days postinoculation and inoculation level were among the top variables with the highest permutation importance scores. Among the environmental variables, days of rain since the previous sampling day and soil moisture content had the largest effect on the predicted average E. coli concentrations in amended soil. Among agricultural variables, predicted average concentrations of E. coli were mostly influenced by amendment type.

Fig. 2 displays the PDPs for selected explanatory variables listed in Table I. The predicted average concentrations of E. coli O157:H7 and generic E. coli declined with time (i.e., days postinoculation) (Fig. 2A). As expected, higher E. coli O157:H7 and generic E. coli concentrations were predicted when amended soils were inoculated at higher concentration levels (Fig. 2B). For the environmental variables (Figs. 2C and D), higher predicted average E. coli O157:H7 and generic E. coli concentrations in amended soils were generally associated with higher soil moisture content and higher days of rain since the previous sampling day. For the agricultural variables, average concentrations of E. coli O157:H7 and generic E. coli were higher in soils amended with poultry litter, compared with soil amended with horse manure or dairy manure (Fig. 2E). In general, predicted average generic E. coli concentrations in amended soil are also slightly higher compared with E. coli O157:H7 concentrations. As shown in Fig. 2A, despite the slight difference in concentrations, the two curves were almost parallel, indicating generic E. coli and E. coli O157:H7 in amended soils declined over time at a similar rate. Similar parallel patterns were observed in Fig. 2B–D, suggesting the impact of agricultural and environmental variables on predicted *E. coli* concentrations in amended soils is similar between the two different *E. coli* strains.

We also investigated the two-way interaction effects between selected explanatory variables, including days postinoculation, inoculation level $(\log_{10} \text{cfu/g})$, and soil moisture content (%), using the two-dimensional PDPs. As expected, Fig. 3A shows considerable interaction effect between inoculation level and days postinoculation, with inoculation level having its strongest impact on predicted average concentrations of both E. coli O157:H7 and generic E. coli during the first 50 days, and then became less impactful over time. Similarly, there is an interaction effect between days postinoculation and soil-moisture content (Fig. 3B). The predicted average concentrations varied dramatically due to changes in the soil moisture content within the first 50 days. Beyond the first 50 days, changes in the soil moisture content became less impactful through time, resulting in smaller range of variability in predicted concentration of both E. coli strains.

3.2. Validation Results

From the 2,000 repeated validation, the NRMSE values of the predicted average E. coli concentrations in amended soils ranged between 5.2% and 8.6%, with a mean of 6.7%; the R^2 values ranged between 0.79 and 0.91, with a mean of 0.87; and the APZ values ranged between 81% and 92%, with a mean of 88%. Validations performed for the other two Random Forest models representing the 95% variability in within-profiles E. coli concentrations in amended soil also showed narrow ranges for NRMSE, R^2 , and APZ values (results not shown here). These narrow ranges indicate that the model validation is robust and not sensitive to the random partitioning of the survival data set used during the training step of the model development. For the remaining analyses discussed in the article, results were based on the Random Forest model with the lowest NRMSE value.

We found statistically significant correlations $(p < 0.05, R^2 = 0.89$ for both *E. coli* O157:H7 and generic *E. coli*) between the observed and predicted average concentrations, with no significant deviation from the 1:1 line (Fig. 4). Additionally, we compared the observed average *E. coli* concentrations in the validation data set with the predicted average *E. coli* concentrations from the Random Forest

Fig. 1. Variable importance scores for the explanatory variables included in the Random Forest model for predicting average concentration of *E. coli* O157:H7 and generic *E. coli* in the amended soil.



model (Fig. 5). Results showed that the predicted median and the interquartile values were in good agreement with the observed values for both E. coli strains. The validation results also showed that the Random Forest model has an NRMSE of 5.0% and an R^2 value of 0.90 when predicting the average concentration of E. coli O157:H7 and an NRMSE of 5.5% and an R^2 value of 0.90 when predicting the average concentration of generic E. coli. When compared with the observed values, 94% and 91% of the predicted values for E. coli O157:H7 and generic E. coli concentrations were within the acceptable failsafe and fail-dangerous zone (-1.0 to $0.5 \log_{10}$ cfu/g), respectively, which is above the threshold for an acceptable predictive performance (70%) (Oscar, 2005, 2009). In addition, the developed upper and lower bound Random Forest models also showed good performance in predicting the range of E. coli concentrations in amended soil under different conditions (Table II).

