Using farm management practices to predict *Campylobacter* prevalence in pastured poultry farms

Xinran Xu,[*] Michael J. Rothrock Jr.[†] Anand Mohan,[*] Govindaraj Dev Kumar,[‡] and Abhinav Mishra[*,1]

[*]*Department of Food Science and Technology, University of Georgia, Athens, GA, USA;* [†]*Egg Safety and Quality Research Unit, U.S. National Poultry Research Center, Agricultural Research Service, United States Department of Agriculture, Athens, GA, USA; and* [‡]*Center for Food Safety, University of Georgia, Griffin, GA, USA*

ABSTRACT Contamination of poultry products by *Campylobacter* is often associated with farm management practices and processing plant practices. A longitudinal study was conducted on 11 pastured poultry farms in southeastern United States from 2014 to 2017. In this study, farm practices and processing variables were used as predictors for a random forest **(RF)** model to predict *Campylobacter* prevalence in pastured poultry farms and processing environments. Individual RF models were constructed for fecal, soil and whole carcass rinse after processing **(WCR-P)** samples. The performance of models was evaluated by the area under curve **(AUC)** from the receiver operating characteristics curve. The AUC values were 0.902, 0.894, and 0.864 for fecal, soil, and WCR-P models, respectively. Relative importance plots were generated to predict the most important variable in each RF model. Animal source of feces was identified as the most important variable in fecal model and the soy content of the brood feed was the most important variable for soil model. For WCR-P model, the average flock age showed the strongest impact on RF model. These RF models can help pastured poultry growers with food safety control strategies to reduce *Campylobacter* prevalence in pastured poultry farms.

Key words: predictive microbiology, alternative poultry production, food safety, random forest, machine learning

INTRODUCTION

The U.S. Centers for Disease Control and Prevention **(CDC)** estimates that *Campylobacter* causes an estimated 1.5 million illnesses each year in the United States (CDC, 2019). *Campylobacter* spp. are Gram-negative spiral, rod-shaped, nonspore forming bacteria with polar flagella (Kaakoush et al., 2015). Major symptoms of *Campylobacter* infection are gastroenteritis, diarrhea, and sequelae such as Guillain-Barre syndrome (Sahin et al., 2003). From 2010 to 2017, among 236 reported foodborne *Campylobacter* outbreaks, 41 were associated with poultry products (CDC, 2019). In addition, the consumption of chicken and eggs linked to *Campylobacter* account for most laboratory-confirmed cases of bacterial gastroenteritis in the United States (Arsi et al., 2019). Poultry is identified as a major reservoir and source of transmission of campylobacteriosis (Kaakoush et al., 2015). After poultry products contaminated with *Campylobacter* are brought to a consumer's kitchen,

they can cross contaminate utensils and further infect consumers (Nauta et al., 2009). In general, *Campylobacter* contaminates poultry products before or during processing, surviving through the food supply chain to becoming a potential health risk to humans (Newell and Fearnley, 2003). It is important to identify the factors leading to *Campylobacter* transmission on poultry farms and processing plants. Feed, drinking water, soil, other farm animals, biosecurity threats (wildlife species), insects, farm equipment, employees, visitors, and farm vehicles are possible routes of *Campylobacter* transmission on poultry farms (Ghareeb et al., 2019). Many conventional farm-based studies have explored the sources of flock infection, modes of transmission, and the host and environmental factors affecting the spread of *Campylobacter* (as previously reviewed; Sahin et al., 2002); however, there is limited information on *Campylobacter* infection in pastured poultry farms. Identification of potential risk factors at the farm level is vital to prevent *Campylobacter* infection in these alternative poultry management systems.

Pastured poultry is a sustainable agriculture technique where chickens have access to fresh pasture land to graze on a daily basis while residing within a moveable pen (Rothrock et al., 2019a). Compared with conventional poultry production facilities in which tens of thousands of birds are housed in relatively small areas,

pastured flocks have much lower stocking densities (>1.5 ft^2/bird) and have greater access to the farm environment, potentially providing a variety of benefits to the birds, environment, and the consumers. Ponte et al. (2008) reported that the omega-3 fatty acid content in the pastured poultry meat is significantly higher than conventional poultry meat. Karsten et al. (2010) reported an increase of vitamin A, vitamin E, and n-3 fatty acid in eggs produced from pastured poultry farms. Health benefits are a key concern for consumers who have shown a rising interest in the pastured poultry products, leading to an increasing demand for pastured poultry products in the U.S. in recent years (Hilimire, 2011). The increased interest in pastured poultry products presents food safety challenges since guidance for pastured poultry growers and scientific research is limited (Elkhoraibi et al., 2017). Karsten et al. (2010) reported that food safety concerns, biosecurity, and a lack of food safety knowledge are the three main challenges faced by pastured poultry growers. Therefore, an increased understanding of food safety issues, and the environmental/management variables that influence food safety issues, is vital to the sustainability of pastured poultry operations.

The environmental factors and farm management practices may have impact on the prevalence of foodborne pathogens in the food products. One method that has been used in the food industry to identify significant environmental or management variables that are correlated to foodborne pathogen prevalence is the use of predictive, machine-learning models. Machine-learning is a set of methods that can automatically detect patterns in data and then use the uncovered patterns to predict outcomes in future data sets (Murphy, 2012). Random forest (**RF**) model is a commonly used machine-learning method that has been used to predict or track pathogen prevalence in several food safety studies (Smith et al., 2010; Pang et al., 2017; Golden et al., 2019a, 2019b). The RF model is an ensemble of classification and regression trees (Breiman, 2001). After fitting with training data, the prediction is made by averaging the outcome between all the trees. An advantage of this algorithm is its ability to handle complex and high-dimensional data (Breiman, 2001). The goal of this study was to identify farm practices variables associated with *Campylobacter* prevalence in pastured poultry farms using RF machine-learning algorithms.

MATERIALS AND METHODS

Sample Collection

A longitudinal study was conducted on 42 flocks of broilers across 11 pastured poultry farms in the southeastern U.S. from March 2014 to November 2017. All 11 farms reared their broiler flocks in movable pens that were moved to fresh pastured daily. The shape, number, and use of temporary fencing around the houses varied among the farms included in the study. A brief description of the size and scale of each farm, as well as other major characterization data, is listed in Table 1. Data were collected for 40 major farm practice variables (Table 2) over a flock's lifecycle and all samples were evaluated for the presence of *Campylobacter*.

The following samples were collected from each flock to analyze the presence of *Campylobacter*: 1) feces, 2) pasture soil, 3) whole carcass rinse directly after processing (**WCR-P**), 4) final product whole carcass rinse after chilling and storage time (**WCR-F**), and 5) ceca samples collected during processing from each farm. If a farm was multiuse and contained pastured layers, swine or cattle (see Table 1), fecal and soil samples were collected from the area where these animals resided at the time of sampling. Fecal and soil samples were taken 3 times throughout a flock's lifecycle: 1) within a few days of being placed on pasture, 2) halfway through their time on the pasture, and (iii) on the day the flock was processed. In all, 2,305 samples consisted of 815 fecal samples, 815 soil samples, 235 WCR-P samples, 230 WCR-F samples, and 210 ceca samples.

On each sampling day, the moveable pens were moved before sampling, and fecal and soil samples were

Table 1. Comparison of the 11 all-natural, antibiotic free, pastured broiler farms included in this study.

