IDENTIFYING RISK FACTORS ASSOCIATED WITH *SALMONELLA* PREVALENCE IN SOUTHEASTERN UNITED STATES PASTURED POULTRY FARMS

by

DAIZY HWANG

(Under the Direction of Abhinav Mishra)

ABSTRACT

Consumer demand has increased for pastured poultry products in the recent years. It is necessary to identify the meteorological factors and farm management and processing practices associated with the prevalence of *Salmonella* on pastured poultry farms. Presence of *Salmonella* in the environment could lead to contamination of the final product. The objective of this study was to develop predictive models that identify the specific meteorological, farm management and processing factors that contribute to the presence of *Salmonella*, samples including soil, feces, and whole carcass rinses. Random forest method was used to develop the models, and receiver operating characteristic (ROC) curves were used to evaluate the performances of these models. All models generated in this study had predicting abilities with the area under the ROC curve values above 0.87. The predictive models developed in this study can provide users practical and effective tool to make informed decisions based on scientific evidence.

INDEX WORDS: Pastured poultry farm, *Salmonella*, random forest

IDENTIFYING RISK FACTORS ASSOCIATED WITH *SALMONELLA* PREVALENCE IN SOUTHEASTERN UNITED STATES PASTURED POULTRY FARMS

by

DAIZY HWANG

B.S.F.C.S., The University of Georgia, 2016

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2019

IDENTIFYING RISK FACTORS ASSOCIATED WITH *SALMONELLA* PREVALENCE IN SOUTHEASTERN UNITED STATES PASTURED POULTRY FARMS

by

DAIZY HWANG

Major Professor:	Abhinav Mishra
Committee:	Mark Harrison
	Michael Rothrock Jr.

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2019

DEDICATION

I want to thank my mom in supporting my decisions in life. I also want to thank Eric, for your constant encouragement and love.

ACKNOWLEDGEMENTS

I would like to thank my major professor Dr. Abhinav Mishra for guidance and support throughout my research process. I truly have learned a lot through him. I would like to thank Dr. Mark Harrison for taking me on to the graduate school path and serving on my committee, as well as Dr. Michael Rothrock Jr. for serving on my committee. I would like to thank Dr. Hao Pang for statistical guidance and encouragement. Finally, I want to thank friends Gwen Hirsch and Merlyn Thomas for being my support system, help when needed, and all my wonderful office peeps Brittany Magdovitz, Chase Golden, Sam Cimowsky and Lizzy White for the fun times.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES viii

LIST OF FIGURES ix

CHAPTER

1 INTRODUCTION 1

References 4

2 LITERATURE REVIEW 5

Salmonellosis and Recent Outbreaks 5

Sources of *Salmonella* in the Environment 6

Meteorological Factors on *Salmonella* and Other Pathogens 7

Pastured Poultry Farm 9

Random Forest Analysis 11

References 14

3 PREDICTING *SALMONELLA* PREVALENCE ASSOCIATED WITH METEOROLOGICAL FACTORS IN PASTURED POULTRY FARM IN SOUTHEASTERN UNITED STATES 24

Abstract 25

Introduction 25

Materials and Methods 27

Results....31
Discussion....33
Conclusion....36
References....37
4 FARM PRACTICE VARIABLES THAT EFFECT THE PREVALENCE OF *SALMONELLA* IN PASTURED POULTRY FARMS....50
Abstract....51
Introduction....51
Materials and Methods....53
Results....56
Discussion....58
References....62
5 CONCLUSIONS....87
APPENDICES....88
Appendix A....88
Appendix B....91

LIST OF TABLES

Page

Table 2.1: Summary of poultry related salmonellosis hospitalization and deaths in the United States 2015-2018 22

Table 3.1: Comparison of the 11 all-natural, antibiotic-free, pastured broiler farms included in this study 43

Table 3.2: Effect of season on prevalence of *Salmonella* in pastured poultry farms 44

Table 3.3: Predictive performance of the soil and feces random forest models 45

Table 4.1: The farm management variables included in the random forest model 72

Table 4.2: Storage temperature and average storage time in days for each farm 76

Table 4.3: Predictive performance of random forest models and the confusion matrices of the models 77

Table 4.4: Seasonal effect on *Salmonella* presence in pastured poultry farms 78

LIST OF FIGURES

Page

Figure 3.1: Top 10 Relative importance variables for feces and soil sample 46

Figure 3.2: Partial dependency plots of soil model for top 1 and 2 predicting variables ..47

Figure 3.3: Partial dependency plots of feces model for top 1 and 2 predicting variables 48

Figure 3.4: ROC curves for soil model (a) and feces model (b).....49

Figure 4.1: Relative importance plots for feces, soil, WCR-p and WCR-f samples 79

Figure 4.2: Partial dependency plots for top 2 feces model.....81

Figure 4.3: Partial dependency plots for top 2 soil model 82

Figure 4.4: Partial dependency plots for top 2 WCR-p model 83

Figure 4.5: Partial dependency plots for top 2 WCR-f model 84

Figure 4.6: Receiver operating characteristic curve for random forest model. (a) feces (b) soil (c) WCR-p, (d) WCR-f models 85

CHAPTER 1

INTRODUCTION

Salmonella associated with poultry and poultry products is often identified as the source of foodborne illness. *Salmonella* is responsible for causing salmonellosis from ingesting contaminated foods. Common sources of *Salmonella* include raw poultry meats, shell eggs, and also other processed poultry products. *Salmonella* is also known to survive different environmental conditions, including low water activity or relative humidity. Routes of entry of *Salmonella* include various places from the farm environments to food processing facilities.

There has been an increase in consumer demand for pastured poultry products in the United States due to awareness and concern for welfare of the birds (Sossidou, Dal Bosco, Elson, & Fontes, 2011). Consumers pay premium prices for pastured poultry products, and their willingness to pay has driven the niche market into a bigger scale in the recent decade (Hilimire, 2012). Pastured poultry are exposed to environmental influences due to continuous access to the outdoors. The birds freely forage for fresh grass and insects that can supplement their diet. It is believed that the exposure to the outside environment can provide the birds healthy lifestyle compared to the conventional system; however, pastured poultry farming has higher mortality rates due to greater exposure to predators and disease (Sossidou, Dal Bosco, Elson, & Fontes, 2011). Major concerns in the pastured poultry systems include the uncontrollable weather events as well as need for protection from other wildlife around the pasture. Special attention

toward the microbial safety in this type of farming system is necessary, since foodborne pathogens like *Salmonella* are present in the environment.

It is recognized that weather conditions and farming practices can influence the distribution and prevalence of foodborne pathogens in the poultry farms (Ivenek et al., 2009; Park et al., 2014). Meteorological factors such as humidity, precipitation, wind speed, and temperature can affect the *Salmonella* presence in pastured poultry farms. For example, heavy rain events can create hazardous environments for the birds (Sossidou, Dal Bosco, Elson, & Fontes, 2011). Farm management practices such as types of feed used, breed of birds, and cleaning frequencies have been previously found to impact the prevalence of *Salmonella* in poultry flocks (Siemon, Bahnson, & Gebreyes, 2007). Given many factors that can affect pathogen prevalence, predictive models, such as decision trees can be used to identify the risk factors.

The objectives of this research were 1) to identify the risk factors, specifically meteorological variables and farm management practices, that are associated with the prevalence of *Salmonella*, and 2) to develop models that can predict the presence of *Salmonella* in the pastured poultry farms.

This thesis includes five chapters. The first chapter introduces the research topics of the thesis and the research objectives. The second chapter provides a literature review of related topics including information about salmonellosis and recent outbreaks, pastured poultry farms, sources of *Salmonella* in the environment, meteorological factors on *Salmonella* and other pathogens, and the random forest analysis. The third chapter identifies the meteorological factors that affect the *Salmonella* presence in the pastured poultry farms. The fourth chapter investigates the farm management and processing

practices that contribute to the prevalence of *Salmonella* in the farm and processing samples. The predictive models were developed and validated for estimating their performances. The models identified the important explanatory variables as well as the relationship that the variables have on the presence of *Salmonella* in the farm environment. The final (fifth) chapter includes the conclusion of the research done and suggested future research avenues.

References

Hilimire, K. (2012). The grass is greener: Farmers' experiences with pastured poultry. *Renewable Agriculture and Food Systems, 27*(3), 173-179.

Ivanek, R., Gröhn, Y., Wells, M., Lembo, A., Sauders, B., & Wiedmann, M. (2009). Modeling of spatially referenced environmental and meteorological factors influencing the probability of *Listeria* species isolation from natural environments. *Applied and Environmental Microbiology, 75*(18), 5893-5909.

Park, S., Navratil, S., Gregory, A., Bauer, A., Srinath, I., Szonyi, B., Nightingale, K., Anciso, J., Jun, M., Han, D. (2014). Farm management, environment and weather factors jointly affect the probability of spinach contamination with generic *Escherichia coli* at the preharvest level. *Applied and Environmental Microbiology*, 03643-03613.

Siemon, C. E., Bahnson, P. B., & Gebreyes, W. A. (2007). Comparative investigation of prevalence and antimicrobial resistance of *Salmonella* between pasture and conventionally reared poultry. *Avian Diseases, 51*(1), 112-117.

Sossidou, E., Dal Bosco, A., Elson, H., & Fontes, C. (2011). Pasture-based systems for poultry production: implications and perspectives. *World's Poultry Science Journal, 67*(1), 47-58.

CHAPTER 2

LITERATURE REVIEW

Salmonellosis and Recent Foodborne Outbreaks

Salmonella is a Gram-negative, non-spore forming bacterium that can cause foodborne illness. *Salmonella* is responsible for causing approximately 1.1 million cases of illness in the United States annually (CDC, 2019). There are two types of salmonellosis, non-typhoidal gastroenteritis and typhoidal fever. Depending on the serotype of *Salmonella*, the disease can cause different symptoms and severity (Jones et al., 2008). Salmonellosis generally has a low mortality rate (1%); however, in immunocompromised groups, the mortality rates can be higher. Symptoms of salmonellosis include nausea, vomiting, cramps, diarrhea, and fever that occur at 12 to 72 hours after consumption of contaminated food (CDC, 2019).

Salmonella infects the gut in the host. It survives the acidic environment of the stomach (Garcia-del Portillo, Foster, & Finlay, 1993), colonizing the intestines then producing enterotoxins resulting in the host (Bäumler, Tsolis, & Heffron, 1996). *Salmonella* has been recognized a serious food safety issue in food products such as eggs, raw poultry and meat. In the recent decade (2009-2019), *Salmonella* from identifiable food sources were responsible for over 1,500 outbreaks, 40,000 illnesses, and 69 deaths in the United States according to the National Outbreak Reporting System (CDC, 2018a). The salmonellosis outbreaks indicate that the infectious dose is low, but there is high variability depending on type of food consumed, and the age and health of an individual

(Finstad, O'Bryan, Marcy, Crandall, & Ricke, 2012). A foodborne outbreak is when there are two or more individuals sharing the same illness from consuming the same contaminated food. Recent poultry-related outbreaks from 2015 to 2018 are summarized in Table 2.1. In 2018, there were six *Salmonella* outbreaks related to poultry in the United States, all of them being multistate outbreaks. Of those outbreaks, two were caused by shell eggs, resulting in 23 hospitalizations. Raw poultry products caused three different outbreaks with 143 hospitalizations and 3 deaths (CDC, 2018b). The most common serotype that was associated with outbreaks related with chicken was *S.* Enteritidis. *S.* Enteritidis accounted for 20% of the isolations, whereas *S.* Typhimurium accounted for 17% of the isolations. However, the most common serotype that was found in poultry farms was *S.* Kentucky, which is not a major pathogen in humans in comparison with other *Salmonella* serotypes (Finstad et al., 2012).

Sources of *Salmonella* in the Environment

The environmental sources for the pathogen entry into the poultry management chain include the henhouse, poultry, and humans (Guard-Petter, 2001). In a processing plant, some of the major steps where *Salmonella* contamination may occur are evisceration and head pulling, intestine could rupture and splash feces onto the carcass (Smith et al., 2007). There are measures in the poultry industry to control the growth of pathogens like *Salmonella* on the carcass by chilling the carcasses to below 4°C. Commercial chilling methods include air chilling and immersion in chilled water baths containing sanitizers such as chlorine or peroxyacetic acid (James, Vincent, de Andrade Lima, & James, 2006).

Soil provides a medium for bacterial survival and the survival depends on numerous factors associated with the soil such as temperature, moisture, nutrient availability and the initial microbial load (Jacobsen & Bech, 2012). Soil type can influence *Salmonella* survival, with higher survival rates in clay-type soil than loamy soil, due to the high moisture holding capacity of the clay (Holley, Arrus, Ominski, Tenuta, & Blank, 2006). In moist soil, *Salmonella* was able to survive 45 days on inoculated soil (Guo, Chen, Brackett, & Beuchat, 2002). The nutrient rich soil provides a niche where *Salmonella* can survive and even possibly replicate (Baloda, Christensen, & Trajcevska, 2001; Winfield & Groisman, 2003).

In the farm environment, the exposure to feces, manure, and compost can affect the prevalence of *Salmonella.* Manure-amended soil reduced *Salmonella* Typhimurium due to competition of the soil microbial community that was present (Garcia, Baelum, Fredslund, Santorum, & Jacobsen, 2010). *Salmonella* has also been isolated in different water sources such as rivers, surface waters and irrigation water (Cherry et al., 1972; Norman & Kabler, 1953). In the case of surface water, *Salmonella* can contaminate produce by its use in irrigation water, as well as, during flooding after an extreme weather event such as storms (Cooley et al., 2007; Liu, Hofstra, & Franz, 2013). Wildlife such as insects, rodents and birds may carry pathogens such as *Salmonella* and act as a vector to infect the livestock (Cooley et al., 2007; Liu et al., 2013).