In cross-validation, the NRMSE, R^2 , and APZ values of the predicted average concentrations of *E. coli* in amended soils ranged from 9% to 13.5%, 0.60 to 0.75, and 66% to 82%, respectively, when

 Table II.
 Predictive Performance Measures for Random Forest

 Models Predicting (i) Average; (ii) Lower Bound; and (iii) Upper
 Bound Concentrations of *E. coli* in Amended Soil

	Prediction Performance Measures ^a			
Models	NRMSE (%)	R^2	APZ (%)	
Model (i): average	5.2	0.90	93	
Model (ii): lower bound	6.1	0.84	90	
Model (iii): upper bound	6.3	0.87	90	

^aNRMSE, normalized root mean squared error; R^2 , coefficient of determinations; APZ, acceptable prediction zone.

data from one experiment location were left out from the model development and used as a validation set (Table III). The developed models showed satisfactory predictive performance when the UMES data set was among the two data sets used for training the Random Forest model (e.g., R^2 value of 0.75 with an APZ value of 80% when PA and UMES data sets were used). The model performance was less satisfactory when this experiment location was not used to train the model. No survival trial was conducted during the summer season at PA



Fig. 2. Partial dependence plots for selected explanatory variables: (A) days postinoculation; (B) inoculation level; (C) days of rain since the previous sampling day; (D) soil moisture content; (E) amendment type; and (F) *E. coli* type.

and BARC experiment locations, whereas survival trials were conducted in all four seasons at UMES. Thus, the Random Forest model developed using data from the PA and BARC experiment locations was not trained with information required to predict *E. coli* survival patterns during summer. Overall,

predictions from the cross-validation models were less accurate compared to the model trained using data from all three experiment locations (Tables II and III). Reduction in prediction accuracy when two of the three experiment locations were used for model training can also be attributed to the smaller



Fig. 3. Two-variable partial dependence plots representing interaction effects between: (A) inoculation level and days postinoculation; and (B) soil moisture content (%) and days postinoculation on predict average *E. coli* concentrations in amended soils.

size of the training data set compared to the model trained using all three experiment locations.

3.3. Evaluating the Predictive Model Performance in Capturing Distinct Survival Patterns

As indicated in Sharma et al. (2019), the observed *E. coli* concentrations in amended soil from the field experiment showed a variety of survival patterns with respect to time and most survival profiles fell into one of six patterns: (i) exponential decline; (ii) initial increase and exponential decline; (iii) exponential decline with intermediate peak; (iv) exponential decline with one moderate peak; (v) initial increase with exponential decline and one moderate peak; (vi) initial increase with exponential decline and two moderate peaks. We evaluated the performance of our Random Forest models in capturing these distinct survival patterns over time. Fig. 6 shows a comparison between the observed and predicted concentrations of *E. coli* O157:H7 and generic *E. coli* in amended soil for selected profiles (excluded from data set used to develop the models) that showed the distinct survival patterns mentioned above. Fig. 6A shows the Random Forest model predictions for a particular survival profile that

Training Locations ^a	Validation Location ^a	Prediction Performance Measures ^b		
		NRMSE	R^{b}	APZ
PA & UMES	BARC	9.0%	0.75	80%
PARC & UMES	PA	9.1%	0.74	82%
PA & BARC	UMES	13.5%	0.60	66%

Table III. Predictive Performance Measures from Cross-Validation

^aIn each cross-validation, data from one experiment location were left out from model development and used as the validation data set. The other two data sets were used for training the Random Forest model.

^bNRMSE, normalized root mean squared error; R^2 , coefficient of determinations; APZ, acceptable prediction zone. Values were calculated based on the predictive average concentrations in amended soil.



Fig. 4. Plot of the correlation between predicted and observed average concentrations of: *E. coli* O157:H7 (A) and generic *E. coli* (B) in amended soil.

showed an exponential decline observed under the following condition: generic E. coli in conventional soil amended with surface applied dairy manure at UMES experiment location in winter season. A total of eight replicate trials were conducted for this survival profile. As shown in Fig. 6A, the predicted average E. coli concentrations were in good agreement with the observed concentrations in this survival profile and the within-profile variability across the eight replicate trials were captured as most of the observations fell inside the predicted upper and lower bound concentrations. Similar observations were found for other survival profiles (Figs. 6B-F). Overall, the Random Forest models were able to capture the complex, sometimes nonmonotonic relationships and changes in concentrations observed in survival profiles over time. In addition, the Random Forest models also accurately characterized the within-profile variability of E. coli concentrations in amended soil.