Farm	Breed	No. of flocks	Flock size	Multiuse farm?	Animal type(s)	Processing
A	Freedom Ranger	10	>500	Yes	Layers, swine, cattle, sheep	USDA-inspected plant
B	Freedom Ranger, Cornish Cross	5	<50	Yes	Layers, swine, goats	on farm
C	Freedom Ranger	1	<50	No	NA	on farm
D	Freedom Ranger	1	<50	No	NA	on farm
E	Freedom Ranger, Cornish Cross	5	50-100	Yes	Layers, swine, cattle, sheep	on farm
H	Freedom Ranger	2	>500	Yes	Layers	USDA-inspected plant
I	Freedom/Red Ranger, Cornish Cross,	8	100-500	Yes	Layers, swine, goats	USDA-inspected plant
J	Freedom Ranger, Cornish Cross	2	50	Yes	Layers	USDA-inspected plant
K	Freedom Ranger	4	100-500	Yes	Layers, cattle, goats	on farm & USDA-inspected plant
L	Freedom Ranger	2	>500	Yes	Layers, swine, cattle, sheep	USDA-inspected plant
M	Cornish Cross	2	50-100	Yes	Layers, swine	on farm

Table. 2. Predictors used in the fecal, soil, and processing product whole carcass rinse (WCR) random forest model.

Variable	Description	Levels/unit
AvgNumBirds	Average number of birds that the farm handle in 1 year	3 levels: <1000, 1000-10,000, >10000
AvgNumFlocks	Average number of flocks that the farm handle in 1 year	6 levels: 1, 2, 3, 4, 5, 16
YearsFarming	Number of years the farm had been operating at the time of sampling	2 levels: <10, ≥10 (years)
EggSource	Source of broiler eggs	6 levels: company A, B, C, D, E, F
BroodBedding	Type of bedding broilers received during brooding	3 levels: pastured based brooder (PB), wood shavings (WS), sawdust/shredded paper (SDSP)
BroodFeed	Up to top 3 sources of protein in brooding feed	6 levels: barley, wheat, oats (BWO); corn, soy, wheat (CSW); wheat, corn (WC); wheat (W); corn, soy, oats (CSO); peas, corn, oats (PCO)
BrGMOFree	Was the brood feed GMO free?	2 levels: yes (Y), no (N)
BrSoyFree	Was the brood soy free?	2 levels: yes (Y), no (N)
BrMedicated	Was the brood feed medicated?	2 levels: yes (Y), no (N)
BroodCleanFrequency	How often the brooding area was cleaned?	6 levels: 3Days, all in/all out (AIAO), daily, deep litter method (DLM), mobile, weekly, yearly
AveAgeToPasture	Average age broilers were put on pasture	2 levels: 3 weeks, 4 weeks
PastureHousing	Type of pasture housing environment	4 levels: chicken tractor (CT), chicken tractor with fencing (CTF), chicken tractor free ranger (CTFR), chicken tractor with fencing (2 tractors; CTF2)
FreqHousingMove	How often the pasture area was moved?	2 levels: daily, every 2 days
AlwaysNewPasture	Was the pasture always moved to a brand-new pasture area?	2 levels: yes (Y), no (N)
PasturedFeed	Up to top 3 sources of protein in pasture feed	7 levels: barley, wheat, oats (BWO); corn, soy, wheat (CSW); wheat, corn (WC); wheat (W); corn, soy, oats (CSO); corn, cotton seed mill, wheat (CMW); peas, corn, oats (PCO)
PaGMOFree	Was the pasture feed GMO free?	2 levels: yes (Y), no (N)
PaSoyFree	Was the pasture feed soy free?	2 levels: yes (Y), no (N)
PaMedicated	Were broilers medicated while on pasture?	2 levels: yes (Y), no (N)
LayersOnFarm	Were layers present on the farm?	2 levels: yes (Y), no (N)
CattleOnFarm	Were cattle present on the farm?	2 levels: yes (Y), no (N)
SwinOnFarm	Were swine present on the farm?	2 levels: yes (Y), no (N)
GoatsOnFarm	Were goats present on the farm?	2 levels: yes (Y), no (N)
SheepOnFarm	Were sheep present on the farm?	2 levels: yes (Y), no (N)
WaterSource	Water source for broilers during grow-out	3 levels: public, rain, well
FreqBirdHandling	How often chickens were handled on pasture?	2 levels: daily, only if needed (OIN)
AnyABXUse	Were antibiotics ever used on the broilers?	2 levels: yes (Y), no (N)
LengthFeedRestrixProcess	Length of feed restriction before processing	5 levels: 8, 12, 16, 18, 24 (hours)
DayOfYear	Day of the year samples were collected on	Numeric (days)
FlockAgeWeek	Age of flock at time of sampling	Numeric (weeks)
Breed	Breed of broilers used	3 levels: freedom ranger (FR), Cornish cross (CC), red ranger (RR)
Flocksize	Number of birds in the sampled flock	3 levels: 0-100, 100-500, >500
AnimalSource	Type of the animals	4 levels: broiler, layer, swine, cattle
ProcessingType[a]	Where the broilers were processed?	2 levels: farm, plant
SkinOnOff[a]	Skin on or off processing facility	2 levels: on, off
ScalderTempC[ab]	Temperature of water (°C) used during scalding of birds during processing	7 levels: 55, 60, 63, 65, 71, 82, none
RinseWaterSource[ab]	Source of water used for carcass rinsing during process	2 levels: public, well
RinseWaterChlor[ab]	Was the rinse water chlorinated?	2 levels: yes (Y), no (N)
ChillingMethod[a]	Type of chilling method used for carcasses after processing	2 levels: water, air
TransportTime[ab]	Length of time to transport broilers to processors (if necessary)	4 levels: 0.5, 3, 3.5, 5 (hours)
StorageTempC[ab]	Temperature that carcasses were stored before reception by customer	2 levels: −20, 4 (°C)
StorageTimeD[ab]	Amount of time carcasses were stored before reception by customer	Numeric (days)

[a]Variables were used in the WCR-P model.
[ab]Variables were used in the WCR-F model.

collected from the areas where the flock was just moved from. The sampling site was divided into five sections, where five subsamples were collected and pooled from each section. On each sampling site, subsamples were pooled for the high variability expected from each subsample and for the possibility that there would be low numbers of *Campylobacter* (Semenov et al., 2008; Bergholz et al., 2011). Fecal samples were collected by sampling fresh fecal droppings on the sampling site. Soil samples were collected by scooping topsoil (approximately 0-7 cm from the surface) into sterile bags. Sterile scoops were used for each sample, and scoops and gloves were changed after each sample. All pooled samples were at least 25 g. Samples were transferred to a laboratory on ice for processing.

Sample Preparation

Upon arrival in the laboratory, samples were prepared as previously described by Rothrock and Locatelli (2019). Briefly, 3 g from each subsample were combined in a filtered stomacher bag (Seward Laboratory Systems, Inc., Davie, FL) and diluted 1:3 with 10 mM phosphate buffered saline (PBS) and homogenized for 1 min. Next, 100 μL of homogenized sample was plated onto Campy-Cefex agar (Neogen, Lansing MI) and incubated at 42 ± 1°C in microaerophilic conditions (85% N_2, 10% CO_2, 5% O_2) for 36 to 48 h (Stern et al., 1992). Putative *Campylobacter* colonies were enumerated for each plate, and up to 5 suspected colonies were transferred to Brucella agar (Neogen, Lansing MI) supplemented with

10% lysed horse blood (Lampire Biological Laboratories, Pipersville PA) for confirmation and incubated. For model development purposes, samples were classified as positive if countable colonies were found during Campy-Cefex plating.