Meteorological Factors on *Salmonella* and Other Foodborne Pathogens at the Farm Level

Precipitation can influence the prevalence of foodborne pathogens and has been identified as the highest contributing factor for the presence of generic *E. coli* in spinach contamination. Increase in the amount of rain increased the prevalence of foodborne

pathogens (Park et al., 2014). Heavy rainfall results in the diffusion of *Salmonella* into the environment from farms, possibly due to runoffs from other contamination sources that were present such as manure and compost (Gorski et al., 2011). Also, the increased rainfall can lead to higher moisture in the soil, which in turn can increase the survival of microorganisms (Dowe, Jackson, Mori, & Bell, 1997). Surface water can serve as a vehicle for transmission of bacteria, especially if there are livestock, manure applications and wildlife in the farm area (Strawn et al., 2013).

Ambient temperature increases the prevalence of cases of salmonellosis (D'Souza, Becker, Hall, & Moodie, 2004; Kovats et al., 2004). There are some seasonal trends of *Salmonella* prevalence in the farm settings, where the prevalence was higher in warmer months (Baptista, Alban, Ersbøll, & Nielsen, 2009; Pangloli et al., 2008). In free-range layer flocks, seasonality and temperature were significantly associated with *Salmonella* contamination, where summer months had higher prevalence of this pathogen (Wales, Breslin, Carter, Sayers, & Davies, 2007).

Increased wind speeds have been found to affect the prevalence of foodborne pathogens in the farm environment (Pang, McEgan, Mishra, Micallef, & Pradhan, 2017) identified that the increasing wind speed just before sampling increased the risk of finding *Listeria* spp. in mixed farm, a farm that has both produce and farm animals. Dust and other particles within the farm environment were also found to be able to aerosolize and spread to different areas of the farm (Oni, Sharma, & Buchanan, 2015), providing a possible mechanism for foodborne pathogen distribution throughout the farm environment.

Pastured Poultry Farming System

There is an increased interest from the consumers for alternative methods of poultry production. Although it is a fast-growing industry, compared to the conventional methods of poultry farming, the pastured poultry market accounts for less than 1 percent of the total poultry market in U.S. (Trimble et al., 2013). Pastured poultry generally is reared outdoors with movable, floorless pens. This method is also distinguished from organically grown poultry. Pastured poultry live on and forage directly in the pasture, whereas in organic farms, birds only have access to the environment, but primarily feed indoors (Fanatico, Pillai, Emmert, & Owens, 2007; Hilimire, 2012). Terms such as organic, pasture-raised, antibiotic free and free-ranged poultry products may bring confusion to consumers. Pasture-raised chickens spend most of their lives on a pasture with continuous exposure to grass and sunlight; however, this system exposes the birds to various environmental conditions. Weather conditions may affect the microbial safety of the birds. For instance, extreme weather events can lead to conditions that can potentially create hazardous environments for birds (Sossidou, Dal Bosco, Elson, & Fontes, 2011). These conditions can also influence the spread of diseases through runoffs during rain events.

U.S. Department of Agriculture (USDA)-labeled organic broilers are grown where the birds are raised without antibiotics and fed organic feeds and supplements. The birds also have access to the outside (Dimitri & Greene, 2002). The free-range label necessitates proof that the poultry has had access to the outdoors. Natural poultry products have “no artificial ingredient or added color and only is minimally processed” (USDA, 2015).

There are guidelines for European pastured poultry growing systems, but they are not defined in the U.S. by USDA (Fanatico & Born, 2002; Sossidou et al., 2011). The public assumption is that by birds having access to the outside increases their welfare. This would lead to more ethical, better tasting, healthier and safer ways to grow chickens (Brennan, Gallagher, & McEachern, 2003; Hughner, McDonagh, Prothero, Shultz, & Stanton, 2007; Yiridoe, Bonti-Ankomah, & Martin, 2005). However, there is a greater need for more scientific evidence for these claims. *Salmonella* prevalence has been found to be higher in free-range and all-natural chicken than from broilers grown in an enclosed area with the conventional methods (Bailey & Cosby, 2005; McCrea et al., 2006). Exposure to outside water sources, insects and other wildlife could potentially introduce *Salmonella* and other pathogens to chickens.

In previous studies, *Salmonella* prevalences were compared between pastured poultry, free-range and organic farms. There were no significant differences between the farms (Bailey & Cosby, 2005, Siemon, Bahnson, & Gebreyes, 2007). There are limited amounts of scientific studies investigating the microbial safety in pastured poultry farms. In conventional systems, antibiotics are given for therapeutic purposes and the prevention of diseases. Antibiotics also can alter the gut microbiota and result in increased animal growth (Pan & Yu, 2014). In pastured poultry systems, antibiotics are usually not used. The poultry are moved frequently to prevent pecking of fecal material, and sick birds are separated from the flock more often to prevent contamination (Siemon et al., 2007). However, one study reported that the prevalence of *Campylobacter* and *Salmonella* were found to be greater in free-range broilers compared to those reared in conventional, enclosed houses (McCrea et al., 2006).

Random Forest Analysis

Random forest models and other machine learning techniques can be applied to numerous food safety situations. These models can provide useful information in goals to reduce risks from foodborne pathogens, like *Salmonella* prevalence, in pastured poultry farms.

In classification and regression tree (CART) models, there are usually multiple explanatory variables and one dependent variable. CART is a machine learning method where prediction models are generated from the existing data. Classification trees refer to a type of data where the dependent variable is a "class" type of variable. Regression trees are where the dependent variable is a continuous variable.

The random forest model also is a type of classification and regression tree that uses bootstrapping (Breiman, 2001). The model is created as an ensemble, where multiple trees are generated, hence the term "forest" is used. Bootstrapping is a random sampling method in random forest algorithm that protects the model in overfitting to the training set, and this is because it repeatedly selects samples with replacements. In order to improve the model, a tuning parameter (m_{try}) is chosen to find the best split from those predictors. Another method to increase the prediction accuracy is growing the number of trees to improve the model. More stable results of variable importance are obtained with more trees, usually more than 500 trees (Liaw & Wiener, 2002). The variable importance plot ranks the predicting variables based on mean decrease in classification accuracy using the Gini index (Breiman et al., 1984). By creating the variable importance, a ranking system of the importance variables can be generated.

Validation is an important step to measure if the model can predict accurately with an independent data set; however, it is often difficult to have access to an independent data set. This challenge can be overcome by partitioning the data into training and testing sets. The training set is used to build and train the model. The testing set is left out until the end to test the predictive performance of the trained model. The training sets are also tuned, which improves the model performance. The test set serves as the sufficient validation to evaluate the model performance if the model can be applied to an independent data set. In random forest models, the tuning parameter is m_{try}. The m_{try} refers to the number of variables for splitting at each node. For classification, the default value is then calculated as the square root of the number of predicting variables.

In the case of presence or absence studies, where the dependent variable is binary, class imbalance can happen. Imbalance happens when one class of binary data occurs rarely. When developing models, imbalance can lead to low sensitivity. Sensitivity, or the true positive rate, is important in prediction performance where the rate of positives is rare. Random over sampling method (ROSE) can overcome this problem (Lunardon, Menardi, & Torelli, 2014). The ROSE package (in R software) can generate synthetic balanced samples to create more minority case to improve the low sensitivity. In food safety, correctly predicting the positives is vital because positive pathogen presence such as *Salmonella* can lead to illness.

Partial dependency plots (PDPs) are visualizations of the marginal effect that the predictor variable has on the dependent variable (Friedman, 2001). These plots focus on one predicting variable having the effect of average prediction of the dependent variable, therefore increasing the interpretability of the model. The PDPs can visualize if the

relationship between a single predictor variable and the response variable is linear, monotonous, or complex. Monotonous PDPs means that the predicting variable have no effect on the response variable. Predicted probability values, in percent, are often used as the y-axis. This can quantify how much the predicted probability increases or decreases. Partial dependence models can also show a relationship between two variables, creating a 3-dimensional plot (Zhao & Hastie, 2017). The plots created are used for the interpretation of the model predictions.

Receiver operating characteristic (ROC) curves are used to measure predictive performance of a model. The plot is built on the sensitivity and 1- specificity values, which are the rates of true positive detection and the rate of false positive, respectively. Area under the curve (AUC) values are calculated, ranging from 0.5 to 1, where a value of 1 is the maximum theoretical value given that all values are predicted correctly. The AUC values can also be used to compare model performances (Hanley & McNeil, 1983; Harrell Jr, 2015).

References

Alali, W. Q., Thakur, S., Berghaus, R. D., Martin, M. P., & Gebreyes, W. A. (2010). Prevalence and distribution of *Salmonella* in organic and conventional broiler poultry farms. *Foodborne Pathogens and Disease, 7*(11), 1363-1371.

Bailey, J., & Cosby, D. (2005). *Salmonella* prevalence in free-range and certified organic chickens. *Journal of Food Protection, 68*(11), 2451-2453.

Baloda, S. B., Christensen, L., & Trajcevska, S. (2001). Persistence of a *Salmonella enterica* serovar Typhimurium DT12 clone in a piggery and in agricultural soil amended with *Salmonella*-contaminated slurry. *Applied and Environmental Microbiology, 67*(6), 2859-2862.

Baptista, F., Alban, L., Ersbøll, A. K., & Nielsen, L. R. (2009). Factors affecting persistence of high *Salmonella* serology in Danish pig herds. *Preventive Veterinary Medicine, 92*(4), 301-308.

Bäumler, A. J., Tsolis, R. M., & Heffron, F. (1996). Contribution of fimbrial operons to attachment to and invasion of epithelial cell lines by *Salmonella* Typhimurium. *Infection and Immunity, 64*(5), 1862-1865.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Breiman, L., Friedman, J., & Olshen, R. (1984). Stone CJ. Classification and regression trees. Wadsworth Wadsworth statistics/probability series: Belmont, CA: CRC Press.

Brennan, C., Gallagher, K., & McEachern, M. (2003). A review of the 'consumer interest'in organic meat. *International Journal of Consumer Studies, 27*(5), 381-394.

CDC. (2018a). National Outbreak Reporting System (NORS). Retrieved March 16, 2019, from https://wwwn.cdc.gov/norsdashboard/

CDC. (2018b). Reports of selected *Salmonella* outbreak investigations. Retrieved Febuary 10, 2019, from https://www.cdc.gov/salmonella/outbreaks.html

CDC. (2019). *Salmonella* Questions and answers. Retrieved April 10, 2019, from https://www.cdc.gov/salmonella/general/index.html

Cherry, W. B., Hanks, J. B., Thomason, B. M., Murlin, A. M., Biddle, J. W., & Croom, J. M. (1972). S*almonellae* as an index of pollution of surface waters. *Applied and Environmental Microbiology, 24*(3), 334-340.

Cooley, M., Carychao, D., Crawford-Miksza, L., Jay, M. T., Myers, C., Rose, C., Keys, C., Farrar, J., Mandrell, R. E. (2007). Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California. *PloS one, 2*(11), e1159.

D'Souza, R. M., Becker, N. G., Hall, G., & Moodie, K. B. (2004). Does ambient temperature affect foodborne disease? *Epidemiology, 15*(1), 86-92.

Dimitri, C., & Greene, C. R. (2002). *Recent growth patterns in the US organic foods market*. (No. 1474-2016- 12-853).

Dowe, M. J., Jackson, E. D., Mori, J. G., & Bell, C. R. (1997). *Listeria monocytogenes* survival in soil and incidence in agricultural soils. *Journal of Food Protection, 60*(10), 1201-1207.

Fanatico, A., & Born, H. (2002). Label Rouge: Pasture-based poultry production in France. *An ATTRA Livestock Technical Note (*https://attra.ncat.org/attrapub/download.php?id=224*)*.

Fanatico, A., Pillai, P. B., Emmert, J., & Owens, C. (2007). Meat quality of slow-and fast-growing chicken genotypes fed low-nutrient or standard diets and raised indoors or with outdoor access. *Poultry Science, 86*(10), 2245-2255.

Finstad, S., O'Bryan, C. A., Marcy, J. A., Crandall, P. G., & Ricke, S. C. (2012). *Salmonella* and broiler processing in the United States: relationship to foodborne salmonellosis. *Food Research International, 45*(2), 789-794.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.

Garcia, R., Baelum, J., Fredslund, L., Santorum, P., & Jacobsen, C. S. (2010). Influence of temperature and predation on survival of *Salmonella enterica* serovar Typhimurium and expression of invA in soil and manure-amended soil. *Appliedand Environmental Microbiology, 76*(15), 5025-5031.

Garcia-del Portillo, F., Foster, J. W., & Finlay, B. (1993). Role of acid tolerance response genes in *Salmonella* typhimurium virulence. *Infection and Immunity, 61*(10), 4489-4492.

Gorski, L., Parker, C. T., Liang, A., Cooley, M. B., Jay-Russell, M. T., Gordus, A. G., Atwill, E. R., Mandrell, R. E. (2011). Prevalence, distribution, and diversity of *Salmonella enterica* in a major produce region of California. *Applied and Environmental Microbiology, 77*(8), 2734-2748.

Guard-Petter, J. (2001). The chicken, the egg and *Salmonella* enteritidis. *Environmental Microbiology, 3*(7), 421-430.

Guo, X., Chen, J., Brackett, R. E., & Beuchat, L. R. (2002). Survival of *Salmonella* on tomatoes stored at high relative humidity, in soil, and on tomatoes in contact with soil. *Journal of Food Protection, 65*(2), 274-279.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148*(3), 839-843.

Harrell Jr, F. E. (2015). Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: *Springer*.