4. DISCUSSION

Predictive models, which here are used to characterize the relationships among different agricultural and environmental variables and concentrations of enteric bacteria in BSAAO-amended soils, are an integral part of any risk assessment effort to better understand public health risks associated with application of untreated BSAAO. Although the fate of enteric bacteria in amended soils has been the subject of many previous works, only a few have developed predictive models to investigate the impact of agricultural and environmental variables and to predict the concentration of enteric bacteria over time under dynamic conditions (Franz, Schijven, de RodaHusman, & Blaak, 2014; Ongeng, Muyanja, Ryckeboer, Springael, & Geeraerd, 2011; Park et al., 2016; Semenov, Franz, & van Bruggen, 2010).

We report here a predictive model for the fate of E. coli O157:H7 and generic E. coli in amended soil, which determined the relationship between a variety of variables and the persistence of the E. coli under dynamic agricultural and environmental conditions in the U.S. mid-Atlantic region. Many commonly used primary predictive models are static in nature and thus are not appropriate for predicting survival observed under dynamic environmental conditions (Franz, van Hoek, Bouw, & Aarts, 2011; Ongeng et al., 2014; Ongeng et al., 2011; Ongeng et al., 2014). For example, Ongeng et al. (2011) investigated the survival of Salmonella in manureamended soil under tropical agricultural field conditions. Despite being able to develop a predictive model based on laboratory experiments performed under constant temperature conditions, the isothermally derived model underestimated the survival of Salmonella in manure-amended soil under dynamic conditions (Ongeng et al., 2011). This indicated that survival of enteric bacteria in amended soils under dynamic conditions can be influenced by changing environmental conditions, and thus cannot be accurately estimated from survival observed under isothermal conditions (Ongeng et al., 2011; Ongeng et al., 2011; Semenov et al., 2010). By using Random Forest, we were able to model and evaluate the effect of the observed dynamic changes in environmental conditions (e.g., temperature and precipitation) on the complex E. coli survival patterns(e.g., nonmonotonic) in amended soil.

Predictive models derived from monotonic primary models (e.g., log-linear and Weibull survival models) do not consider oscillations in bacteria concentrations, thus are not suitable for prediction of pathogen survival in agricultural environment where such oscillations are often observed (Ongeng et al., 2014). In an effort to overcome this limitation, Semenov et al. (2010) developed a COLIWAVE simulation model to predict the survival of E. coli O157:H7 in manure-amended soils under dynamic environmental conditions. By separating the relative growth and death rates, the COLIWAVE model was able to mimic the oscillations (nonmonotonic behavior) in E. coli O157:H7 observed in the survival data and improved the prediction accuracy by 10-15% comparing to that of the monotonic Weibull model (Semenov et al., 2010). As a semi-mechanistic model, COLIWAVE model was developed based on variables corresponded to existing mechanistic knowl-

edge about microbial survival. Although a mechanistic model can describe known causal relationships between selected variables and pathogen survival in amended soil, it typically involves simplified assumptions that may not hold true under realistic and highly variable field conditions (Baker, Pena, Jayamohan, & Jerusalem, 2018; Ongeng et al., 2014; Semenov et al., 2010). Additionally, the specific nature of mechanistic models prohibits the inclusion of potential influential variables (where underlying mechanisms of their impact were not fully established) to achieve more universal predictions (Baker et al., 2018; Ongeng et al., 2014). Machine learning models, such as Random Forest, focus on providing accurate predictions through the use of large-scale data sets (Baker et al., 2018). Because of the availability of the large-scale survival data set collected through the comprehensive long-term field study conducted by Sharma et al. (2019), we were able to utilize the power of machine learning to improve the prediction of our model. Our predictive model also considered a comprehensive list of agricultural and environmental variables, including those that have been suggested for inclusion in predictive survival models, such as soil moisture (Franz et al., 2014; Semenov et al., 2010), as well as precipitation, manure-application methods, and amendment type (Park et al., 2016).