Model Development

Random forest models were developed for fecal, soil and WCR-P sample data to predict the presence or absence of *Campylobacter* based on the predictors presented in Table 2. As the data for WCR-F and ceca samples were extremely imbalanced, random forest models were not developed. For the WCR-P model, predictors specified for processing samples were included where nonprocessing predictors were used only for fecal and soil models. Before model fitting, each data set was split into training and testing sets. The training and testing sets contained 80% and 20% of the data, respectively. Random forest models were trained with the training set and the test set was served as an independent data set to evaluate the performance of the training model.

Statistical Analysis

All statistical analyses were performed in R (Version 3.4.0; R Foundation for Statistical Computing, Vienna, Austria). The Chi-square test and Fisher's exact test were used to compare the prevalence of *Campylobacter* across different sample types. Results with P value less than 0.05 were considered statistically significant.

The RF model is an ensemble consisting of a collection of classification trees where each tree is independent identically distributed random vectors and these classification trees vote for the most popular class (Breiman, 2001). The classification trees are built based on bootstrap sampling method using a training set which is constructed for drawing observations from the whole data set one at a time and returning them after they have been chosen (Efron, 1979). At each split, the training set is a random subset of all variables, which is a successful approach for assembling unstable leaners (Breiman, 1996; Hastie et al., 2001). With the combination of bagging (bootstrap aggregation) and random variable selection for tree building, each tree in the RF model is unpruned, low-bias tree resulting in low correlation of the individual trees (Diaz-Uriarte and Alvarez de Andres, 2006). However, RF model based on a classification and regression tree (Breiman, 1984) is biased when selecting the important variables, and it favors continuous variables and variables with many categories (Strobl et al., 2007). Strobl et al. (2008) provided 'cforest' function to address the issue, which is based on unbiased conditional inference trees by substituting bootstrap sampling with sampling without replacement (Hothorn et al., 2006; Strobl et al., 2007).

The *party* and *caret* package were used for model training and analysis (Kuhn, 2008; Hothorn et al., 2010). All models were built using "cforest" function with "replace = FALSE" and default option "controls = cforest_unbiased()." To choose the suitable value for m_{try} and n_{tree}, RF models were trained using various m_{try} and n_{tree} values. The values with highest receiver operating characteristic (ROC) statistic were chosen and the chosen values were implemented in the final model. Variable importance was determined using the mean decrease in accuracy. Variables were ranked by relative importance from low to high where the variable with highest value represents the most important variable. Partial dependence plots (PDP) were built for the 2 most important variables in each model using *pdp* package (Greenwell, 2017).

To address the imbalance of negative and positive observations of soil and WCR-P samples, the synthetic minority over-sampling technique (SMOTE) was used (Chawla et al., 2002). The SMOTE method applies a mix of over-sampling minority class and under-sampling majority class to make a balanced training set. After model construction, test set was used to validate the performance of each model. Models were evaluated using area under the ROC curves (AUC; Bradley, 1997), sensitivity and specificity.

RESULTS

Of the 2,305 total samples collected, 910 (39.5%) were *Campylobacter* spp. positive (Table 3). For all the samples collected, the five sample types showed significantly different *Campylobacter* prevalence ($\chi^2 = 728.06$; $df = 4$; $P < 0.0001$). To compare the difference between every 2 groups, Fisher's exact test was performed. *Campylobacter* prevalence was significantly higher in fecal samples (61.1%), compared to the soil (21.1%, $P < 0.0001$), WCR-P (15.6%, $P < 0.0001$), and WCR-F (2.2%, $P < 0.0001$) samples, but the highest *Campylobacter* prevalence was found in the ceca samples (94.3%, $P < 0.0001$). The only pair-wise comparison without a significant difference in *Campylobacter* prevalence was between soil and WCR-P samples ($P = 0.0779$).

Random forest models were constructed for fecal, soil, and WCR-P samples. For the fecal model, the relative importance plot containing the top 10 most important predictor variables is illustrated in Figure 1. The model predicted that source of animal feces was the most important variable in predicting *Campylobacter* prevalence in pastured poultry farms. The mean decrease

Table 3. Effect of sample type on prevalence of *Campylobacter* spp. in pastured poultry samples.

Sample type	No. of samples	No. (%) of positive samples[3]
Fecal	815	498 (61.1) a
Soil	815	172 (21.1) b
WCR-P[1]	235	37 (15.7) b
WCR-F[2]	230	5 (2.2) c
Ceca	210	198 (94.3) d
Total	2305	910 (39.5)

[1]Whole carcass rinse after processing (WCR-P).
[2]Final product whole carcass rinse after chilling and storage (WCR-P).
[3]Different letters represent statistically significant different values when comparing sample types ($P < 0.05$ as determined by the Fisher's exact test).
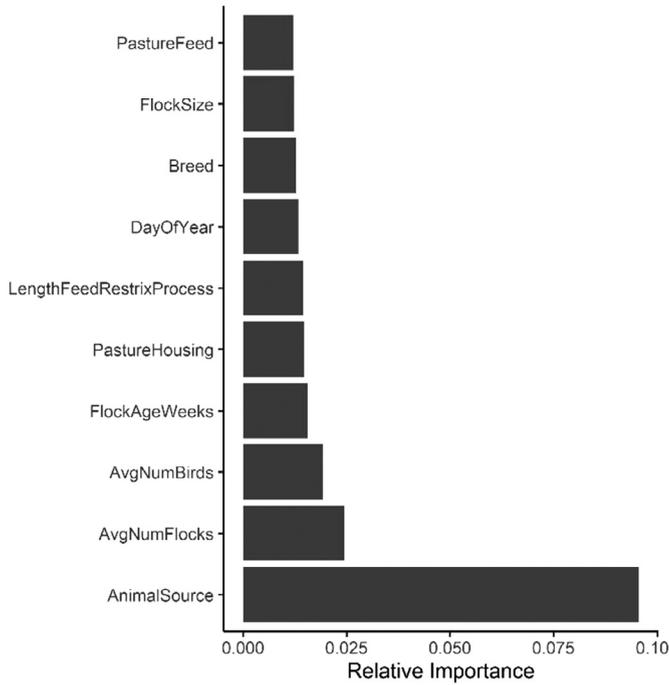
Figure 1. Relative importance plot for fecal models.
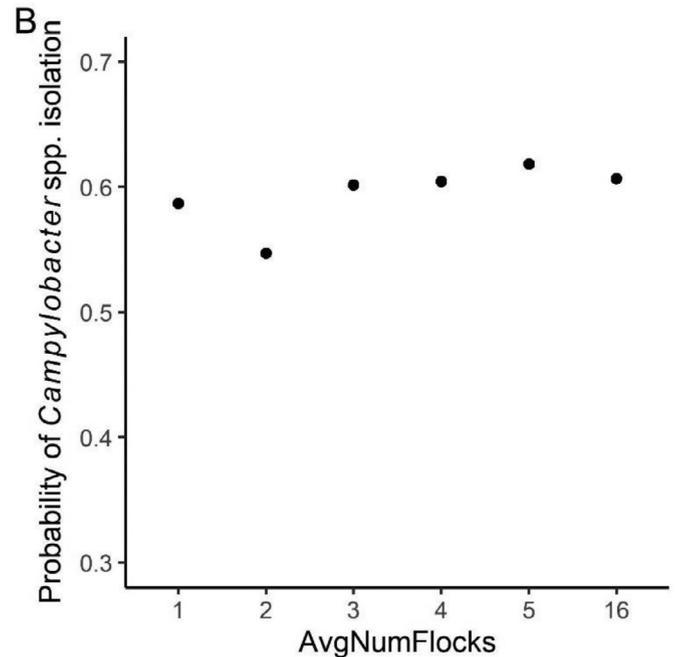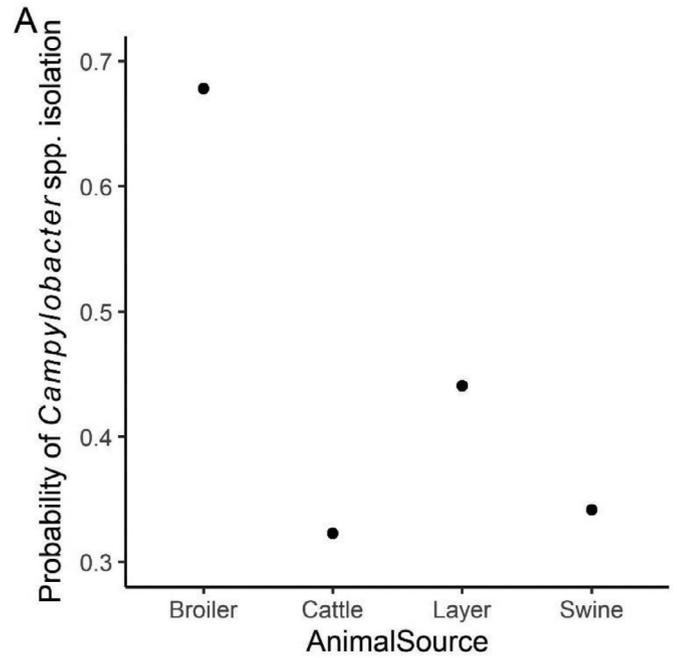
accuracy (MDA) value of animal source was 0.096, compared to 0.025 and 0.019 for average number of flocks and average number of birds a farm handles every year, respectively. No other predictor variables were found to have a MDA value over 0.02. Partial dependency plots (PDPs) were built for animal feces source and average number of flocks (Figure 2). As shown in Figure 2A, fecal samples collected from broiler chickens appeared to have the highest probability of *Campylobacter* isolation. Fecal samples collected from cattle and swine had the lowest and the second lowest probability of *Campylobacter* isolation, respectively. Though there was some variance within Figure 2B, the model suggested an increasing trend of isolating *Campylobacter* as the average number of flocks increased.