Hilimire, K. (2012). The grass is greener: Farmers' experiences with pastured poultry. *Renewable Agriculture and Food Systems, 27*(3), 173-179.

Holley, R. A., Arrus, K. M., Ominski, K. H., Tenuta, M., & Blank, G. (2006). *Salmonella* survival in manure-treated soils during simulated seasonal temperature exposure. *Journal of Environmental Quality, 35*(4), 1170-1180.

Hughner, R. S., McDonagh, P., Prothero, A., Shultz, C. J., & Stanton, J. (2007). Who are organic food consumers? A compilation and review of why people purchase organic food. *Journal of Consumer Behaviour: An International Research Review, 6*(2-3), 94-110.

Ivanek, R., Gröhn, Y., Wells, M., Lembo, A., Sauders, B., & Wiedmann, M. (2009). Modeling of spatially referenced environmental and meteorological factors influencing the probability of *Listeria* species isolation from natural environments. *Applied and Environmental Microbiology, 75*(18), 5893-5909.

Jacobsen, C. S., & Bech, T. B. (2012). Soil survival of *Salmonella* and transfer to freshwater and fresh produce. *Food Research International, 45*(2), 557-566.

James, C., Vincent, C., de Andrade Lima, T., & James, S. (2006). The primary chilling of poultry carcasses—a review. *International Journal of Refrigeration, 29*(6), 847-862.

Jones, T. F., Ingram, L. A., Cieslak, P. R., Vugia, D. J., Tobin-D'Angelo, M., Hurd, S., Medus, C., Cronquist, A., Angulo, F. J. (2008). Salmonellosis outcomes differ substantially by serotype. *The Journal of Infectious Diseases, 198*(1), 109-114.

Kovats, R. S., Edwards, S. J., Hajat, S., Armstrong, B. G., Ebi, K. L., & Menne, B. (2004). The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten European countries. *Epidemiology and Infection, 132*(3), 443-453.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1-26.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18-22.

Liu, C., Hofstra, N., & Franz, E. (2013). Impacts of climate change on the microbial safety of pre-harvest leafy green vegetables as indicated by *Escherichia coli* O157 and *Salmonella* spp. *International Journal of Food Microbiology, 163*(2-3), 119-128.

Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1*(1), 14-23.

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: A Package for Binary Imbalanced Learning. *R Journal, 6*(1).

McCrea, B., Tonooka, K., VanWorth, C., Boggs, C., Atwill, E., & Schrader, J. (2006). Prevalence of *Campylobacter* and *Salmonella* species on farm, after transport, and at processing in specialty market poultry. *Poultry Science, 85*(1), 136-143.

Norman, N. N., & Kabler, P. W. (1953). Bacteriological study of irrigated vegetables. *Sewage and Industrial Wastes*, 605-609.

Oni, R. A., Sharma, M., & Buchanan, R. L. (2015). Survival of *Salmonella enterica* in dried turkey manure and persistence on spinach leaves. *Journal of Food Protection, 78*(10), 1791-1799.

Pan, D., & Yu, Z. (2014). Intestinal microbiome of poultry and its interaction with host and diet. *Gut Microbes, 5*(1), 108-119.

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.

Pangloli, P., Dje, Y., Ahmed, O., Doane, C., Oliver, S., & Draughon, F. (2008). Seasonal incidence and molecular characterization of *Salmonella* from dairy cows, calves, and farm environment. *Foodborne Pathogens and Disease, 5*(1), 87-96.

Park, S., Navratil, S., Gregory, A., Bauer, A., Srinath, I., Szonyi, B., Nightingale, K., Anciso, J., Jun, M., Han, D. (2014). Farm management, environment and weather factors jointly affect the probability of spinach contamination with generic *Escherichia coli* at the preharvest level. *Applied and Environmental Microbiology*, 03643-03613.

Siemon, C. E., Bahnson, P. B., & Gebreyes, W. A. (2007). Comparative investigation of prevalence and antimicrobial resistance of *Salmonella* between pasture and conventionally reared poultry. *Avian Diseases, 51*(1), 112-117.

Smith, D., Northcutt, J., Cason, J., Hinton Jr, A., Buhr, R., & Ingram, K. (2007). Effect of external or internal fecal contamination on numbers of bacteria on prechilled broiler carcasses. *Poultry Science, 86*(6), 1241-1244.

Sossidou, E., Dal Bosco, A., Elson, H., & Fontes, C. (2011). Pasture-based systems for poultry production: implications and perspectives. *World's Poultry Science Journal, 67*(1), 47-58.

Strawn, L. K., Fortes, E. D., Bihn, E. A., Nightingale, K. K., Gröhn, Y. T., Worobo, R. W., Wiedmann, M., 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.

Trimble, L. M., Alali, W. Q., Gibson, K. E., Ricke, S. C., Crandall, P., Jaroni, D., Berrang, M., Habteselassie, M. Y. (2013). Prevalence and concentration of *Salmonella* and *Campylobacter* in the processing environment of small-scale pastured broiler farms. *Poultry Science, 92*(11), 3060-3066.

USDA. (2015). Meat and Poultry Labeling Terms. Retrieved March 4, 2019, from https://www.fsis.usda.gov/wps/portal/fsis/topics/food-safety-education/get-answers/food-safety-fact-sheets/food-labeling/meat-and-poultry-labeling-terms/meat-and-poultry-labeling-terms.

Winfield, M. D., & Groisman, E. A. (2003). Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Applied and Environmental Microbiology, 69*(7), 3687-3694.

Yiridoe, E. K., Bonti-Ankomah, S., & Martin, R. C. (2005). Comparison of consumer perceptions and preference toward organic versus conventionally produced foods: A review and update of the literature. *Renewable Agriculture and Food Systems, 20*(4), 193-205.

Zhao, Q., & Hastie, T. (2017). Causal interpretations of black-box models. *Journal of Business & Economic Statistics,* Press.

Table 2.1. Summary of poultry related salmonellosis hospitalizations and deaths in the United States during 2015-2018 (CDC, 2018b)

Food source	Year	Serotype	Hospitalization	Deaths
Raw chicken products	2018	*Salmonella* Infantis	32	1
Gravel Ridge farms shell eggs	2018	*Salmonella* Enteritidis	12	0
Chicken	2018	*Salmonella* I 4,[5]],12:i:-	11	1
Raw turkey products	2018	*Salmonella* serotype not specified	107	1
Shell eggs, Rose acre	2018	*Salmonella* Braenderup	11	0
Chicken salad	2018	*Salmonella* Typhimurium	94	1
Live poultry	2017	*Salmonella* serotype not specified	249	1
Shell eggs	2016	*Salmonella* Oranienburg	2	0
Live poultry	2016	*Salmonella* serotype not specified	48	3
Raw, frozen, stuffed chicken entrees by Aspen foods	2015	*Salmonella* Enteritidis	1	2
Raw, frozen, stuffed chicken entrees by Barber foods	2015	*Salmonella* Enteritidis	4	0
Live poultry	2015		63	0

Salmonella serotype
not specified

CHAPTER 3

PREDICTING *SALMONELLA* PREVALENCE ASSOCIATED WITH METEOROLOGICAL FACTORS IN PASTURED POULTRY FARMS IN SOUTHEASTERN UNITED STATES[1]

[1] Hwang, D., Rothrock, M. Jr., Pang, H., Mishra, A. To be submitted to *Food Control*

Abstract

Consumer demand has increased for pastured poultry products as the drive for sustainable farming practices and ethical treatments of livestock have become popular in the popular press. It is necessary to identify the important meteorological factors associated with the prevalence of *Salmonella* in the pastured poultry settings since the presence of *Salmonella* in the environment could lead to contamination of the final product. The objective of this study was to develop a model to describe the relationship between meteorological factors and the presence of *Salmonella* on the pastured poultry farms. The random forest method was used to develop a model where 83 meteorological factors were included as the predicting variables. The soil model identified humidity as the most important variable associated with *Salmonella* prevalence, while high wind gust speed and average temperature were identified as important meteorological variables in the feces model. The developed models were robust in predicting the prevalence of *Salmonella* in pastured poultry farms with the area under ROC curve values of 0.872 and 0.884 for feces and soil models, respectively. The predictive models developed in this study can provide users with practical and effective tools to make informed decisions with scientific evidence regarding the meteorological parameters that are important to monitor for increased on-farm *Salmonella* prevalence.

Introduction

Salmonella is a foodborne pathogen that is frequently associated with poultry meat and products. The Centers for Disease Control and Prevention (CDC) reports that non-typhoidal *Salmonella* causes 1.1 million foodborne illnesses and 450 deaths every year in the United States (CDC, 2018). *Salmonella* is non-spore forming, gram negative

bacteria that can survive in many environmental settings outside of an animal host (Franz & van Bruggen, 2008). Salmonellosis, infection with *Salmonella* spp., has been historically linked to consumption of contaminated poultry and eggs, but is not limited to these products, with contamination occurring in produce and dry spices in recent years (Fatica & Schneider, 2011; Liu, Hofstra, & Franz, 2013; Van Doren, Kleinmeier, Hammack, & Westerman, 2013). Finding measures to control the presence of *Salmonella* at the farm level to ultimately ensure safety of consumption is desired. It is important to raise awareness of the environmental exposure in the pastured poultry management system.

Pastured poultry meat, eggs and other products have increased in popularity over recent years. Increased production efforts toward organic, local, and sustainable practices in poultry products have been commercialized due to consumer demand (Oberholtzer, Greene, & Lopez, 2006; Sossidou, Dal Bosco, Elson, & Fontes, 2011). It is important to analyze the risk associated with these types of farming practices since pathogens such as *Salmonella* can be present in these environments, and even from biofilms (Kumar, Willians, Srianganathan, Boyerm & Eifert, 2018; Micallef & Kumar, 2017). According to the American Pastured Poultry Producers Association (APPPA), pastured poultry farming practices involve floorless pens that provides protection. The pens are on the rotation system where they are moved daily, and the chickens have continuous access to fresh pasture (APPA, 2019). To our knowledge, limited research has been conducted in the microbial safety of pastured poultry farms (Baron, 2016; Siemon, 2007). Even fewer have used predictive modeling, (Golden, Rothrock, & Mishra, 2019) and none involving modeling *Salmonella* prevalence in pastured poultry farms.

Predictive models such as random forest models can be applied to numerous food safety situations. These models can provide useful information in goals to reduce risks from foodborne pathogens, like *Salmonella*, in pastured poultry farms. Random forest is an ensemble method consisting of numerous classification and regression trees (Breiman, 2001). By averaging the predictions across all individual trees, random forest predictions are robust and not prone to overfitting. This study used predictive models in random forest method for prevalence of *Salmonella* in pastured poultry farms. The objectives of this study were to: 1) develop a model to predict the presence of *Salmonella* in pastured poultry farms in the Southeastern U.S. and 2) identify the specific meteorological factors contributing to the presence of *Salmonella* in pastured poultry farms.

Material and Methods

Description of the Study

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 with temporary fences. A brief description of the size and scale of each farm is provided in Table 3.1.

Sample Collection

The following samples were collected for each flock: (i) feces, and (ii) pasture soil. All samples were collected in the field and returned to the lab in a cooler packed with ice. At each sampling time, the pasture area was divided into 5 separate sections and 5 subsamples in each section were pooled into a single sample for each section (5 total fecal and 5 total soil samples were collected on each sampling day). The total amount of sample collected for each field sample was at least 25 g. To prepare the

environmental samples for homogenization, 3 g (feces, soil) was combined within filtered stomacher bags and diluted 1:3 using 1X phosphate-buffered saline (PBS). All samples were homogenized for 60 sec and these homogenates were used for all downstream cultural isolations.

Cultural Isolation Methods

As a pre-enrichment step, the stomached homogenates remained in the filtered stomacher bags and incubated overnight at 35 °C. Two different enrichment broths were used to isolate *Salmonella* spp. from these environmental samples: tetrathionate (TT; Becton-Dickinson, Sparks, MD) broth and Rappaport-Vassiliadis (RV; Becton Dickinson) media. After overnight incubation at 42 °C in both of these enrichment broths, 1 loopful from each enrichment broth was spread on two different differential media: brilliant green sulfa with novobiocin (BGS; Becton Dickinson) agar and xylose lysine tergitol-4 (XLT-4; Becton Dickinson) agar. These plates were incubated overnight at 35 °C, and on each plate, 3 *Salmonella*–like colonies per subsample were picked and confirmed using triple sugar iron agar (TSI; Becton-Dickinson) and lysine iron agar fermentation (LIA; Becton-Dickinson) using an incubation period of 18-24 hours at 35 °C. Final confirmation of suspect TSI/LIA isolates was performed using *Salmonella* polyvalent O antiserum agglutination (Becton-Dickinson), using manufacturer's specifications. The sample collection and cultural isolation methods were adopted from Rothrock, Hiett, Guard, & Jackson (2016).

Meteorological Data Collection

The meteorological data for each farm was collected using Weather Underground (www.wunderground.com) and National Oceanic and Atmospheric Administration

(NOAA) databases from the closest weather stations (www.noaa.gov). For each respective sampling date, weather data was collected which included the temperature (minimum, maximum and average °C), humidity (minimum, maximum and average %), precipitation (cm), wind speed (minimum, maximum and average m/s), and maximum gust speed (m/s). The weather data was collected retrospectively up to 7 days prior to the sampling date, averaging the prior dates to account for the lasting effect of the meteorological events. A total of 83 meteorological variables were included in these models. These variables are found on the supplementary table in the appendix B.