We were able to identify influential agricultural and environmental variables and characterized their impact on E. coli survival in amended soils using our predictive model. E. coli O157:H7 and generic E. coli concentrations in amended soils were shown to be related to the amendment type, with higher predicted concentrations of E. coli O157:H7 and generic E. coli in poultry-litter-amended soils, compared with concentrations in soil amended with dairy or horse manure. This finding is in agreement with previous work conducted in outdoor lysimeter systems, which observed a slower decline of E. coli O157:H7 in soil amended with poultry litter than in soil amended with cattle manure (Nyberg, Vinneras, Ottoson, Aronsson, & Albihn, 2010). In a published greenhouse study, poultry litter-amended soil was also reported to have supported significantly higher concentrations of E. coli O157:H7 and generic E. coli compared with soils amended with dairy manure liquids or horse manure (Sharma et al., 2016). The authors from these two studies suggested that higher nutrient content (nitrogen and phosphorus) provided by poultry litter may have supported the extended survival observed in poultry litter-amended soil (Nyberg et al., 2010; Sharma et al., 2016). Soil moisture

Fig. 5. Comparison of the average observed *E. coli* O157:H7 and generic *E. coli* concentrations in the survival data set with predicted values using the Random Forest model. The bottom and top of the boxes represent the first and the third quartiles, whereas the bars inside the whiskers represent the median values. The upper and lower ends of the whiskers (the vertical lines) represent the largest and smallest values in $1.5 \times IQR$ (inter-quartile range, the distance between the first and the third quartile), and the dots represent values outside the $1.5 \times IQR$.

content and days of rain since the previous sampling day were also identified as influential environmental factors affecting survival of E. coli in amended soil especially in the first 50 days after inoculation. Sharma et al. (2019), in a descriptive analysis of the same data set used in the current study, also showed in specific seasonal trials where initial soil moisture content was between 11% and 12.1% that E. coli survival was supported for longer durations than in trials that did not have a soil moisture content within this range. Our finding is also consistent with other studies that have demonstrated that soil moisture content (Mubiru et al., 2000; van Elsas et al., 2011; Vidovic et al., 2007; Williams et al., 2015) and rainfall (and high humidity associated with rain events) (Pang, McEgan, Micallef, & Pradhan, 2018; Pang, McEgan, Mishra, Micallef, & Pradhan, 2017; Park et al., 2015; Strawn et al., 2013; Weller, Wiedmann, & Strawn, 2015) are variables influencing the survival of pathogenic bacteria in (amended) soil.

As generic *E. coli* strains are sometimes used as surrogate for pathogens such as *E. coli* O157:H7, it is imperative to understand their behavior relative to *E. coli* O157:H7 in environmental reservoirs such as amended soils (Franz et al., 2014). Our model results indicated that *E. coli* O157:H7 generally declined over time at a rate similar to that of generic *E. coli* in amended soils. In addition, under dynamic field conditions, the impact of environmental factors on the survival of *E. coli* O157:H7 was similar compared to that of generic *E. coli*. These findings suggested that generic *E. coli* can be potentially used as a surrogate to predict survival patterns of *E. coli* O157:H7 in amended soils.

We also identified several areas that can be improved in our modeling approach. Our predictive model included 17 potential agricultural and environmental variables that may influence the survival of E. coli in amended soil under field conditions. However, model predictions could potentially be improved if additional agricultural and environmental variables, such as pH, soil type, activity and diversity of soil microbiota, and soil chemical properties, were included when such data become available. Optimally, we would develop a generalizable predictive survival model by training the model using a diverse data set that contains information from multiple different geographic regions across the United States. As such a comprehensive data set is not currently available, we investigated whether the current model was generalizable by conducting a cross-validation using the three geographic locations and the inherent heterogeneity in our survival data sets across different experiment locations. Results



Fig. 6. Comparison between the observed and the predicted concentrations of *E. coli* O157:H7 and generic *E. coli* in amended soil for the distinct patterns identified in the survival data set: (A) exponential decline; (B) initial increase and exponential decline; (C) exponential decline with intermediate peak; (D) exponential decline with one moderate peak; (E) initial increase with exponential decline and one moderate peak; (F) initial increase with exponential decline and two moderate peaks.

from model cross-validations based on the three experiment locations were generally satisfactory, indicating that the survival model could be potentially used in other geographic locations. Our results also demonstrated that the representativeness and size of the training data set were crucially impacting the survival model prediction accuracy. This indicates that the Random Forest model can be validated for other geographic regions and predictive performance can be enhanced by using additional data that contain information from different geographic regions when such data become available. In addition, Random Forest, as a machine learning algorithm, focuses on prediction but does not explicitly contribute to our knowledge of the underlying causal mechanisms for bacterial population dynamics in amended soil (Clark & Golder, 2015; Jones & Linder, 2015). By including variable importance and partial dependence measurements that demonstrate the relationship between influential variables and microbial survival in amended soil, our analysis provided insights about the underlying mechanisms and enhanced the interpretability of machine learning model results. As suggested by Baker et al. (2018), future research should be directed toward enabling a symbiotic relationship between mechanistic modeling and machine learning to embrace the strengths of both approaches.