For the soil RF model, the relative importance plot is shown in Figure 3. The soy content (soy-containing versus soy-free) of the brood feed was ranked as the most important variable to predict *Campylobacter* prevalence of soil samples in pastured poultry farm with a MDA value of 0.093. The second most important indicator was types of pasture housing used when rearing the chickens with a MDA value of 0.058. The PDP plot of brood feed soy content suggested that brood feed without soybeans had a higher probability of isolating *Campylobacter* in soil samples than feed with soybeans (Figure 4A). The model predicted that *Campylobacter* prevalence was higher when a mobile chicken tractor with fencing was used than other types of pasture housing (Figure 4B).

The relative importance plot for WCR-P model showed that flock age (weeks) was the most important variable in predicting *Campylobacter* prevalence in processing carcasses samples. The MDA value of flock age was 0.053 followed by day of the year and average number of flocks as the second and third important predictors with MDA

Figure 2. Partial dependency plots for the two most important predicting variables in fecal model. Animal source (A) was the most important variable, which represented the types of samples based on animal source, and Average Number of Flocks (B) was the second important variable, which indicated the average number of flocks the farms handle each year.

values of 0.047 and 0.045, respectively (Figure 5). Compared to fecal and soil RF models, the MDA values for WCR-P model were lower. As shown in Figure 6A, the model predicted *Campylobacter* prevalence was higher when flock age was less than 10 weeks and a drop of *Campylobacter* isolation was observed after 10 weeks. Similarly, the probability of *Campylobacter* isolation was higher during the Spring and Summer months (<210 days into the calendar year) and a drop was observed around the late Fall/Winter months.
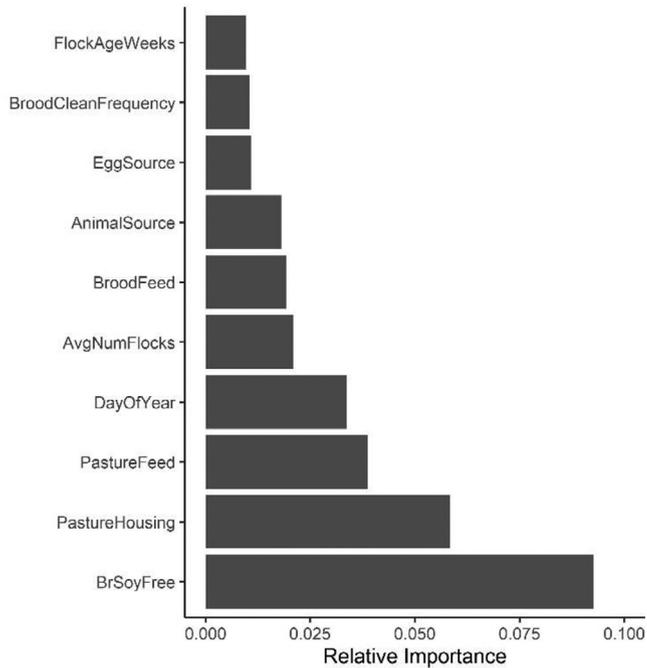
Figure 3. Relative importance plot for soil models.

Figure 4. Partial dependency plots for the two most important predicting variables in soil model. Brood Soy Free (A) was the most important variable, which represented whether the brood feed is soy free or not, and Pasture Housing (B) was the second important variable. Abbreviations: CT, chicken tractor; CTF, chicken tractor with fencing; CTFR, chicken tractor free ranger; CTF2, chicken tractor with fencing (2 tractors).

For all 3 models, the performance was evaluated on a test set separated from the original data set. The test set was not used in training the model but aimed to verify model outcomes. Confusion matrices were generated for the 3 models and were used to demonstrate the ability of models' prediction and observed results (Table 4). Sensitivity is using the correctly predicted positive results divided by the actual positive results. Similarly, specificity represents correctly predicted negatives divided by the true positives. In predicting pathogen prevalence, false negatives cost more than false positives. Thus, sensitivity is of great importance for the models. If the sensitivity of a model is close to one, it means the model has low false negative ratio which indicates good predicting ability of a model. The fecal model showed a sensitivity of 0.8692 and specificity of 0.7705. It suggested that the fecal model correctly predicted 93 positive results out of 107 actual positives whereas 47 negative results were correctly predicted out of 61 true negatives. To further present these results, ROC curves and area under the curve (AUC) were used (Figure 7A-C). An AUC value close to one indicated a good performance of a model. The soil model had a sensitivity of 0.7407 and specificity of 0.8784. Its AUC value was 0.894 (Table 4). For WCR-P model, it obtained a sensitivity of 0.7778 and specificity of 0.8529. The AUC value for WCR-P model is 0.864. Overall, the three models received acceptable sensitivity and specificity values that were above 0.7. In the meantime, the models also achieved AUC values above 0.85.