Model Development

The statistical analysis was performed using R (version 3.4.0; R Foundation for Statistical Computing, Vienna, Austria). Random forest method was used to assess the impact of meteorological factors (explanatory variables) and to predict the presence of *Salmonella* (response variable) in the soil and feces samples using these meteorological factors. In this study, random forest models were developed using the "randomForest" R-package (Liaw & Wiener, 2002). To build and train the model, 70% of the study data were randomly selected by the algorithms built on the random forest package, then the remaining 30% of the data was later used for testing the developed random forest models.

Models for soil and feces samples were developed separately. The models were developed using the best tuning parameter m_{try} to improve the model, where the m_{try} value is the optimum number of splits from a node. The variable importance plot was then generated for ranking the most important explanatory variables. Since our study had a smaller number of positive samples, we used the oversampling method with the R-package (Random Over-Sampling Examples) (ROSE) (Lunardon, Menardi, & Torelli,

2014). The oversampling from the positive samples was necessary due to low sensitivity of our model and the ROSE method addresses this limitation.

Random forest models can be difficult to interpret, as it is referred to as a black box method. Partial dependency plot (PDPs) and the variable importance plot were used to provide better interpretation of the model results. The relative importance plot ranked the predicting variables based on the mean decrease in classification accuracy (Breiman, Friedman, & Olshen, 1984). The top variable had the relative importance score of 100, which indicates that the variable chosen had the most influence in *Salmonella* presence out of the other factors included in this study.

The impact of the predicting variables has in the presence of *Salmonella* in soil and feces samples was evaluated using the relative importance plots and the scores calculated. To further understand the relationship between a single predicting variable and the presence of *Salmonella*, partial dependency plots (PDPs) were generated. The PDPs were only generated for the top two important variables identified from the relative importance plots.

Model Performance Assessment

The test set of remaining 30% of the data (those not used in the development of the models) were used to test the model performances. The predictive performances of the developed models were evaluated using values of sensitivity and specificity, and the area under ROC curves. Receiver operating characteristic (ROC) curves were generated by plotting 1- specificity by the sensitivity. The area under ROC curve values were then calculated.

Results

There were total 693 soil samples with 56 of them being *Salmonella* positive (8.1%). In the feces sample, there were total 779 samples and 125 of them were *Salmonella* positive (16.0%). The prevalence data was divided into four seasons (Table 3.2). For the soil samples, prevalence of *Salmonella* was maximum in summer (11%), whereas for feces samples, the maximum positive samples were in the spring season (22%). For both sample types, there were no positive samples for the winter season (five samples collected for each type of sample).

Each random forest model developed in this study incorporated all 83 meteorological variables. Variable importance plots provided a ranking system for estimating the most important predicting variables (Fig. 3.1). The relative importance was calculated and shown where the most important variable has an importance score of 100. The five most important factors for predicting the prevalence of *Salmonella* in soil samples were associated with humidity, where the most important variable was average humidity seven days prior to the sampling day. The top five most important factors for predicting *Salmonella* prevalence in feces samples included wind gust speed, temperature, humidity and seasons. All of the meteorological variables were averaged values from the previous days to capture the lasting effects of meteorological events.

In the soil model, average humidity of seven days prior sampling date was the most important variable, followed by the average humidity six days prior to the sampling date. The importance scores were 100 and 62, respectively. Both variables had sustained high predicted probability when the average humidity was between 20 to 70%. Decrease

in predicted probability was observed when the average humidity values were 70 to 90%. The prevalence also increased as the average humidity values were 90 to 100% (Fig. 3.2).

The most important variable identified from the feces model was maximum gust speed six days prior to the sampling date, and the second most important variable was average temperature seven days prior to the sampling date. The relative importance scores were 100 and 93, respectively (Fig. 3.1). Partial dependency plots were then generated for the top two most important variables to depict the dependency of the individual predicting variables for prevalence of *Salmonella* (Fig. 3.3).

The predicted probability of *Salmonella* prevalence was low when the maximum wind gust speed was less than 10 m/s, and the model-estimated probability was higher when the gust speed was more than 11 m/s. On the average temperature seven days prior to sampling, the predicted probability of the *Salmonella* prevalence increased when the temperature exceeded higher than 28° C.

The predictive performances and the confusion matrices of the developed models are presented in Table 3.3. The testing set left out in the beginning was used to test the prediction with models trained with the training data. The sensitivity, or the true positive detection rate, and specificity, or the true negative detection rate, of the feces model were 0.811 and 0.894, respectively. For the soil model, sensitivity was 0.875 and the specificity was 0.844. These values were used in evaluating the predictive performance. The receiver operating characteristic curves (ROC curves) were generated, where the true positives and false negatives are plotted (Fig. 3.4). The area under the ROC curve values were calculated, which were 0.872 for feces model and 0.884 for the soil model.

Discussion

Managing foodborne pathogens incorporating holistic approach from farm to fork has been effective risk management strategy to reduce contamination. Presence of *Salmonella* in pastured poultry farms may result in contamination of the final product. Therefore, reducing the initial risk of *Salmonella* from the farm is recommended to prevent foodborne illness. Models developed in this research identified important meteorological factors that contribute to the prevalence of *Salmonella* within the preharvest environmental (soil and fecal) samples. The information gathered from the models can be used to educate users such as farmers and poultry producers in farming practices. Although meteorological factors cannot be controlled, some practice measures can be implemented to reduce the prevalence of *Salmonella* such as wind screens.

The low prevalence of *Salmonella* in feces and soil samples indicates that the data was imbalanced in terms of having more negative cases than positives. The models were improved by using the Random Over-Sampling Examples (ROSE) technique that generates synthetic random samples of the positives to balance the underrepresented data (Lunardon et al., 2014). ROSE was appropriate to use since the data was binary classification for presence or absence of *Salmonella*. This resampling method enhanced the model performance by balancing the number of positive and negative cases to increase the sensitivity, which represented the ability of the model to detect the true positive outcomes of *Salmonella*. Even though the resampling did not greatly affect the AUC values, it improved the sensitivity of the model. High true positive detection rate was essential to building a robust model when the positive cases are rare, since identifying *Salmonella* presence was the goal of these models.

Classification trees have been used to predict the effects of meteorological factors on finding foodborne pathogens in produce and mixed farm environments (Ivanek et al., 2009; Pang, McEgan, Mishra, Micallef, & Pradhan, 2017; Strawn et al., 2013). Since pastured poultry reside freely outdoors, the birds are heavily affected by meteorological factors. It was expected that there would be similar exposures to microbial contamination in the pastured poultry farms. AUC values from ROC curves assessed the models' predictive abilities (Fig. 3.4). In this study, both soil and feces models created had AUC values above 0.87, indicating that random forest models in this study have effectively identified the important variables that contribute to *Salmonella* prevalence.

The management of the birds in the free-range system can be complex since there are uncontrollable factors such as extreme weather events and wildlife exposure. For birds raised outdoors, the spread of disease was higher when the weather was warm, as well as when the ground is wet from precipitation (Sossidou et al., 2011, Jacobsen & Bech, 2012). A previous study observed that seasonal patterns were more apparent among free-range flocks then the caged flocks (Wales et al., 2007). Exposure to outside water sources, insects and other wildlife can easily introduce *Salmonella* and other pathogens to chickens. *Salmonella* prevalence was higher in free-range flocks than the prevalence from those that were grown enclosed area, the conventional method (Bailey & Cosby, 2005; McCrea et al., 2006).

Although, in this study, precipitation was not one of top 10 predicting variables, rainfall affected the prevalence of foodborne pathogens in mixed and produce farms in previous studies (Pang et al., 2017; Park et al., 2014). Heavy precipitation that causes flooding and runoff may disperse *Salmonella* to new areas. These meteorological events

can therefore introduce *Salmonella* to new areas such as water sources and soil. Results from this study indicate that meteorological factors affect the prevalence of *Salmonella* in pastured poultry farms. *Salmonella* can potentially survive in preharvest broiler farm environments such as in soil and poultry feces (Greene et al., 2008; Holley, Arrus, Ominski, Tenuta, & Blank, 2006).

In the soil model, average humidity seven and six days prior to the sampling date were the top two predicting variables (Fig. 3.1). The PDPs show high predicted probability for detecting *Salmonella* until the percent humidity reached 70 to 90% in average humidity 7 days prior, and 70 to 80% in average humidity 6 days prior to sampling day (Fig. 3.2). The atmospheric humidity did not correlate to the soil moisture (data not shown); however, there was an increase in predicted probability after average humidity exceeded 90%. *Salmonella* is known to survive environments outside the host. These environments include lower water activity environments (Winfield & Groisman, 2003). Previous research has reported longer survival of *Salmonella* in low moisture soil (Chandler & Craven, 1980).

In the feces model, the maximum gust speed 6 days prior to sampling day was the most important predicting variable. Wind gust by definition is a sudden increase in windspeed, when the speed is at least over 8.2 meters per second (NOAA, 2019). The predicted probability of *Salmonella* increased when the maximum gust speed seven days prior to the sampling day was more than 10 m/s (Fig. 3.2). This may be due to aerosolized *Salmonella* along with other dust particles rapidly moving through the air with the high wind speed (Dungan, 2010; Oliveira, Carvalho, & Garcia, 2006; Oni, Sharma, & Buchanan, 2015; Kumar, Williams, Al Qublan, Srirangnathan, Boyer, &

Eifert, 2017). Average temperature 7 days prior to sample day was identified as the second most important predicting variable in the feces model. Temperatures exceeding 28 °C had increased predicted probability of *Salmonella* in feces samples. This supports the behavior of *Salmonella* which survives and grows better at warmer temperatures greater than 25 °C (D'Souza, Becker, Hall, & Moodie, 2004; ICMF, 1996, Liu et al., 2013).

Conclusion

In the context of food safety, biological significance was focused on assessing the usefulness of the model along with the prediction performance. The random forest method has been used in solving many food safety problems to make casual inferences. Careful handling in processing is recommended due to higher prevalence of *Salmonella* during warmer months. Having a routine cleaning schedule may reduce the *Salmonella* presence by preventing the transfer of fecal matter from strong winds throughout the farm. Knowledge of the important factors will provide farmers, food producers, and risk managers with practical information to make informed decisions with scientific evidence.

References

APPA. (2019). American Poultry Producers Association. Retrieved April 15, 2019, from https://www.apppa.org/

Baron, P. (2016). Awareness and perceptions of food safety risks and risk management in poultry production and slaughter: A qualitative study of direct-market poultry producers in Maryland. *PloS One., 11*(6), E0158412.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Breiman, L., Friedman, J., & Olshen, R. (1984). Stone CJ. Classification and regression trees. Wadsworth Wadsworth statistics/probability series: Belmont, CA: CRC Press.

CDC. (2018). Reports of Selected *Salmonella* outbreak investigations. Retrieved November, 2018, from https://www.cdc.gov/salmonella/outbreaks.html.

Chandler, D. S., and J. A. Craven. (1980) Relationship of soil moisture to survival of *Escherichia coli* and *Salmonella* Typhimurium in soils. *Australian Journal of Agricultural Research* 31(3) 547-555.

D'Souza, R. M., Becker, N. G., Hall, G., & Moodie, K. B. (2004). Does ambient temperature affect foodborne disease? *Epidemiology, 15*(1), 86-92.

Dungan, R. (2010). Board-invited review: Fate and transport of bioaerosols associated with livestock operations and manures. *Journal of Animal Science, 88*(11), 3693-3706.

Fatica, M. K., & Schneider, K. R. (2011). *Salmonella* and produce: survival in the plant environment and implications in food safety. *Virulence, 2*(6), 573-579.

Franz, E., & van Bruggen, A. H. (2008). Ecology of *E. coli* O157:H7 and *Salmonella enterica* in the primary vegetable production chain. *Critical Reviews in Microbiology, 34*(3-4), 143-161.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.

Golden, C. E., Rothrock, M. J., & Mishra, A. (2019). Using farm practice variables as predictors of *Listeria* spp. prevalence in pastured poultry farms. *Frontiers in Sustainable Food Systems*, *3*, 15.

Greene, S., Daly, E., Talbot, E., Demma, L., Holzbauer, S., Patel, N., Hill, T., Walderhaug, M., Lynch, M. (2008). Recurrent multistate outbreak of *Salmonella* Newport associated with tomatoes from contaminated fields, 2005. *Epidemiology & Infection, 136*(2), 157-165.

Greenwell, B. M. (2017). pdp: An R Package for Constructing Partial Dependence Plots. *R Journal, 9*(1).

Guo, X., Chen, J., Brackett, R. E., & Beuchat, L. R. (2002). Survival of *Salmonella* on tomatoes stored at high relative humidity, in soil, and on tomatoes in contact with soil. *Journal of Food Protection, 65*(2), 274-279.

Holley, R. A., Arrus, K. M., Ominski, K. H., Tenuta, M., & Blank, G. (2006). *Salmonella* survival in manure-treated soils during simulated seasonal temperature exposure. *Journal of Environmental Quality, 35*(4), 1170-1180.

ICMSF, 1996. Salmonellae. In: ICMSF (Ed.), *Microorganisms in Foods 5*, Chapman and Hall, London, UK, p. 126–140.

Ivanek, R., Gröhn, Y., Wells, M., Lembo, A., Sauders, B., & Wiedmann, M. (2009). Modeling of spatially referenced environmental and meteorological factors influencing the probability of *Listeria* species isolation from natural environments. *Applied and Environmental Microbiology, 75*(18), 5893-5909.

Jacobsen, C. S., & Bech, T. B. (2012). Soil survival of *Salmonella* and transfer to freshwater and fresh produce. *Food Research International, 45*(2), 557-566.

Kumar, G. D., Williams, R. C., Al Qublan, H. M., Sriranganathan, N., Boyer, R. R., & Eifert, J. D. (2017). Airborne soil particulates as vehicles for *Salmonella* contamination of tomatoes. *International Journal of Food Microbiology*, *243*, 90-95.