In conclusion, we developed and validated a Random Forest model that can predict changes in concentrations of E. coli in amended soil observed over time under dynamic agricultural and environmental conditions. The modeling approach reported here also can be used to predict the survival of other enteric bacteria, such as Salmonella, in amended soils, if data elements similar to those reported in this article are available for the pathogens in question. Cross-validation results indicated that our model can be potentially generalized to other geographic regions and incorporated into a risk assessment model for evaluating the risks associated with application of untreated BSAAO. Data sets that contain information from additional geographic regions, when available, can be used to validation the Random Forest model and also can be incorporated to enhance model predictive performance.

ACKNOWLEDGMENTS

The extensive data set utilized for the generation and validation of this predictive model was funded through an Interagency Agreement between the USDA Agricultural Research Service (ARS) and the U.S. FDA (IAA 224-11-2046). We acknowledge Dr. Bryan Vinyardfrom USDA ARS for statistical consultation; Dr. FawzyHashem from University of Maryland Eastern Shore (UMES) for laboratory and field research contributions; Eric Handy, Cheryl East, Dr. Kate White, Russell Reynnells, Corrie Cotton, Dr. Lorna Graham, Kelly Domesle, and Dr. June deGraft Hansen for field and laboratory technical assistance. This research was supported in part by an appointment to the Research Participation Program at the U.S. Food and Drug Administration administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement (IAA 224-17-20935) between the U.S. Department of Energy and the U.S. Food and Drug Administration. Dr. Hao Pang was an ORISE research participant at the time this research was conducted. This research was also funded in part by the University of Maryland Joint Institute for Food Safety and Applied Nutrition through a cooperative agreement (#FDU001418) with the U.S. Food and Drug Administration.

REFERENCES

- Atwill, E. R., Chase, J. A., Oryang, D., Bond, R. F., Koike, S. T., Cahn, M. D., ... Dennis, S. (2015). Transfer of *Escherichia coli* 0157:H7 from simulated wildlife scat onto romaine lettuce during foliar irrigation. *Journal of Food Protection*, 78(2), 240–247.
- Avery, L. M., Hill, P., Killham, K., & Jones, D. L. (2004). Esherichia coli O157 survival following the surface and sub-surface application of human pathogen contaminated organic waste to soil. Soil Biology and Biochemistry, 36(12), 2101–2103.
- Baker, R. E., Pena, J. M., Jayamohan, J., & Jerusalem, A. (2018). Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, 14(5), 20170660
- Bezanson, G., Delaquis, P., Bach, S., McKellar, R., Topp, E., Gill, A., ... Gilmour, M. (2012). Comparative examination of *Escherichia coli* O157:H7 survival on romaine lettuce and in soil at two independent experimental sites. *Journal of Food Protection*, 75(3), 480–487.
- Bolton, D. J., Byrne, C. M., Sheridan, J. J., McDowell, D. A., & Blair, I. S. (1999). The survival characteristics of a non-toxigenic strain of *Escherichia coli* O157:H7. *Journal of Applied Microbiology*, 86(3), 407–411.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Cevallos-Cevallos, J. M., Danyluk, M. D., Gu, G., Vallad, G. E., & van Bruggen, A. H. (2012). Dispersal of *Salmonella* Typhimurium by rain splash onto tomato plants. *Journal of Food Protection*, 75(3), 472–479.
- Cevallos-Cevallos, J. M., Gu, G., Danyluk, M. D., & van Bruggen, A. H. (2012). Adhesion and splash dispersal of *Salmonella enteric* Typhimurium on tomato leaflets: Effects of rdar morphotype and trichome density. *International Journal of Food Mi crobiology*, 160(1), 58–64.
- Chase, J. A., Partyka, M. L., Bond, R. F., & Atwill, E. R. (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.
- Clark, W., & Golder, M. (2015). Big data, causal inference, and formal theory: Contradictory trends in political science? *PS: Political Science & Politics*, 48(1), 65–70.
- Cools, D., Merckx, R., Vlassak, K., & Verhaegen, J. (2001). Survival of *E. coli* and *Enterococcs* spp. derived from pig slurry in soils of different texture. *Applied Soil Ecology*, 17, 53–62.
- Coroller, L., Jeuge, S., Couvert, O., Christieans, S., & Ellouze, M. (2015). Extending the gamma concept to non-thermal inactivation: A dynamic model to predict the fate of *Salmonella* during the dried sausages process. *Food Microbiology*, 45, 266–275.
- Erickson, M. C., Habteselassie, M. Y., Liao, J., Webb, C. C., Mantripragada, V., Davey, L. E., & Doyle, M. P. (2014). Examination of factors for use as potential predictors of human

Predictive model for Escherichia coli survival in amended-soil

enteric pathogen survival in soil. *Journal of Applied Microbiology*, *116*(2), 335–349.