DISCUSSION

One of the key aspects in reducing *Campylobacter* infection of broiler chickens is to limit transmission pathways. Flies, water source, feces, wild and domestic animal activities on or near the farm, feed, and employee boots have been identified as possible pathways for *Campylobacter* transmission to broilers in conventional farms in the Netherlands, United Kingdom and United States (Humphrey et al., 1993; Pearson et al., 1993; Jacobs-Reitsma et al., 1995; Gregory et al., 1997; Shreeve et al., 2000; Hald et al., 2004). Moreover, bird-to-bird infection within flocks is another potential path of *Campylobacter* transmission. Shreeve et al. (2000) reported that once *Campylobacter* colonization was first
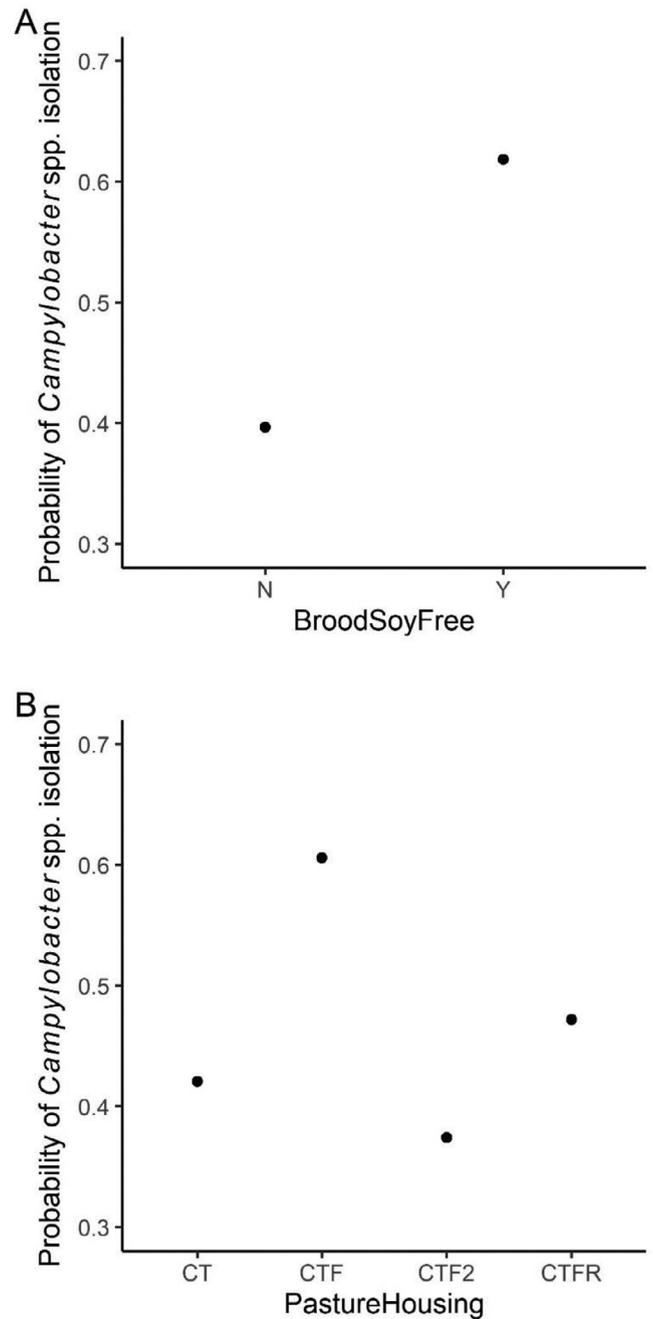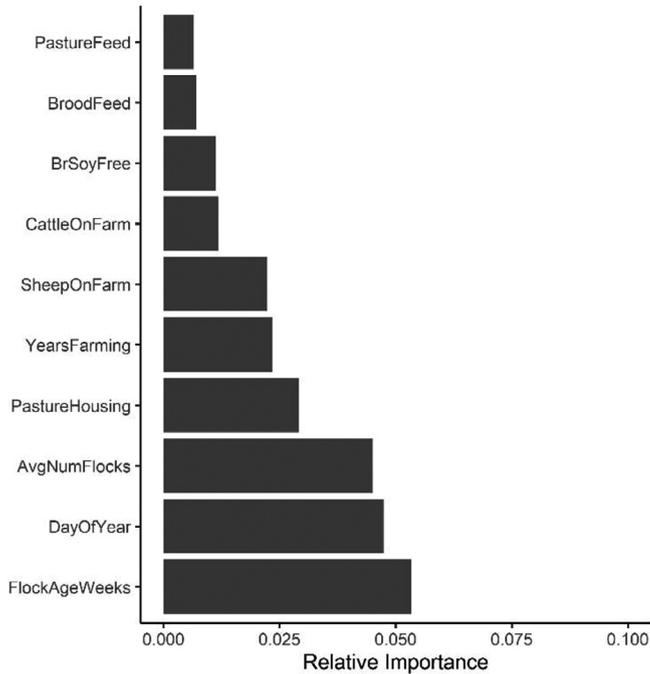
Figure 5. Relative importance plot for WCR-P models.

Figure 6. Partial dependency plots for the two most important predicting variables in WCR-P model. Flock Age Weeks (A) was the most important variable, which represented the age of a flock at the time of sampling and Day of Year (B) was the second important variable, which means the day of year when samples were collected on.

detected in one of the flocks, nearly all flocks tested positive for *Campylobacter* one to eight days later. Similarly, Evans and Sayers (2000) reported that it only took three weeks to spread from 40% of the flocks to over 90%. These findings point out the concerns for alternative systems such as pastured poultry farms that are characterized by the exposure to the natural farm environment. Some studies have analyzed the relationship between environmental samples and carcass rinse samples from broiler chicken flocks on conventional farms (Berghaus et al., 2013; Trimble et al., 2013; Schroeder et al., 2014). More research that relate environmental and carcass rinse samples with pastured poultry farm practices are needed. This study utilized the RF algorithm, a machine learning method used to detect the potential patterns between environmental and carcass rinse samples and pastured poultry farm practices. The results from this study could offer guidelines to pastured poultry growers from both statistical and practical standpoint for their farm operations.