Kumar, G. D., Williams, R. C., Sriranganathan, N., Boyer, R. R., & Eifert, J. D. (2018). Survival of tomato outbreak associated *Salmonella* serotypes in soil and water and the role of biofilms in abiotic surface attachment. *Foodborne Pathogens and Disease*, *15*(9), 548-553.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18-22.

Liu, C., Hofstra, N., & Franz, E. (2013). Impacts of climate change on the microbial safety of pre-harvest leafy green vegetables as indicated by *Escherichia coli* O157 and *Salmonella* spp. *International Journal of Food Microbiology, 163*(2-3), 119-128.

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: A Package for Binary Imbalanced Learning. *R Journal, 6*(1).

Micallef, S. A., & Kumar, G. D. (2017). Biofilms: a community-based strategy for bacterial persistence and relevance to food safety. In *Trends in Food Safety and Protection* (pp. 107-130). CRC Press.

NOAA, Wind Gust. Retrieved Febuary 15, 2019, from https://graphical.weather.gov/definitions/defineWindGust.html

Oberholtzer, L., Greene, C., & Lopez, E. (2006). Organic poultry and eggs capture high price premiums and growing share of specialty markets. Washington, DC: US Department of Agriculture, Economic Research Service. *Citeseer.*

Oliveira, C., Carvalho, L., & Garcia, T. (2006). Experimental airborne transmission of *Salmonella* Agona and *Salmonella* Typhimurium in weaned pigs. *Epidemiology & Infection, 134*(1), 199-209.

Oni, R. A., Sharma, M., & Buchanan, R. L. (2015). Survival of *Salmonella enterica* in dried turkey manure and persistence on spinach leaves. *Journal of Food Protection, 78*(10), 1791-1799.

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., Nightingale, K., Anciso, J., Jun, M., Han, D. (2014). Farm management, environment and weather factors jointly affect the probability of spinach contamination with generic *Escherichia coli* at the preharvest level. *Applied and Environmental Microbiology*, 03643-03613.

Rothrock, M. J., Hiett, K. L., Guard, J. Y., & Jackson, C. R. (2016). Antibiotic resistance patterns of major zoonotic pathogens from all-natural, antibiotic-free, pasture-raised broiler flocks in the southeastern United States. *Journal of Environmental Quality, 45*(2), 593-603.

Siemon, C. (2007). Comparative investigation of prevalence and antimicrobial resistance of Salmonella between pasture and conventionally reared poultry. *Avian Diseases., 51*(1), 112.

Sossidou, E., Dal Bosco, A., Elson, H., & Fontes, C. (2011). Pasture-based systems for poultry production: implications and perspectives. *World's Poultry Science Journal, 67*(1), 47-58.

Strawn, L. K., Fortes, E. D., Bihn, E. A., Nightingale, K. K., Gröhn, Y. T., Worobo, R. W., Wiedmann, M., 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.

Trimble, L. M., Alali, W. Q., Gibson, K. E., Ricke, S. C., Crandall, P., Jaroni, D., Berrang, M., Habteselassie, M. Y. (2013). Prevalence and concentration of *Salmonella* and *Campylobacter* in the processing environment of small-scale pastured broiler farms. *Poultry Science, 92*(11), 3060-3066.

Van Doren, J. M., Kleinmeier, D., Hammack, T. S., & Westerman, A. (2013). Prevalence, serotype diversity, and antimicrobial resistance of *Salmonella* in imported shipments of spice offered for entry to the United States, FY2007–FY2009. *Food Microbiology, 34*(2), 239-251.

Winfield, M., & Groisman, E. (2003). Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Applied and Environmental Microbiology., 69*(7), 3687-3694.

Table 3.1. Comparison of the 11 all-natural, antibiotic-free, pastured broiler farms included in this study.

Farm	Breed	No. of flocks	Flock size	Multi-use farm?	Animal types	Processing
A	Freedom Ranger	10	>500	Yes	Layers, Swine, Beef cattle, Sheep	USDA-inspected facility
B	Freedom Ranger, Cornish Cross	5	50-75	Yes	Layers, Swine, Horses, Goats	On-farm (skin-off)
C	Freedom Ranger	1	50-75	No	n/a	On-farm (skin-on)
D	Freedom Ranger	1	50-75	No	n/a	On-farm (skin-on)
E	Freedom Ranger, Cornish Cross	5	50-75	Yes	Layers, Swine, Beef cattle, Sheep	On-farm (skin-on)
F	Freedom Ranger	2	>500	Yes	Layers	USDA-inspected facility
G	Freedom Ranger, Cornish Cross	9	100-500	Yes	Layers, Swine, Goats	USDA-inspected facility
H	Freedom Ranger, Cornish Cross	2	50-75	Yes	Layers	On-farm (skin-on)
I	Freedom Ranger	4	100-500	Yes	Layers, Beef cattle, Goats	USDA-inspected facility and on-farm (skin-on)
J	Freedom Ranger	2	>500	Yes	Layers, Swine, Beef cattle, Sheep	USDA-inspected facility
K	Cornish Cross	2	50-75	Yes	Layers, Swine	On-farm (skin-on)

Table 3.2. Effect of season on prevalence of *Salmonella* in pastured poultry farms

Season	Month	No. (%) of positive *Salmonella*	
		Soil samples	Fecal samples
Spring[a]	March-May	2/91 (2)	29/131(22)
Summer[b]	June-August	47/433 (11)	83/473(18)
Fall[b]	September-November	7/164(4)	13/170(8)
Winter[ab]	December-February	0/5(0)	0/5(0)

[a] The different letters represent statically different values, p values < 0.05 from Bonferroni corrected Fisher's exact test

Table 3.3. Predictive performance of the soil and feces random forest models

Models		Actual +	Actual -	Sensitivity	Specificity	AUC*
Soil	+	14	28	0.875	0.844	0.884
Predicted	-	2	151			
Feces	+	30	20	0.811	0.894	0.873
Predicted	-	7	169			

*AUC is the area under receiver operating characteristic curve
"+": *Salmonella* isolated
"-": *Salmonella* not isolated

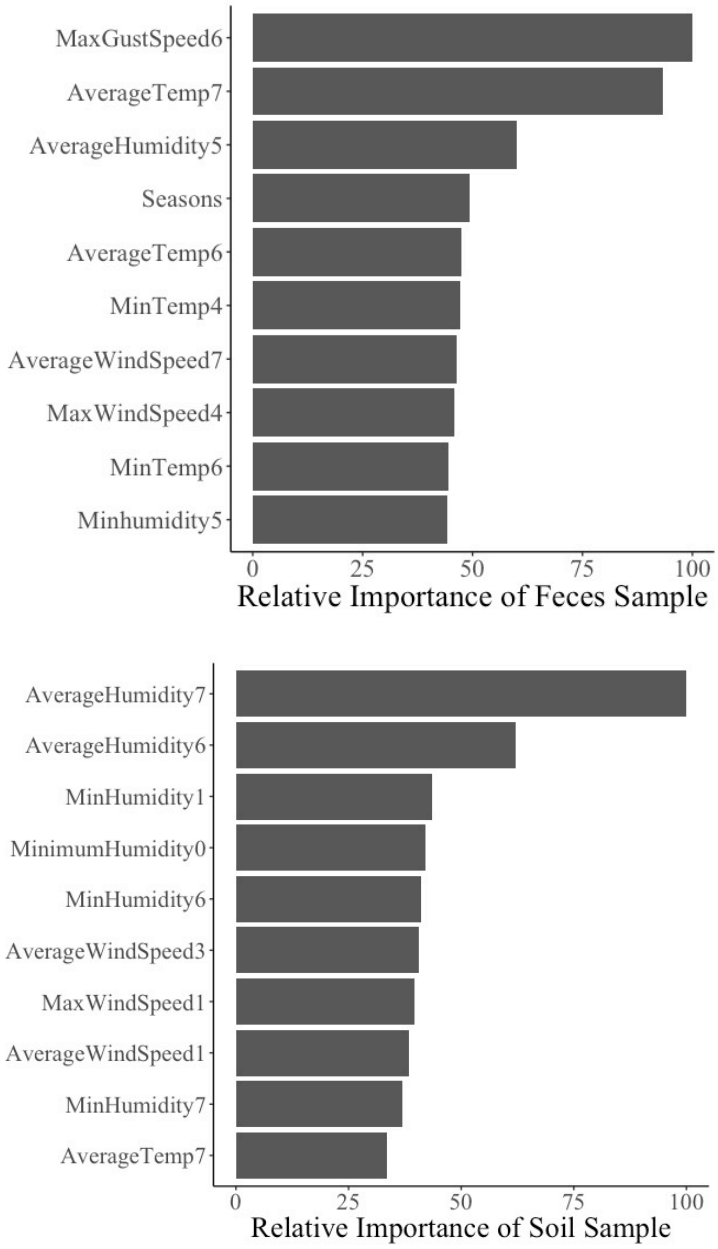


Figure 3.1. Top 10 Relative importance variables for feces and soil model were ranked from the most important variable.

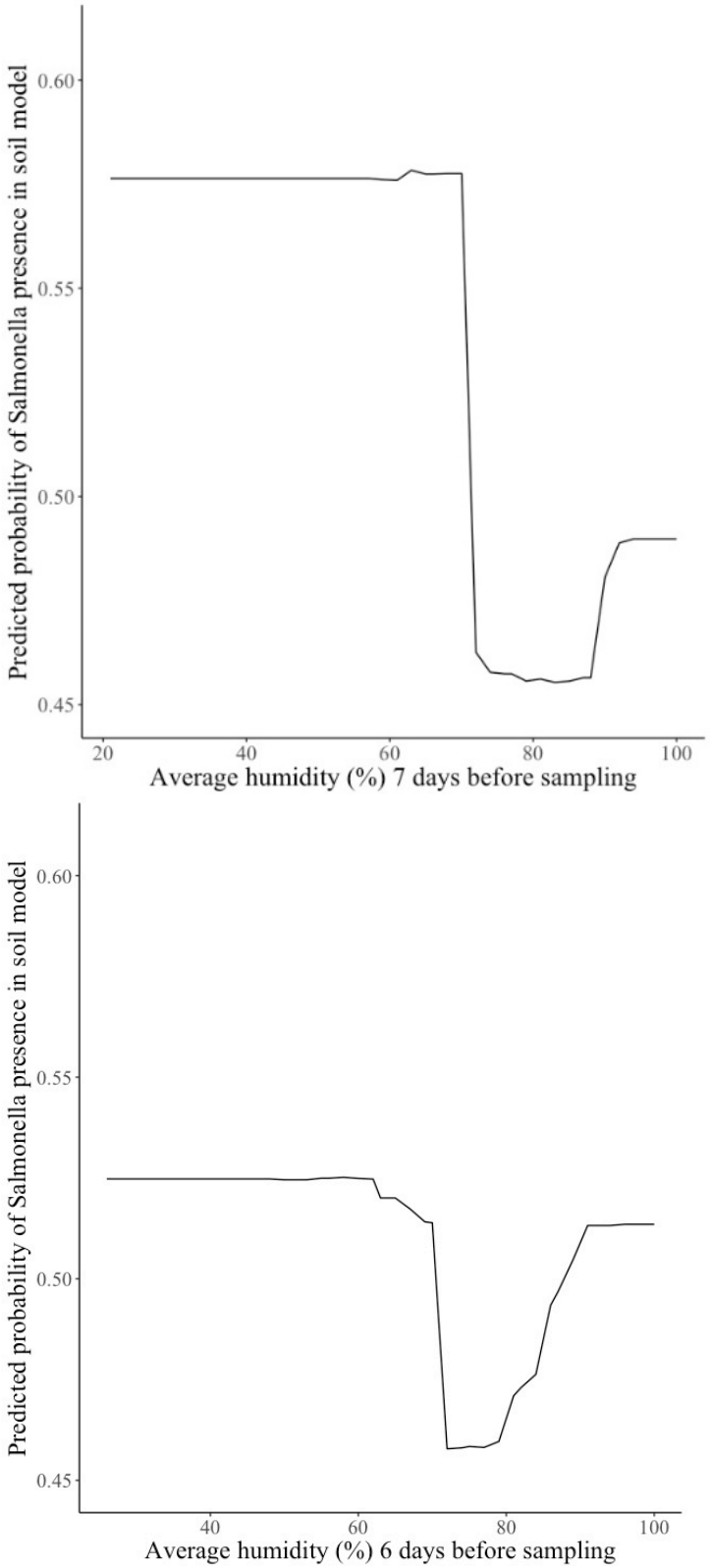


Figure 3.2. Partial dependency plots of soil model for top 1 and 2 predicting variables, average humidity 7 days prior to sampling date and average humidity 6 days prior to the sampling date.