- Esser, D. S., Leveau, J. H., & Meyer, K. M. (2015). Modeling microbial growth and dynamics. *Applied Microbiology and Biotechnology*, 99(21), 8831–8846.
- FAO/WHO. (2008). *Microbiological hazards in fresh leafy* vegetables and herbs. Meeting Report. Retrieved from http://www.fao.org/3/a-i0452e.pdf
- Farakos, S. M. S., Frank, J. F., & Schaffner, D. W. (2013). Modeling the influence of temperature, water activity and water mobility on the persistence of *Salmonella* in low-moisture foods. *International Journal of Food Microbiology*, 166(2), 280–293.
- U.S. Food and Drug Administration (FDA). (2015). Final qualitative assessment of risk to public health from on-farm contamination of produce. Retrieved from https://www.fda.gov/downloads/Food/FoodScienceResearch/RiskSafetyAssessment/UCM470780.pdf
- Franz, E., Schijven, J., de RodaHusman, A. M., & Blaak, H. (2014). Meta-regression analysis of commensal and pathogenic *Escherichia coli* survival in soil and water. *Environmental Science & Technology*, 48(12), 6763–6771.
- Franz, E., Semenov, A. V., Termorshuizen, A. J., de Vos, O. J., Bokhorst, J. G., & van Bruggen, A. H. (2008). Manureamended soil characteristics affecting the survival of *E. coli* O157:H7 in 36 Dutch soils. *Environmental Microbiology*, 10(2), 313–327.
- Franz, E., van Hoek, A. H., Bouw, E., & Aarts, H. J. (2011). Variability of *Escherichia coli* O157 strain survival in manureamended soil in relation to strain origin, virulence profile, and carbon nutrition profile. *Applied and Environmental Microbi*ology, 77(22), 8088–8096.
- Fremaux, B., Prigent-Combaret, C., Delignette-Muller, M. L., Mallen, B., Dothal, M., Gleizal, A., & Vernozy-Rozand, C. (2007). Persistence of Shiga toxin-producing *Escherichia coli* O26 in various manure-amended soil types. *Journal of Applied Microbiology*, 104(1), 296–304.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.
- Genereux, M., Breton, M. J., Fairbrother, J. M., Fravalo, P., & Cote, C. (2015). Persistence of indicator and pathogenic microorganisms in broccoli following manure spreading and irrigation with fecally contaminated water: Field experiment. *Journal of Food Protection*, 78(10), 1776–1784.
- Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225–2236.
- Jiang, X., Morgan, J., & Doyle, M. P. (2002). Fate of Escherichia coli O157:H7 in manure-amended soil. Applied and Environmental Microbiology, 68(5), 2605–2609.
- Jones, Z., & Linder, F. (2015). Exploratory data analysis using random forests. Retrived from http://zmjones.com/ static/papers/rfss_manuscript.pdf
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2/3, 18–22.
- Ma, J., Ibekwe, A. M., Crowley, D. E., & Yang, C. H. (2012). Persistence of *Escherichia coli* O157:H7 in major leafy green producing soils. *Environmental Science & Technology*, 46(21), 12154–12161.
- Ma, J., Ibekwe, A. M., Yi, X., Wang, H., Yamazaki, A., Crowley, D. E., & Yang, C. H. (2011). Persistence of *Escherichia coli* 0157:H7 and its mutants in soils. *PLoS One*, 6(8), e23191.
- Ma, J., Mark Ibekwe, A., Crowley, D. E., & Yang, C. H. (2014). Persistence of *Escherichia coli* O157 and non-O157 strains in agricultural soils. *Science of the Total Environment*, 490, 822– 829.
- McMeekin, T. A., Olley, J., Ratkowsky, D. A., & Ross, T. (2002). Predictive microbiology: Towards the interface and beyond. *International Journal of Food Microbiology*, 73(2-3), 395–407.