The fecal RF model predicted feces source (broilers, layers, swine, and cattle) as the most important variable in predicting *Campylobacter* prevalence in pastured poultry farms. Broilers have the highest probability for carrying *Campylobacter* whereas swine and cattle samples were least likely to have *Campylobacter* (Figure 2A). Out of the 630 fecal samples taken from broilers, 460 (73%) were *Campylobacter* positive. For 50 swine and 45 cattle fecal samples collected, 6 and 0 positive samples were isolated, respectively. This was not surprising, since Rothrock et al. (2019b) previously showed that *Campylobacter* was part of the core microbiome of the pastured poultry gastrointestinal tract. These results indicate that *Campylobacter* infection of broilers are less likely to be due to other agricultural
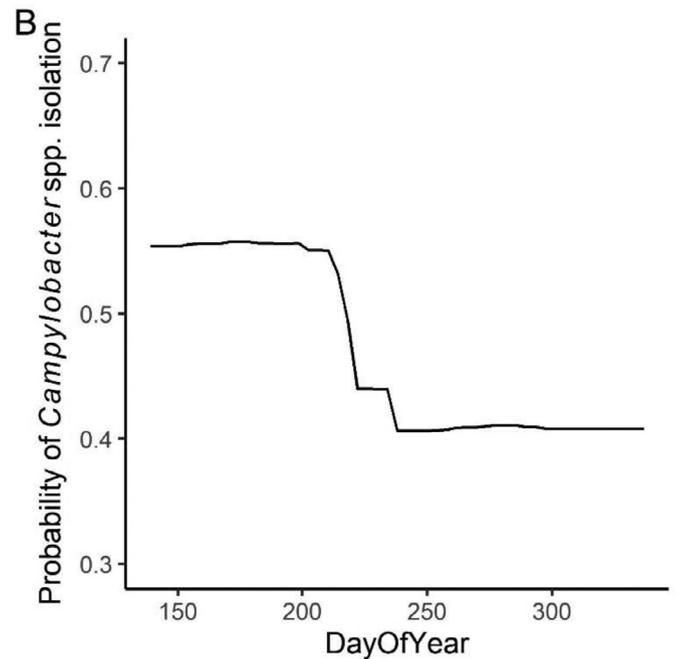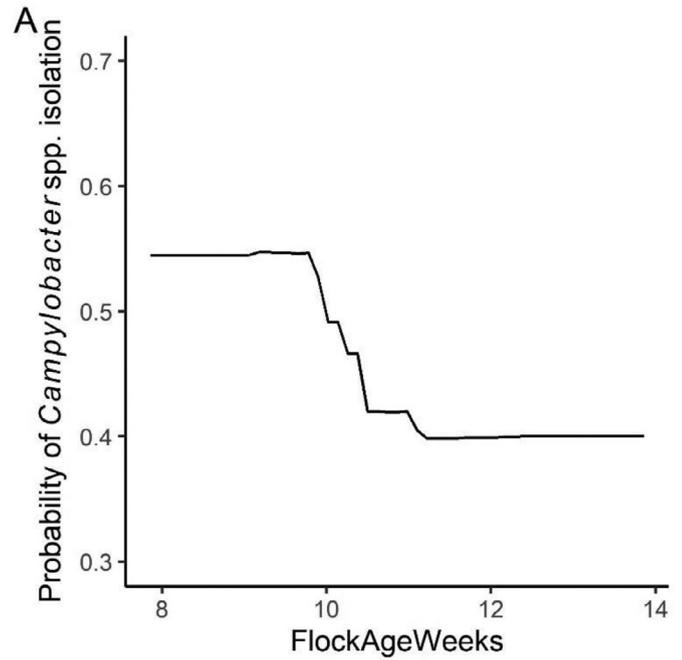
animals on farm. This is in agreement with another study Jacobs-Reitsma et al. (1995), which reported that *Campylobacter* serotypes from layers, swine, sheep, and cattle were different from the ones isolated from broilers. Similarly, another study that investigated ten broiler farms in United Kingdom showed that fecal samples from dogs, sheep, horses, and mammals were *Campylobacter* negative (Bull et al., 2006), indicating a lack of transmission from these animals to broilers. However, Gregory et al. (1997) reported that broiler houses with cattle nearby all tested positive for *Campylobacter*. This
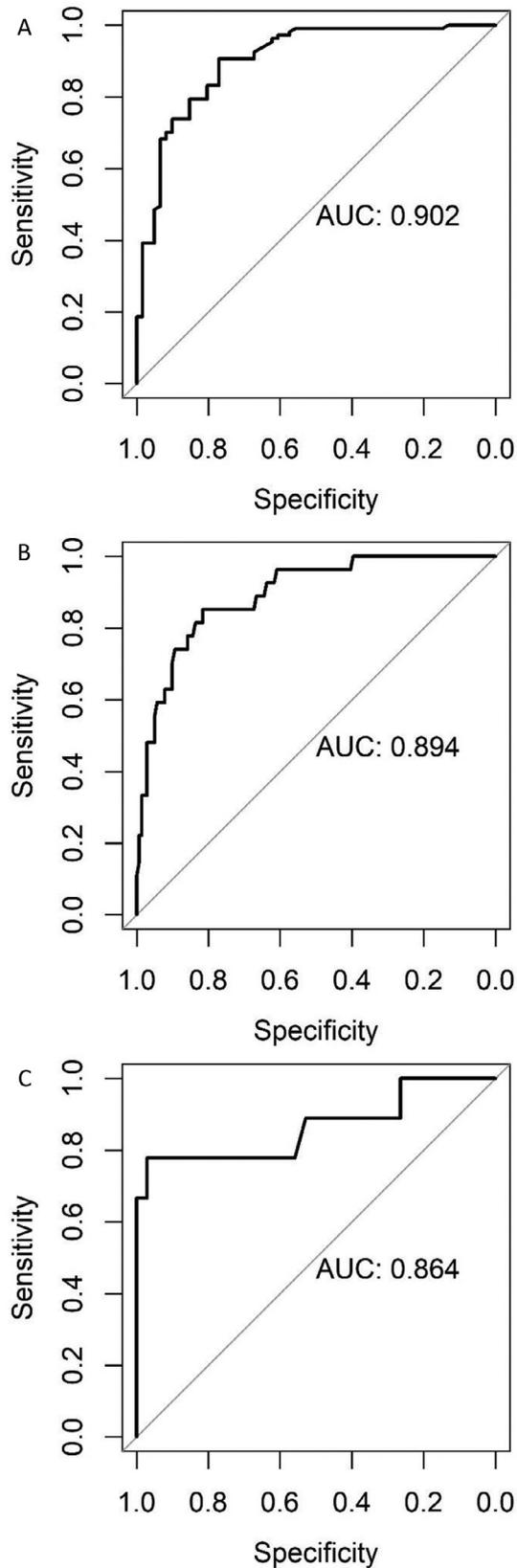
Figure 7. Receiver operating characteristic (ROC) curve for A (Fecal model), B (Soil model), C (WCR-P model).

suggests that cattle are the reservoirs that maintain the organism on the farm. Similar results were presented by Zweifel et al. (2008), who reported that identical *Campylobacter* genotypes were isolated from broilers as well as other farm animals (swine, cattle, and layers),

indicating their role as pathogen reservoirs. The contradiction of these results can be due to different farm practices implemented at each farm. For example, poor control of personnel movement between infected and noninfected areas could lead to cross-contamination between animals. These findings indicate that on a complicated farm environment, a well-designed biosecurity protocol is necessary. The average number of flocks handled on farm was identified as the second highest important predicting variable for the fecal model. In Figure 2B, average number of flocks ranged from 1 to 16, all showed a high probability of *Campylobacter* isolation. This trend is in agreement with another study (Newell and Fearnley, 2003) which reported that once flocks were infected, *Campylobacter* spread to other flocks rapidly usually within a week.