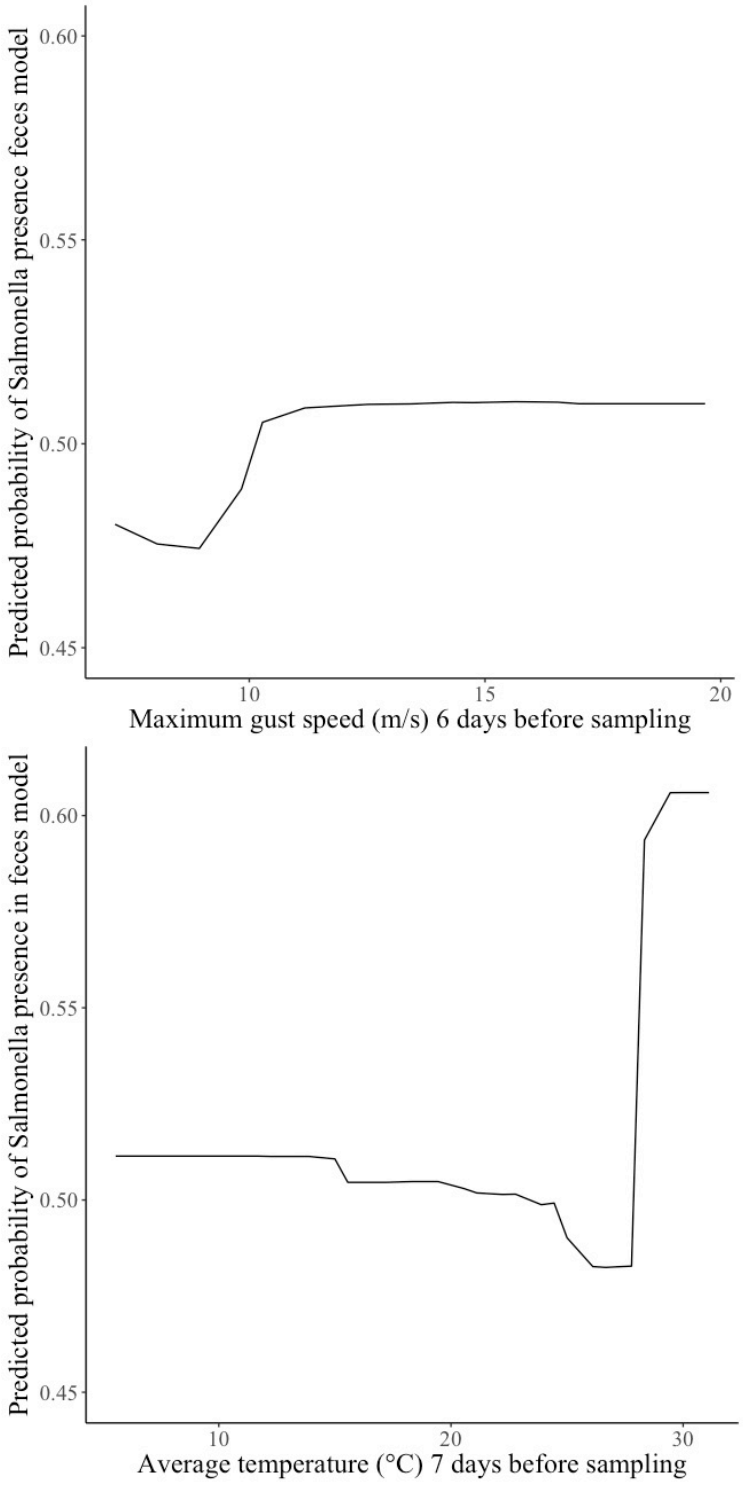


Figure 3.3. Partial dependency plots of feces model for top 1 and 2 predicting variables, Maximum gust speed (m/s) 6 days prior sampling date and Average temperature 7 days prior sampling date.

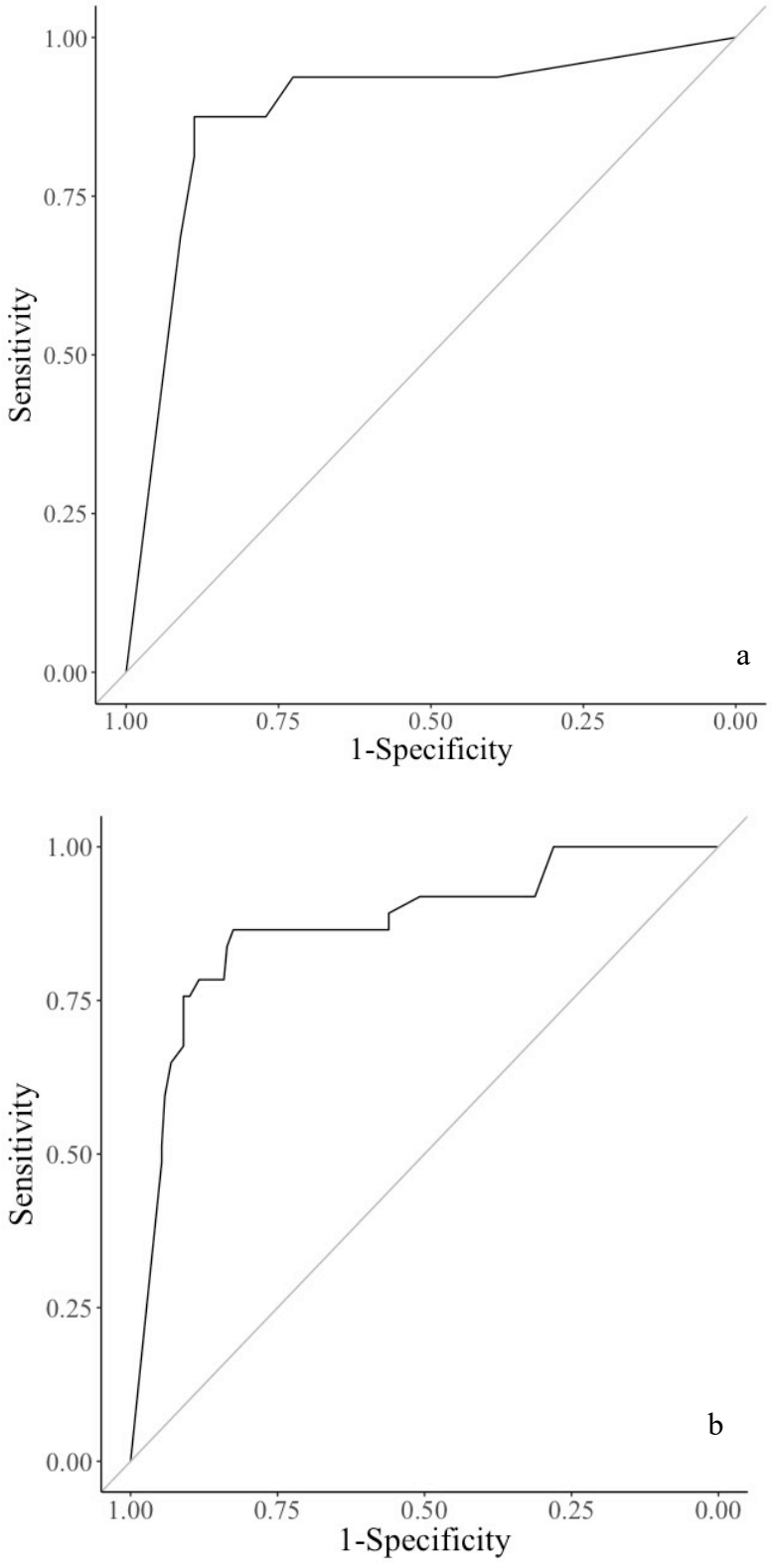


Figure 3.4. Receiver operating characteristic curves for soil model (a) and feces model (b)

CHAPTER 4

FARM MANAGEMENT PRACTICES THAT EFFECT THE PREVALENCE OF *SALMONELLA* IN PASTURED POULTRY FARMS[2]

[2] Hwang, D., Rothrock, M. Jr., Pang, H., Mishra, A. To be submitted *LWT*

Abstract

Salmonella is a foodborne pathogen that has been identified as a public health threat in the poultry industry. Farm practices can affect the prevalence of *Salmonella* in the final product when poultry are exposed to the outside environment. Pastured poultry farms in the Southeastern United States were investigated in this study. Farm practice and processing variables that may affect the presence of *Salmonella* were determined by developing predictive models using the random forest method. Important variables were identified in preharvest (feces and soil), and postharvest (whole carcass rinses) samples. Predictive models were generated with each type of sample, and the models were tested with the corresponding test set. The model performances were measured by the area under curve (AUC) values from the receiver operating characteristic (ROC) curve. All models developed in this study were robust in predicting *Salmonella* presence, with AUC values above 0.83. It was found that as the number of years spent farming increased, there was increase in predicted probability of finding *Salmonella*. First 3 ingredients in the feed type was identified as the top predicting variables for both preharvest and postharvest variables. The predicted probability of *Salmonella* presence decreased when the flock age was between 70 to 85 days.

Introduction

Salmonella is responsible for causing approximately 1.1 million cases of infection in the United States annually (CDC, 2010). Salmonellosis is a human illness that is caused by *Salmonella* when contaminated food or water is ingested. Live poultry has

been identified to be a reservoir of *Salmonella* where it is found in the gastrointestinal tract. The bacteria then are shed through feces, which can become source for contamination in foods (Thorns, 2000). It has been known to survive multiple farm environments outside of a host, such as farm environments like irrigation water, soil, manure and compost (Winfield & Groisman, 2003). According to the CDC, live poultry has also been identified as a route to infect humans, causing outbreaks that hospitalized 360 and death of 4 people from 2015 to 2017. These outbreaks can be traced back to small-scale farms where safety practices in poultry production can often be neglected. Some other recent outbreaks from poultry sources include those from raw chicken, turkey, poultry products, and shell eggs (CDC, 2018b).

Pastured poultry farms fall into a category of small alternative farms. The organic and pastured poultry products have increased in popularity among American consumers. The popularity rose from the attention on sustainability and welfare of the birds (Fanatico & Born, 2002). There are no United States Department of Agriculture (USDA) labeling guidelines for pastured poultry currently. Pastured poultry farms rear the birds in movable pens outdoors. These pens are moved daily, and the birds are continuously exposed to the fresh pasture (APPA 2019, Sossidou, 2011). The public assumption is that having access to outside where the birds freely roam would increase their welfare, therefore leading to more ethical, better tasting, and safer ways to grow chickens (Brennan et al., 2003; Hughner et al., 2007; Yiridoe et al., 2005). However, there is little scientific evidence that these poultry products are safer than the conventional equivalents. By having continuous exposure to the pasture, many extraneous variables could influence

the microbial safety of the pastured poultry (Kumar, Williams, Qublan, Sriranganathan, Boyer, & Eifert, 2017; Trimble, 2013, Siemon, 2007).

Machine learning techniques such as random forest use predictive modeling that can be applied in the food safety context. Important factors contributing to the presence of foodborne pathogens such as *Salmonella* in the food production environments can be identified using these methods. Farm management practices and processing variables were included in the models to evaluate which of the factors are associated with *Salmonella* presence in the pastured poultry farms. Knowing these factors can aid the decision-making process in the management, ultimately to reduce the *Salmonella* contamination. The objective of this study was to develop models to predict the prevalence of *Salmonella* in the pastured poultry farms using farm management practices as predicting variables. The random forest method was used in the models to identify which variables contributed to the prevalence of *Salmonella*.

Materials and Methods

Sample Collection

A longitudinal study was conducted on 43 flocks of broilers across 11 pastured poultry farms in the southeastern United States from March 2014 to November 2017. All 11 farms reared their broiler flocks in movable pens with temporary fences. A brief description of the size and scale of each farm is presented in Table 3.1. Data were collected for major farm practice variables (Table 4.1) over a flock's lifecycle and all samples were evaluated for the presence of *Salmonella*.

The following samples were collected for each flock: (i) feces, (ii) pasture soil, (iii) whole carcass rinse (WCR) directly after processing, and (iv) final product WCR after

chilling and storage time. The carcass had been stored at the temperature and time that is included in Table 4.2. All samples were collected in the field and returned to the lab in a cooler packed in ice. For the grow-out samples (feces and soil), samples were collected from the pasture where the flock was residing at the time of sampling. At each sampling time, the pasture area was divided into five separate sections and five subsamples in each section were pooled into a single sample for each section (five fecal and five soil samples were collected on each sampling day). The total amount of sample collected for each field sample was at least 25 g. For the processing and final product samples (WCR), five pooled samples were collected for each sample type, with each pooled sample containing WCR from five carcasses. WCR-p samples were the samples after whole carcass rinse but prior to the storage period. WCR-f samples were the final product after the storage period.

To prepare the environmental samples for homogenization, 3 g (feces, soil) were combined within filtered stomacher bags and diluted 1:3 using 1X phosphate-buffered saline (PBS). All samples were homogenized for 60 sec and these homogenates were used for all downstream cultural isolations. For the WCR, 100 ml of 1X PBS was added to each carcass within the storage bag, and the bags were vigorously shaken for 60 sec. Rinsates were collected from 25 carcasses, with five WCR being pooled into a single filtered stomaching bag a total of five times. No further dilution in 1X PBS was required for the WCR samples.

Cultural Isolation Methods

Sample collection and cultural isolation methods were adapted from Rothrock et al., (2016). Briefly, as a pre-enrichment step, the stomached homogenates remained in the filtered stomacher bags and were incubated overnight at 35 °C. Two different enrichments

broths were used to isolate *Salmonella* from these environmental samples: tetrathionate (TT; Becton-Dickinson, Sparks, MD) broth and Rappaport-Vassiliadis (RV; Becton Dickinson) media. After overnight incubation at 42 °C in both of these enrichment broths, one loopful from each enrichment broth was spread on two different differential media: brilliant green sulfa with novobiocin (BGS; Becton Dickinson) agar and xylose lysine tergitol-4 (XLT-4; Becton Dickinson) agar. These plates were incubated overnight at 35 °C, and on each plate, three *Salmonella*–like colonies per subsample were selected and confirmed using triple sugar iron agar (TSI; Becton-Dickinson) and lysine iron agar fermentation (LIA; Becton-Dickinson) using an incubation period of 18-24 hours at 35°C, Final confirmation of suspect TSI/LIA isolates was performed using *Salmonella* polyvalent O antiserum agglutination (Becton-Dickinson), using manufacturer's specifications.