- Moynihan, E. L., Richards, K. G., Ritz, K., Tyrrel, S. F., & Brennan, F. P. (2013). Impact of soil type, biology and temperature on the survival of non-toxigenic *Escherichia coli* O157. *Biology & Environment: Proceedings of the Royal Irish Academy*, 113B, 41–46.
- Mubiru, D. N., Coyne, M. S., & Grove, J. H. (2000). Mortality of *Escherichia coli* O157:H7 in two soils with different physical and chemical properties. *Journal of Environmental Quality*, 29(6), 1821–1825.
- Naganandhini, S., Kennedy, Z. J., Uyttendaele, M., & Balachandar, D. (2015). Persistence of pathogenic and non-pathogenic *Esherichia coli* strains in various tropical agricultural soils of India. *PLoS One*, 10(6), e0130038.
- Nicholson, F. A., Groves, S. J., & Chambers, B. J. (2005). Pathogen survival during livestock manure storage and following land application. *Bioresource Technology*, 96(2), 135–143.
- Nyberg, K. A., Vinneras, B., Ottoson, J. R., Aronsson, P., & Albihn, A. (2010). Inactivation of *Escherichia coli* O157:H7 and *Salmonella* Typhimurium in manure-amended soils studied in outdoor lysimeters. *Applied Soil Ecology*, 46(3), 398–404.
- Ongeng, D., Haberbeck, L. U., Mauriello, G., Ryckeboer, J., Springael, D., & Geeraerd, A. H. (2014). Modeling the fate of *Escherichia coli* O157:H7 and *Salmonella enterica* in the agricultural environment: Current perspective. *Journal of Food Science*, 79(4), R421–R427.
- Ongeng, D., Muyanja, C., Geeraerd, A. H., Springael, D., & Ryckeboer, J. (2011). Survival of *Escherichia coli* O157:H7 and *Salmonella* enteric serovar Typhimurium in manure and manure-amended soil under tropical climatic conditions in Sub-Saharan Africa. *Journal of Applied Microbiology*, 110(4), 1007– 1022.
- Ongeng, D., Muyanja, C., Ryckeboer, J., Springael, D., & Geeraerd, A. H. (2011). Kinetic model-based prediction of the persistence of *Salmonella enteric* serovar Typhimurium under tropical agricultural field conditions. *Journal of Applied Microbiology*, 110(4), 995–1006.
- Ongeng, D., Vasquez, G. A., Muyanja, C., Ryckeboer, J., Geeraerd, A. H., & Springael, D. (2011). Transfer and internalisation of *Escherichia coli* O157:H7 and *Salmonella enteric* serovar Typhimurium in cabbage cultivated on contaminated manureamended soil under tropical field conditions in Sub-Saharan Africa. *International Journal of Food Microbiology*, 145(1), 301–310.
- Oscar, T. P. (2005). Validation of lag, time and growth rate models for *Salmonella* typhimurium: Acceptable prediction zone method. *Journal of Food Science*, *70*(2), M129–M137.
- Oscar, T. P. (2009). General regression neural network and Monte Carlo simulation model for survival and growth of *Salmonella* on raw chicken skin as a function of serotype, temperature, and time for use in risk assessment. *Journal of Food Protection*, 72(10), 2078–2087.
- Pang, H., McEgan, R., Micallef, S. A., & Pradhan, A. K. (2018). Evaluation of meteorological factors associated with pre-harvest contamination risk of generic *Escherichia coli* in a mixed produce and dairy farm. *Food Control*, 85, 135–143
- Pang, H., McEgan, R., Mishra, A., Micallef, S. A., & Pradhan, A. K. (2017). Identifying and modeling meteorological risk factors associated with pre-harvest contamination of *Listeria* species in a mixed produce and dairy farm. *Food Research International*, 102, 355–363.
- Park, S., Navratil, S., Gregory, A., Bauer, A., Srinath, I., Szonyi, B., ... Ivanek, R. (2015). Multifactorial effects of ambient temperature, precipitation, farm management, and environmental factors determine the level of generic *Escherichia coli* contamination on preharvested spinach. *Applied and Environmental Microbiology*, 81(7), 2635–2650.
- Park, Y., Pachepsky, Y., Shelton, D., Jeong, J., & Whelan, G. (2016). Survival of manure-borne and fecal coliforms in soil:

Temperature dependence as affected by site-specific factors. *Journal of Environmental Quality*, 45(3), 949–957.