The soy content of the brood feed was identified as the most important variable in detecting *Campylobacter* for the soil RF model. The soy-free brood feed showed higher *Campylobacter* prevalence compared to brood feed with soybeans (Figure 4A). This suggests that soybean, as one of the main sources of plant protein, can lower *Campylobacter* contamination of pasture soils. The removal of soy from pastured poultry diets has been previously shown to reduce *Campylobacter* abundance in fecal and WCR samples, and it was hypothesized that *Campylobacter* were able to metabolize soy components and hence the removal of this nutrient source in soy-free feed led to a decrease in *Campylobacter* (Lourenco et al., 2019). If soy-free feed decreased *Campylobacter* loads in pastured broiler gastrointestinal tracts, then it is possible that *Campylobacter* is passed through in the feces to the soil, resulting in the higher *Campylobacter* prevalence in the pasture soils of broilers fed a soy-free diet during the brood phase. However, soybean-based diets did not show a significant difference compared to other protein source diets in the occurrence of *Campylobacter* in conventionally-reared broiler chickens (Visscher et al., 2017). Other foodborne pathogen data from this same study has shown brood feed composition as a major management variable. Golden et al. (2019b) studied *Listeria* spp. prevalence on these pastured poultry farms using farm management practices as predicting variables and found brood feed composition to be important. The diet consisted of corn, soy, and wheat showed high probability of isolating *Listeria* spp. Similarly, Hwang et al. (2020) studied *Salmonella* prevalence on these pastured poultry farms, and they also found pasture feed and brood feed important variables in predicting *Salmonella* prevalence. These findings, in conjunction with the findings from this current work, highlight the importance of feed composition during the brood phase in predicting foodborne pathogen prevalence throughout the preharvest and postharvest stages of pastured poultry management. Pasture housing was the second most important predictor in detecting *Campylobacter* in the soil model. Four types of pasture housings, chicken tractor, no ranging (CT), single chicken tractor with fencing (CTF), chicken tractor free range (no fence; CTFR), multiple chicken tractors within

Table 4. Predictive performance of random forest models and the confusion matrix of the models.

Models	Predictions	Actual		Sensitivity	Specificity	AUC[a]
		Positive	Negative			
Fecal	Positive	93	14	0.8692	0.7705	0.902
	Negative	14	47			
Soil	Positive	20	17	0.7407	0.8784	0.894
	Negative	7	124			
WCR-P	Positive	7	5	0.7778	0.8529	0.864
	Negative	2	29			

[a]AUC is the area under receiver operating characteristic (ROC) curve.

same fencing (CTF2), were used among the 11 farms investigated. CTF and CTFR presented the highest and second highest probability of isolating *Campylobacter*. The different housing systems might have an impact on isolating *Campylobacter*. However, it was difficult to control the factors that might affect *Campylobacter* prevalence given the observational design used in this study. Experiments that control housing systems would need to be performed to better explain and understand the relationships.

Whole carcass rinse samples collected during processing (postchill, prepackaging, and storage) (WCR-P) were found to have a *Campylobacter* prevalence of 37%, whereas *Campylobacter* prevalence of whole carcass rinse sample final product (WCR-F) was 2.2%. Effective safety practices can prevent *Campylobacter* from spreading during processing. For WCR-P RF model, flock age (weeks) at sampling and day of the year at sampling were predicted as the top 2 most important variables in predicting *Campylobacter* prevalence. As shown in Figure 6A and B, the models indicated that probability of *Campylobacter* isolation was higher in WCR-P samples in broilers processed at an earlier age (8-9 weeks) and then decreased as broiler chickens grow older (10-12 weeks). However, the risk of *Campylobacter* contamination of broiler carcasses increased as the age of broilers increased in a European Union-wide study of 561 slaughterhouses (EFSA, 2010), so these differences may be more related to the breed of chicken (fast-growing Cornish Cross versus slow-growing Freedom Ranger) than the age of the flock.

CONCLUSION

In conclusion, 3 random forest models were generated to predict *Campylobacter* prevalence in fecal, soil, and WCR-P samples collected from 11 southeastern United States pastured poultry farms. Our model identified the type of feces from farm animals as the most important predictors in predicting fecal *Campylobacter* prevalence. Additionally, soy-containing brood feed was associated with higher probability of isolating *Campylobacter* in soil samples. As predicted by the WCR-P model, flock age was the top variable that affected *Campylobacter* prevalence in WCR-P samples. This study showed the use of RF model in predicting bacteria prevalence and the results should assist in identifying factors that are

associated with risks of isolating *Campylobacter*. The generated models will help provide pastured poultry growers and processors with recommendations in implementing farm practices when trying to control or reduce *Campylobacter*.

DISCLOSURES

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Arsi, K., D. J. Donoghue, K. Venkitanarayanan, and A. M Donoghue. 2019. Reducing foodborne pathogens in organic poultry: challenges and opportunities. Pages 25−46 in K. Venkitanarayanan, S. Thakur and S. C. Ricke eds. Food Safety in Poultry Meat Production. Springer International Publishing, Cham.

Berghaus, R. D., S. G. Thayer, B. F. Law, R. M. Mild, C. L. Hofacre, and R. S. Singer. 2013. Enumeration of *Salmonella* and *Campylobacter* spp. in environmental farm samples and processing plant carcass rinses from commercial broiler chicken flocks. Appl. Environ. Microbiol. 79:4106−4114. doi:10.1128/AEM.00836-13

Bergholz, P. W., J. D. Noar, and D. H. Buckley. 2011. Environmental patterns are imposed on the population structure of *Escherichia* coli after fecal deposition. Appl. Environ. Microbiol. 77:211−219.

Bradley, A. P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 30:1145−1159.

Breiman, L. 1984. *Classification and Regression Trees*. Wadsworth International Group.

Breiman, L. 1996. Bagging predictors. Mach. Learn. 24:124−140.

Breiman, L. 2001. Random forests. Mach. Learn. 45:5−32.

Bull, S. A., V. M. Allen, G. Domingue, F. Jorgensen, J. A. Frost, R. Ure, R. Whyte, D. Tinker, J. E. Corry, J. Gillard-King, and

T. J. Humphrey. 2006. Sources of Campylobacter spp. colonizing housed broiler flocks during rearing. Appl. Environ. Microbiol. 72:645–652.

CDC. 2019. The Centers for Disease Control and Prevention. Reports of Selected Campylobacter Outbreak Investigations. Retrieved from https://www.cdc.gov/campylobacter/outbreaks/outbreaks.html.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: synthetic minority oversampling technique. J. Artif. Intell. Res. 16:321–357.

Diaz-Uriarte, R., and S. Alvarez de Andres. 2006. Gene selection and classification of microarray data using random forest. BMC Bioinform. 7:3.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7:1–26.

EFSA. 2010. Analysis of the baseline survey on the prevalence of Campylobacter in broiler batches and of Campylobacter and Salmonella on broiler carcasses, in the EU, 2008. EFSA J. 8. doi:10.2903/j.efsa.2010.1522.

Elkhoraibi, C., M. Pitesky, N. Dailey, and D. Niemeier. 2017. Operational challenges and opportunities in pastured poultry operations in the United States. Poult. Sci. 96:1648–1650.

Evans, S. J., and A. R. Sayers 2000. A longitudinal study of Campylobacter infection of broiler flocks in Great Britain. Prev. Vet. Med. 46:209–223.

Ghareeb, K., W. A. Awad, M. Mohnl, G. Schatzmayr, and J. Böhm. 2019. Control strategies for Campylobacter infection in poultry production. World's Poult. Sci. J. 69:57–76.

Golden, C. E., M. J. Rothrock, and A. Mishra. 2019. Comparison between random forest and gradient boosting machine methods for predicting Listeria spp. prevalence in the environment of pastured poultry farms. Food. Res. Int. 122:47–55.

Golden, C. E., M. J. Rothrock, and A. Mishra. 2019. Using farm practice variables as predictors of Listeria spp. prevalence in pastured poultry farms. Front. Sustain. Food Syst. 3:1–11. doi:10.3389/fsufs.2019.00015.

Greenwell, B. M. 2017. pdp: An R package for constructing partial dependence plots. R J. 9:421–436.

Gregory, E., H. Barnharr, D. W. Dreesen, N. J. Stern, and J. L. Corn. 1997. Epidemiological study of Campylobacter spp. in broilers: source, time of the colonization, and prevalence. Avian Dis. 41:890–898.