Statistical Analyses and Modeling

All statistical analyses were performed by using the R software (Version 3.4.0; R Foundation for Statistical Computing, Vienna, Austria). Random forest models were generated with the farm practice variables using the caret package in R (Kuhn, 2008). There were four sample types: feces, soil, WCR-p, and WCR-f. There were four models developed in total for each of the sample types. The farm practice variables included preharvest variables such as the water source, feed used for brooding and in the pasture, and age of flocks. The final product, post-harvest variables included in the model were chilling time, chilling temperature, and chilling methods (Table 4.2).

The variable importance plots were created with the top 10 predicting variables that contribute to the prevalence of *Salmonella* in each type of sample. Partial dependency plots were then generated to show the relationship between the most important variable and the

probability of predicting *Salmonella*. Receiver operating characteristic (ROC) curves were plotted using the sensitivity and specificity values from the models generated. Area under curve (AUC) values from the ROC curve were retrieved to evaluate the performance of each model. The AUC values and the confusion matrix were presented in Table 4.3 (Golden, Rothrock, & Mishra, 2019; Golden, Rothrock, & Mishra, 2019b).

Results

Of the 726 fecal samples collected, 119 of them were *Salmonella* positive (16.4%), while only 12.3% (89/721) of soil samples possessed *Salmonella.* There were two types of whole carcass rinse (WCR) samples, post-processing (WCR-p) and final product (WCR-f). Depending on the farm and its practices, there were differences in the post-processing storage time and storage temperature (Table 4.2). *Salmonella* was found in 23.5% (119/235) and 17.5% (40/229) of the WCR-p samples had WCR-f samples, respectively.

The *Salmonella* prevalence data were divided into four seasons, and season was included as one of the predictor variables in the developed models. Overall, within the cultural data collected, the summer months had the highest *Salmonella* presence (Table 4.4). The highest prevalence, in percentage, was observed in WCR-f samples, during winter months (40%). However, there were only 20 samples collected during the winter months, which is much smaller number of samples collected compared to other seasons. In the models generated, the seasons were one of the top 10 predicting variable from the relative importance plots for the whole carcass rinse samples.

Random forest models were created for each of the four sample types. The relative importance plots with the top ten predicting variables are illustrated in Figure 4.1.

The relative importance plot ranks the variable with the highest importance score of 100. There were 31 predicting preharvest variables that were included in the model development, and nine additional variables (processing related) for post-harvest samples. The explanation of each farm practice variables and the levels incorporated have been included in Table 4.1. The number of years farming and pasture feed composition (as defined by the top three ingredients) were the top two variables for predicting *Salmonella* presence in fecal model (Fig. 4.1A). The years farming and brood feed composition (as defined by the top three ingredients) were the variables most associated with *Salmonella* presence for the soil model (Fig. 4.1B). While brood feed was also an important factor for predicting *Salmonella* presence in post-harvest samples, flock age was more important for the WCR-p samples (Fig 4.1C). Also, the years farming, again, was associated with *Salmonella* presence in WCR-f samples (Fig. 4.1D).

Partial dependency plots (PDPs) were created from the top two important variables for each sample type. The PDPs can be used to understand the effect of an individual predicting variable has on the predictive probability of detecting *Salmonella* presence. For all the models that predicted years farming as variable of importance, as the years of farming increased, the probability of positive prediction from the model increased. These include feces, soil and WCR-f models (Figs. 4.2, 4.3 ,4.5, respectively).

In the fecal model, pasture feed was also identified as an important variable, where the ingredients “corn, cotton seed mill, grain product (CMGP)”, “corn, soy, sorghum (CSH)”, “corn, soy, wheat (CSW)”, and “pea, corn, oats (PCO)” had higher predicted probability than other feeds (Fig. 4.2). Soil, WCR-p and WCR-f models identified the brood feed as the important variable. There were similarities between the

brood and pasture feed, but brood feed was only fed to the chicken during the brooding period (0 to 21, or 28 days old, depending on time of year). The feeds CSH, CSW, and PCO were consistently identified as the feeds that were associated with *Salmonella* prevalence. Corn was the most popular first ingredient of choice of all the feeds used by the farmers.

Receiver operating characteristics (ROC) curves were used to evaluate the predictive performances of each model (Fig. 4.6). The sensitivity (the true positive detection rate) and specificity (the true negative detection rate) of the model predictions were used in generating the ROC curves. The ideal model will produce sensitivity, specificity, and AUC under ROC curve values of 1.00. The preharvest models, feces and soil, had AUC value of 0.93 and 0.92, respectively. The postharvest models, WCR-p and WCR-f had AUC value of 0.87 and 0.83, respectively.

Discussion

Different environmental factors contribute to *Salmonella* presence in pastured poultry farm settings. The birds exposed to the outside environment are susceptible to increased risks from predators, pests, and weather conditions. This study investigated the different farm management practices as the factors that affect the presence of *Salmonella* in the pastured poultry environmental and processing samples. The survival and growth of *Salmonella* varies depending on the soil qualities such as temperature, moisture and nutrient availability (Jacobsen & Bech, 2012); however, once *Salmonella* is found in the environment, it can survive and even multiply when conditions are met (Chandler & Craven, 1980). Therefore, having a better understanding of how different farming practices that affect the presence of *Salmonella* can reduce the risk of foodborne illness.

The developed random forest models showed the predictive capability where the models can be applied to predict the prevalence with specific management practices in the pastured poultry farms.

Random forest models were used previously to make casual inferences to predict the prevalence of foodborne pathogens (Chapin, Nightingale, Worobo, Wiedmann, & Strawn, 2014; Pang et al., 2017). The models can be interpreted using the variable importance plots and the partial dependency plots. The relative importance plots provide the ranking system of the variables, where the top variable is given an importance score of 100. The PDPs show the marginal effect of the single important variable chosen, visualizing the relationship between the single predicting variable and the predicted probability. In this study, the variables years farming, flock age, brood feed and the pasture feed were identified as the contributing variables for *Salmonella* prevalence. The predicting performances of the models were evaluated with the area under ROC curve. The models created in this study were robust in predicting the presence of *Salmonella* in pastured poultry settings.

The variable "years farming" highly influenced the presence of *Salmonella* for all samples except the WCR-p sample. Eleven farms in the southeast U.S. were included in this study, and the farms had up to 16 years of farming experience. For the feces and soil models, the predicted probability increases and peaks when years farming reached 9 years. In the WCR-p samples, the predicted probability was observed to gradually increase and reached maximum at 16 years of farming. This finding is probably due to *Salmonella* accumulating and adapting to the environment over time as well as forming biofilms (Kumar, Willians, Srianganathan, Boyerm & Eifert, 2018; Micallef & Kumar,

2017). Therefore, the prevalence increases as the number of years of farming increases. *Salmonella* prevalence from the environment increased over time in a previous study, in layer flocks in the United Kingdom (Wales et al., 2007).

Additional factors that were identified to be highly associated with *Salmonella* presence were pasture feed and brood feed. Among these feeds, corn was the most popular choice used as a first ingredient, 5 out of 8 feed types. Other top ingredients included pea and wheat. The top 3 ingredients sources are included in the Table 4.1. The models identified the feed types PCO, CSH, and CSW to be higher association with the presence of *Salmonella*. The fecal model also identified the same ingredients, in addition to CMGP, where the WCR-f model additionally identified barley, wheat, oats, (BWO). Although pasture feed was provided, it was also likely that the birds foraged and grazed from the pasture itself. The source of *Salmonella* may have been from food birds foraged themselves as well as the contaminated feed that may have been provided to them. *Salmonella* has previously been isolated in the commercial chicken feed, which introduces the bacteria to the birds and the environment (Bucher et al., 2007).

Since the pastured poultry are continuously exposed to the outdoors, the seasonal temperatures and other meteorological factors can affect the *Salmonella* exposures. The variable "seasons" was included in the top 10 important predicting variables for whole carcass rinse samples. Season may play a role in *Salmonella* presence depending on the time of year sampled. There was also variability in the storage conditions depending on the farm (Table 4.2). *Salmonella* presence during the summer season was the highest. This finding is likely due to warmer temperature and higher amounts of rain fall (Benjamin, 2013). Previous study identified summer was associated higher *Salmonella*

occurance with retail poultry products (Zdragas, Mazaraki, Vafeas, Giantzi, Papadopoulos & Ekateriniadou, 2012). The flock age was also identified as important variable for WCR-p sample. In a previous study, age of flocks was significantly associated with *Salmonella* prevalence in the litter in broiler farm (Renwick et al., 1992). In this study, decrease in predicted probability was observed for flock age 75 to 84 days; however, it is difficult to make the comparison from previous findings, since the commercial broilers harvested earlier in market age than broilers from this study.

Salmonella prevalence increased as the years spent farming increased. To reduce the prevalence of *Salmonella*, it is recommended to have a regular practice of cleaning after each moving of the pens and sanitation of the equipment. Also, worker training in food safety as well as biosafety can reduce the spreading of *Salmonella* in farm environments. Type of feed was also frequently identified as the top contributor in *Salmonella* presence in pastured poultry farms. Diets high in corn may affect the microbiota of intestines of poultry. Supplements such as probiotics and prebiotics for poultry can strengthen the digestive tracts and may help to decrease *Salmonella* counts by competition.

Salmonella presence in pastured poultry farms is a serious food safety concern. In conclusion, findings from this study show that the presence of *Salmonella* is affected by the farm management practices such as type of feed given, age of flocks (days), and number of years farming. In order to reduce the incidence of pathogens such as *Salmonella*, having an understanding of models developed in this study can serve as a useful tool for farmers and poultry producers to decrease the probability of finding *Salmonella* in pastured poultry farms.

References

Alali, W. Q., Thakur, S., Berghaus, R. D., Martin, M. P., & Gebreyes, W. A. (2010). Prevalence and distribution of *Salmonella* in organic and conventional broiler poultry farms. *Foodborne Pathogens and Disease, 7*(11), 1363-1371.

APPA. (2019). American Poultry Producers Association. Retrieved April 15, 2019, from https://www.apppa.org/.

Bailey, J., & Cosby, D. (2005). *Salmonella* prevalence in free-range and certified organic chickens. *Journal of Food Protection, 68*(11), 2451-2453.

Baloda, S. B., Christensen, L., & Trajcevska, S. (2001). Persistence of a *Salmonella enterica* Serovar Typhimurium DT12 Clone in a piggery and in agricultural soil amended with *Salmonella*-contaminated slurry. *Applied and Environmental Microbiology, 67*(6), 2859-2862.

Baptista, F., Alban, L., Ersbøll, A. K., & Nielsen, L. R. (2009). Factors affecting persistence of high *Salmonella* serology in Danish pig herds. *Preventive Veterinary Medicine, 92*(4), 301-308.

Bäumler, A. J., Tsolis, R. M., & Heffron, F. (1996). Contribution of fimbrial operons to attachment to and invasion of epithelial cell lines by *Salmonella* Typhimurium. *Infection and Immunity, 64*(5), 1862-1865.

Benjamin, L. (2013). Occurrence of generic *Escherichia coli, E. coli* O157 and *Salmonella spp*. in water and sediment from leafy green produce farms and

streams on the Central California coast. *International Journal of Food Microbiology., 165*(1), 65.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Breiman, L., Friedman, J., & Olshen, R. (1984). Stone CJ. Classification and regression trees. *Wadsworth Wadsworth statistics/probability series*: Belmont, CA: CRC Press.

Brennan, C., Gallagher, K., & McEachern, M. (2003). A review of the 'consumer interest' in organic meat. *International Journal of Consumer Studies, 27*(5), 381-394.

Bucher, O., Holley, R., Ahmed, R., Tabor, H., Nadon, C., Ng, L., & D'AOUST, J.-Y. (2007). Occurrence and characterization of *Salmonella* from chicken nuggets, strips, and pelleted broiler feed. *Journal of Food Protection, 70*(10), 2251-2258.

CDC. (2010). What is Salmonellosis? Retrieved April 15, 2019, from http://www.cdc.gov/*Salmonella*/-general/index.html.

CDC. (2018a). National Outbreak Reporting System (NORS). Retrieved April 15, 2019, from https://wwwn.cdc.gov/norsdashboard/.

CDC. (2018b). Reports of Selected *Salmonella* outbreak investigations. Retrieved April 15, 2019, from https://www.cdc.gov/Salmonella/outbreaks.html.

Chandler, D. S., and J. A. Craven. (1980). Relationship of soil moisture to survival of *Escherichia coli* and *Salmonella* typhimurium in soils. *Australian Journal of Agricultural Research* 31(3), 547-555.

Chapin, T. K., Nightingale, K. K., Worobo, R. W., Wiedmann, M., & Strawn, L. K. (2014). Geographical and meteorological factors associated with isolation of

Listeria species in New York State produce production and natural environments. *Journal of Food Protection, 77*(11), 1919-1928.

Cherry, W. B., Hanks, J. B., Thomason, B. M., Murlin, A. M., Biddle, J. W., & Croom, J. M. (1972). *Salmonella*e as an index of pollution of surface waters. *Applied and Environmental Microbiology, 24*(3), 334-340.

Cooley, M., Carychao, D., Crawford-Miksza, L., Jay, M. T., Myers, C., Rose, C., Mandrell, R. E. (2007). Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California. *PloS One, 2*(11), e1159.

D'Souza, R. M., Becker, N. G., Hall, G., & Moodie, K. B. (2004). Does ambient temperature affect foodborne disease? *Epidemiology, 15*(1), 86-92.

Dimitri, C., & Greene, C. R. (2002). Recent growth patterns in the US organic foods market. Retrieved March 4, 2019, from https://ageconsearch.umn.edu/record/33715/files/ai020777.

Dowe, M. J., Jackson, E. D., Mori, J. G., & Bell, C. R. (1997). *Listeria monocytogenes* survival in soil and incidence in agricultural soils. *Journal of Food Protection, 60*(10), 1201-1207.