- Peleg, M., & Corradini, M. G. (2011). Microbial growth curves: What the models tell us and what they cannot. *Critical Reviews* in Food Science and Nutrition, 51(10), 917–945.
- Philibert, A., Desprez-Loustau, M.-L., Fabre, B., Frey, P., Halkett, F., Husson, C., ... Makowski, D. (2011). Predicting invasion success of forest pathogenic fungi from species traits. *Journal* of Applied Ecology, 48(6), 1381–1390.
- Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and Random Forests for ecological prediction. *Ecosystems*, 9(2), 181–199.
- Ratkowsky, D. A., Olley, J., Mcmeekin, T. A., & Ball, A. (1982). Relationship between temperature and growth-rate of bacterial cultures. *Journal of Bacteriology*, 149(1), 1–5.
- Rogers, S. W., Donnelly, M., Peed, L., Kelty, C. A., Mondal, S., Zhong, Z., & Shanks, O. C. (2011). Decay of bacterial pathogens, fecal indicators, and real-time quantitative PCR genetic markers in manure-amended soils. *Applied and Environmental Microbiology*, 77(14), 4839–4848.
- Semenov, A. V., Franz, E., & van Bruggen, A. H. C. (2010). COL-IWAVE a simulation model for survival of *E. coli* O157:H7 in dairy manure and manure-amended soil. *Ecological Modelling*, 221(4), 599–609.
- Semenov, A. V., Franz, E., van Overbeek, L., Termorshuizen, A. J., & van Bruggen, A. H. (2008). Estimating the stability of *Escherichia coli* O157:H7 survival in manure-amended soils with different management histories. *Environmental Microbiology*, 10(6), 1450–1459.
- Sharma, M., Millner, P. D., Hashem, F., Camp, M., Whyte, C., Graham, L., & Cotton, C. P. (2016). Survival and persistence of nonpathogenic *Escherichia coli* and attenuated *Escherichia coli* 0157:H7 in soils amended with animal manure in a greenhouse environment. *Journal of Food Protection*, 79(6), 913–921.
- Sharma, M., Millner, P. D., Hashem, F., Vinyard, B. T., East, C. L., Handy, E. T., ... Cotton, C. P. (2019). Survival of *Escherichia coli* in manure-amended soils is affected by spatiotemporal, agricultural, and weather factors in the mid-Atlantic

United States. *Applied and Environmental Microbiology*, 85(5), e02392–18

- Strawn, L. K., Fortes, E. D., Bihn, E. A., Nightingale, K. K., Grohn, Y. T., Worobo, R. W., ... Bergholz, P. W. (2013). Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. *Applied and Environmental Microbiology*, 79(2), 588–600.
- van Elsas, J. D., Semenov, A. V., Costa, R., & Trevors, J. T. (2011). Survival of *Escherichia coli* in the environment: Fundamental and public health aspects. *The ISME Journal*, 5(2), 173–183.
- Vidovic, S., Block, H. C., & Korber, D. R. (2007). Effect of soil composition, temperature, indigenous microflora, and environmental conditions on the survival of *Escherichia coli* O157:H7. *Canada Journal of Microbiology*, 53(7), 822–829.
- Wang, H., Zhang, T., Wei, G., Wu, L., Wu, J., & Xu, J. (2014). Survival of *Escherichia coli* O157:H7 in soils under different land use types. *Environmental Science and Pollution Research*, 21(1), 518–524.
- Weller, D., Wiedmann, M., & Strawn, L. K. (2015). Spatial and temporal factors associated with an increased prevalence of *Listeria monocytogenes* in spinach fields in New York State. *Applied and Environmental Microbiology*, 81(17), 6059–6069.
- Williams, M. L., LeJeune, J. T., & McSpadden Gardener, B. (2015). Soil conditions that can alter natural suppression of *Escherichia coli* O157:H7 in Ohio specialty crop soils. *Applied and Environmental Microbiology*, 81(14), 4634–4641.
- Yao, Z., Wei, G., Wang, H., Wu, L., Wu, J., & Xu, J. (2013). Survival of *Escherichia coli* O157:H7 in soils from vegetable fields with different cultivation patterns. *Applied and Environmental Microbiology*, 79(5), 1755–1756.
- You, Y., Rankin, S. C., Aceto, H. W., Benson, C. E., Toth, J. D., & Dou, Z. (2006). Survival of *Salmonella enteric* serovar Newport in manure and manure-amended soils. *Applied and Environmental Microbiology*, 72(9), 5777–5783.
- Zhang, T., Wang, H., Wu, L., Lou, J., Wu, J., Brookes, P. C., & Xu, J. (2013). Survival of *Escherichia coli* O157:H7 in soils from Jiangsu Province, China. *PLoS One*, 8(12), e81178.