Hald, B., H. Skovgard, D. D. Bang, K. Pedersen, J. Dybdahl, J. B. Jespersen, and M. Madsen. 2004. Flies and Campylobacter infection of broiler flocks. Emerg. Infect. Dis. 10:1490.

Hastie, T., R. Tibshirani, and J. H. Friedman. 2001. The Elements of Statistical Learning : Data Mining, Inference, and Prediction : With 200 Full-Color Illustrations. Springer, New York.

Hilimire, K. 2011. The grass is greener: farmers' experiences with pastured poultry. Renew. Agric. Food Syst. 27:173–179.

Hothorn, T., K. Hornik, and A. Zeileis. 2006. Unbiased recursive partitioning: a conditional inference framework. J. Comput. Graph. Statist. 15:651–674.

Hothorn, T., K. Hornik, and A. Zeileis. 2010. Party: A Laboratory for Recursive Partytioning,.

Humphrey, T. J., A. Henley, and D. G. Lanning. 1993. The colonization of broiler chickens with Campylobacter jejuni some epidemiological investigations. Epidemiol. Infect. 110:601–607.

Hwang, D., M. Rothrock, H. Pang, and A. Mishra. 2020. Farm management practices that affect the prevalence of Salmonella in pastured poultry farms. LWT (109423).

Jacobs-Reitsma, W. F., A. W. V. D. Giessen, N. M. Bolder, and R. W. A. W Mulder. 1995. Epidemiology of Campylobacter spp. at two Dutch broiler farms. Epidemiol. Infect. 114:413–421.

Kaakoush, N. O., N. Castano-Rodriguez, H. M. Mitchell, and S. M. Man. 2015. Global ppidemiology of Campylobacter infection. Clin. Microbiol. Rev. 28:687–720.

Karsten, H. D., P. H. Patterson, R. Stout, and G. Crews. 2010. Vitamins A, E and fatty acid composition of the eggs of caged hens and pastured hens. Renew. Agric. Food Syst. 25:45–54.

Kuhn, M. 2008. Building predictive models in R ssing the caret package. J. Stat. Software 28:1–26.

Lourenco, J. M., M. J. Rothrock, Y. M. Sanad, and T. R. Callaway. 2019. The effects of feeding a soybean-based or a soy-free diet on the gut microbiome of pasture-raised chickens throughout their lifecycle. Front. Sustain. Food Syst. 3. doi:10.3389/fsufs.2019.00036.

Murphy, K. P. 2012. Machine Learning a Probabilistic Perspective. MIT press.

Nauta, M., A. Hill, H. Rosenquist, S. Brynestad, A. Fetsch, P. van der Logt, A. Fazil, B. Christensen, E. Katsma, B. Borck, and A. Havelaar. 2009. A comparison of risk assessments on Campylobacter in broiler meat. Int. J. Food Microbiol. 129:107–123.

Newell, D. G., and C. Fearnley. 2003. Sources of Campylobacter colonization in broiler chickens. Appl. Environ. Microbiol. 69:4343–4351.

Pang, H., R. McEgan, A. Mishra, S. A. Micallef, and A. K. Pradhan. 2017. Identifying and modeling meteorological risk factors associated with pre-harvest contamination of Listeria species in a mixed produce and dairy farm. Food Res. Int. 102:355–363.

Pearson, A. D., M. Greenwood, T. D. Healing, D. Rollings, M. Shahamat, J. Donaldson, and R. R. Colwell. 1993. Colonization of broiler chickens by waterborne Campylobacter jejuni. Appl. Environ. Microbiol. 59:987–996.

Ponte, P. I., J. A. Prates, J. P. Crespo, D. G. Crespo, J. L. Mourao, S. P. Alves, R. J. Bessa, M. A. Chaveiro-Soares, L. T. Gama, L. M. Ferreira, and C. M. Fontes. 2008. Restricting the intake of a cereal-based feed in free-range-pastured poultry: effects on performance and meat quality. Poult. Sci. 87:2032–2042.

Rothrock, M. J., K. E. Gibson, A. C. Micciche, and S. C. Ricke. 2019. Pastured poultry production in the United States: strategies to balance system sustainability and environmental impact. Front. Sustain. Food Syst. 3. doi:10.3389/fsufs.2019.00074.

Rothrock, M. J., and A. Locatelli. 2019. Importance of farm environment to shape poultry-related microbiomes throughout the farm-to-fork continuum of pasture-raised broiler flocks. Front. Sustain. Food Syst. 3. doi:10.3389/fsufs.2019.00048.

Rothrock, M. J., A. Locatelli, K. M. Feye, A. J. Caudill, J. Guard, K. Hiett, and S. C. Ricke. 2019. A microbiomic analysis of a pasture-raised broiler flock elucidates foodborne pathogen ecology along the farm-to-fork continuum. Front. Vet. Sci. 6:260.

Sahin, O., P. Kobalka, and Q. Zhang. 2003. Detection and survival of Campylobacter in chicken eggs. J. Appl. Microbiol. 95:1070–1079.

Sahin, O., T. Y. Morishita, and Q. Zhang. 2002. Campylobacter colonization in poultry: sources of infection and modes of transmission. Anim. Health Res. Rev. 3:95–105.

Schroeder, M. W., J. D. Eifert, M. A. Ponder, and D. G. Schmale. 2014. Association of Campylobacter spp. levels between chicken grow-out environmental samples and processed carcasses. Poult. Sci. 93:734–741.

Semenov, A. V., E. Franz, L. van Overbeek, A. J. Termorshuizen, and A. H. van Bruggen. 2008. Estimating the stability of Escherichia coli O157:H7 survival in manure-amended soils with different management histories. Environ. Microbiol. 10:1450–1459.

Shreeve, J. E., M. Toszeghy, M. Pattison, and D. G. Newell. 2000. Sequential spread of Campylobacter infection in a multipen broiler house. Avian. Dis. 44:983–988.

Smith, A., B. Sterba-Boatwright, and J. Mott. 2010. Novel application of a statistical technique, Random Forests, in a bacterial source tracking study. Water Res. 44:4067–4076.

Stern, N. J., B. Wojton, and K. Kwiatek. 1992. A Differential-selective medium and dry ice-generated atmosphere for recovery of Campylobacter jejuni. J. Food Prot. 55:514–517.

Strobl, C., A. L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. Conditional variable importance for random forests. BMC Bioinform. 9:307.

Strobl, C., A. L. Boulesteix, A. Zeileis, and T. Hothorn. 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinform. 8:25.

Trimble, L. M., W. Q. Alali, K. E. Gibson, S. C. Ricke, P. Crandall, D. Jaroni, M. Berrang, and M. Y. Habteselassie. 2013. Prevalence and concentration of Salmonella and Campylobacter in the processing environment of small-scale pastured broiler farms. Poult. Sci. 92:3060–3066.

Visscher, C. F., A. Abd El-Wahab, M. F. E. Ahmed, J. Hankel, V. Taube, and J Kamphues. 2017. Influence of different protein sources in the broiler diet on the presence of Campylobacter spp. in excreta and caecal content. J. Anim. Physiol. Anim. Nutr. (Berl) 101(Suppl 1):95–104.

Zweifel, C., k. D. Scheu, M. Keel, and R Stephan. 2008. Occurrence and genotypes of Campylobacter in broiler flocks, other farm animals, and the environment during several rearing periods on selected poultry farms. Int. J. Food Microbiol. 125:182–187.