Fanatico, A., & Born, H. (2002). Label Rouge: Pasture-based poultry production in France. *Livestock Technical Note* Retrieved March 9, 2019, from (http://www.attra.ncat.org/attra-pub/labelrouge. html).

Fanatico, A., Pillai, P. B., Emmert, J., & Owens, C. (2007). Meat quality of slow-and fast-growing chicken genotypes fed low-nutrient or standard diets and raised indoors or with outdoor access. *Poultry Science, 86*(10), 2245-2255.

Finstad, S., O'Bryan, C. A., Marcy, J. A., Crandall, P. G., & Ricke, S. C. (2012). *Salmonella* and broiler processing in the United States: relationship to foodborne salmonellosis. *Food Research International, 45*(2), 789-794.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.

Garcia, R., Baelum, J., Fredslund, L., Santorum, P., & Jacobsen, C. S. (2010). Influence of temperature and predation on survival of *Salmonella enterica* serovar Typhimurium and expression of invA in soil and manure-amended soil. *Applied and Environmental Microbiology, 76*(15), 5025-5031.

Garcia-del Portillo, F., Foster, J. W., & Finlay, B. (1993). Role of acid tolerance response genes in *Salmonella* Typhimurium virulence. *Infection and Immunity, 61*(10), 4489-4492.

Golden, C. E., Rothrock Jr, M. J., & Mishra, A. (2019b). Comparison between random forest and gradient boosting machine methods for predicting *Listeria* spp. prevalence in the environment of pastured poultry farms. *Food Research International*.

Golden, C. E., Rothrock, M. J., & Mishra, A. (2019). Using farm practice variables as predictors of *Listeria* spp. prevalence in pastured poultry farms. *Frontiers in Sustainable Food Systems*, *3*, 15.

Gorski, L., Parker, C. T., Liang, A., Cooley, M. B., Jay-Russell, M. T., Gordus, A. G., Atwill, E.R., & Mandrell, R. E. (2011). Prevalence, distribution, and diversity of *Salmonella enterica* in a major produce region of California. *Applied and Environmental Microbiology, 77*(8), 2734-2748.

Guard-Petter, J. (2001). The chicken, the egg and *Salmonella enteritidis*. *Environmental Microbiology, 3*(7), 421-430.

Guo, X., Chen, J., Brackett, R. E., & Beuchat, L. R. (2002). Survival of *Salmonella* on tomatoes stored at high relative humidity, in soil, and on tomatoes in contact with soil. *Journal of Food Protection, 65*(2), 274-279.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148*(3), 839-843.

Harrell Jr, F. E. (2015). Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: *Springer*.

Hilimire, K. (2012). The grass is greener: Farmers' experiences with pastured poultry. *Renewable Agriculture and Food Systems, 27*(3), 173-179.

Holley, R. A., Arrus, K. M., Ominski, K. H., Tenuta, M., & Blank, G. (2006). *Salmonella* survival in manure-treated soils during simulated seasonal temperature exposure. *Journal of Environmental Quality, 35*(4), 1170-1180.

Hughner, R. S., McDonagh, P., Prothero, A., Shultz, C. J., & Stanton, J. (2007). Who are organic food consumers? A compilation and review of why people purchase organic food. *Journal of Consumer Behaviour: An International Research Review, 6*(2-3), 94-110.

Ivanek, R., Gröhn, Y., Wells, M., Lembo, A., Sauders, B., & Wiedmann, M. (2009). Modeling of spatially referenced environmental and meteorological factors influencing the probability of *Listeria* species isolation from natural environments. *Applied and Environmental Microbiology, 75*(18), 5893-5909.

Jacobsen, C. S., & Bech, T. B. (2012). Soil survival of *Salmonella* and transfer to freshwater and fresh produce. *Food Research International, 45*(2), 557-566.

James, C., Vincent, C., de Andrade Lima, T., & James, S. (2006). The primary chilling of poultry carcasses—a review. *International Journal of Refrigeration, 29*(6), 847-862.

Jones, T. F., Ingram, L. A., Cieslak, P. R., Vugia, D. J., Tobin-D'Angelo, M., Hurd, S., Medus, C., Cronquist, A., & Angulo, F. J. (2008). Salmonellosis outcomes differ substantially by serotype. *The Journal of Infectious Diseases, 198*(1), 109-114.

Kovats, R. S., Edwards, S. J., Hajat, S., Armstrong, B. G., Ebi, K. L., & Menne, B. (2004). The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten European countries. *Epidemiology and Infection, 132*(3), 443-453. doi:10.1017/s0950268804001992

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1-26.

Kumar, G. D., Williams, R. C., Al Qublan, H. M., Sriranganathan, N., Boyer, R. R., & Eifert, J. D. (2017). Airborne soil particulates as vehicles for *Salmonella* contamination of tomatoes. *International Journal of Food Microbiology*, *243*, 90-95.

Kumar, G. D., Williams, R. C., Sriranganathan, N., Boyer, R. R., & Eifert, J. D. (2018). Survival of tomato outbreak associated *Salmonella* serotypes in soil and water and the role of biofilms in abiotic surface attachment. *Foodborne Pathogens and Disease*, *15*(9), 548-553.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18-22.

Liu, C., Hofstra, N., & Franz, E. (2013). Impacts of climate change on the microbial safety of pre-harvest leafy green vegetables as indicated by *Escherichia coli* O157 and *Salmonella* spp. *International Journal of Food Microbiology, 163*(2-3), 119-128.

Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1*(1), 14-23.

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: A Package for binary imbalanced learning. *R Journal, 6*(1), 79-89.

McCrea, B., Tonooka, K., VanWorth, C., Boggs, C., Atwill, E., & Schrader, J. (2006). Prevalence of *Campylobacter* and *Salmonella* species on farm, after transport, and at processing in specialty market poultry. *Poultry science, 85*(1), 136-143.

Micallef, S. A., & Kumar, G. D. (2017). Biofilms: a community-based strategy for bacterial persistence and relevance to food safety. In *Trends in Food Safety and Protection* (pp. 107-130). CRC Press.

Norman, N. N., & Kabler, P. W. (1953). Bacteriological study of irrigated vegetables. *Sewage and Industrial Wastes*, 605-609.

Oni, R. A., Sharma, M., & Buchanan, R. L. (2015). Survival of *Salmonella enterica* in dried turkey manure and persistence on spinach leaves. *Journal of Food Protection, 78*(10), 1791-1799.

Pan, D., & Yu, Z. (2014). Intestinal microbiome of poultry and its interaction with host and diet. *Gut microbes, 5*(1), 108-119.

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.

Pangloli, P., Dje, Y., Ahmed, O., Doane, C., Oliver, S., & Draughon, F. (2008). Seasonal incidence and molecular characterization of *Salmonella* from dairy cows, calves, and farm environment. *Foodborne Pathogens and Disease, 5*(1), 87-96.

Park, S., Navratil, S., Gregory, A., Bauer, A., Srinath, I., Szonyi, B., Nightingale, K.,Anciso, J., Jun, M., & Han, D. (2014). Farm management, environment and weather factors jointly affect the probability of spinach contamination with generic *Escherichia coli* at the preharvest level. *Applied and Environmental Microbiology.* 03643-03613.

Renwick, S. A., Irwin, R. J., Clarke, R. C., McNab, W. B., Poppe, C., & McEwen, S. A. (1992). Epidemiological associations between characteristics of registered broiler chicken flocks in Canada and the *Salmonella* culture status of floor litter and drinking water. *The Canadian Veterinary Journal, 33*(7), 449.

Rothrock, M. J., Hiett, K. L., Guard, J. Y., & Jackson, C. R. (2016). Antibiotic resistance patterns of major zoonotic pathogens from all-natural, antibiotic-free, pasture-raised broiler flocks in the southeastern United States. *Journal of Environmental Quality, 45*(2), 593-603.

Siemon, C. E., Bahnson, P. B., & Gebreyes, W. A. (2007). Comparative investigation of prevalence and antimicrobial resistance of *Salmonella* between pasture and conventionally reared poultry. *Avian Diseases, 51*(1), 112-117.

Smith, D., Northcutt, J., Cason, J., Hinton Jr, A., Buhr, R., & Ingram, K. (2007). Effect of external or internal fecal contamination on numbers of bacteria on prechilled broiler carcasses. *Poultry Science, 86*(6), 1241-1244.

Sossidou, E., Dal Bosco, A., Elson, H., & Fontes, C. (2011). Pasture-based systems for poultry production: implications and perspectives. *World's Poultry Science Journal, 67*(1), 47-58.

Strawn, L. K., Fortes, E. D., Bihn, E. A., Nightingale, K. K., Gröhn, Y. T., Worobo, R. W., Wiedmann, M., & 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.

Thorns, C. (2000). Bacterial food-borne zoonoses. *Revue Scientifique et Technique-Office International des Epizooties, 19*(1), 226-239.

Trimble, L. M., Alali, W. Q., Gibson, K. E., Ricke, S. C., Crandall, P., Jaroni, D., Berrang, M., & Habteselassie, M. Y. (2013). Prevalence and concentration of *Salmonella* and *Campylobacter* in the processing environment of small-scale pastured broiler farms. *Poultry Science, 92*(11), 3060-3066.

USDA. (2015). Meat and Poultry Labeling Terms. Retrieved March 1, 2019, from https://www.fsis.usda.gov/wps/portal/fsis/topics/food-safety-education/get-answers/food-safety-fact-sheets/food-labeling/meat-and-poultry-labeling-terms/meat-and-poultry-labeling-terms.

Wales, A., Breslin, M., Carter, B., Sayers, R., & Davies, R. (2007). A longitudinal study of environmental *Salmonella* contamination in caged and free-range layer flocks. *Avian Pathology, 36*(3), 187-197.

Winfield, M. D., & Groisman, E. A. (2003). Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli. Applied and Environmental Microbiolology, 69*(7), 3687-3694.

Yiridoe, E. K., Bonti-Ankomah, S., & Martin, R. C. (2005). Comparison of consumer perceptions and preference toward organic versus conventionally produced foods: A review and update of the literature. *Renewable Agriculture and Food Systems, 20*(4), 193-205.

Zdragas, A., Mazaraki, K., Vafeas, G., Giantzi, V., Papadopoulos, T., & Ekateriniadou, L. (2012). Prevalence, seasonal occurrence and antimicrobial resistance of *Salmonella* in poultry retail products in Greece. *Letters in Applied Microbiology*, *55*(4), 308-313.

Zhao, Q., & Hastie, T. (2017). Causal interpretations of black-box models. *Journal of Business & Economic Statistics,* Retrieved April 1, 2019, from http://www-stat.wharton.upenn.edu/~qyzhao/papers/pdp.pdf.

Table 4.1. The predicting variables considered in Random forest model

Variable	Description	Levels
AvgNumBirds	Average number of birds on farm in a year	Numeric
AvgNumFlocks	Average number of flocks on farm in a year	Numeric
YearsFarming	Number of years of operation of the farm	Numeric in years
EggSource	Source of broiler eggs	6 levels: Companies A, B, C, D, E, F
BroodBedding	Type of bedding used for brooding	3 levels: pasture-based brooder (PB), wood shavings (WS), saw-dust/shredded paper (SDSP)
BroodFeed	Top 3 protein source for brooding feed	9 ingredients: barley (B), corn (C), oats (O), wheat (W), soy (S), cotton seed mill (M), sorghum (H), peas (P), Grain Products (GP)
BrGMOFree	Was the brood feed GMO free?	2 levels: yes (Y), no (N)
BrSoyFree	Was the brood feed soy free?	2 levels: yes (Y), no (N)
BrMedicated	Was the brood feed medicated?	2 levels: yes (Y), no (N)
BroodCleanFrequency	Brooding area cleaning frequency	7 levels: 3Days, all in/all out (AIAO), daily, deep litter method (DLM), mobile, weekly, yearly
AvgAgeToPasture	Average age broilers were put on pasture	Numeric in weeks

PastureHousing	Type of pasture housing	4 levels: chicken tractor (CT), chicken tractor with fencing (CTF), chicken tractor free range (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 new area?	2 levels: yes (Y), no (N)
PastureFeed	Top 3 protein source for pasture feed	9 ingredients: barley (B), corn (C), oats (O), wheat (W), soy (S), cotton seed mill (M), peas (P), Grain Products (GP), sorghum (H)
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	Was the pasture feed medicated?	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)
SwineOnFarm	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 the grow-out	3 levels: public, rain, well

FreqBirdHandling	Frequency of birds were handled	2 levels: daily, only if needed (OIN)
AnyABXUse	Any antibiotics used for broilers	2 levels: yes (Y), no (N)
LengthFeedRestrixProcess	Length of feed restriction before processing	Numeric in hours
seasons	season of sample collection	4 levels: spring (1), summer (2), fall (3), winter (4)
FlockAgeDays	Age of flock at sample collection	Numeric in days
Breed	Breed of broilers	2 levels: Freedom Ranger (FR), Cornish Cross (CC)
FlockSize	Number of birds in the sample flock	Numeric
ProcessingType[ab]	Location the broilers were processed	2 levels: farm, plant
SkinOnOff[ab]	Skin-on or off	2 levels: on, off
ScalderTempC[b]	Temperature of water (°C) used during scalding of birds during processing	7 levels: 55, 60, 63, 65, 71, 82, none
RinseWaterSource[b]	Source of water used for carcass rinsing	2 levels: public, well
RinseWaterChlor[b]	Was rinse water chlorinated?	2 levels: yes (Y), no (N)
ChillingMethod[ab]	Chilling method used	2 levels: water, air
TransportTime[b]	Length of transport tome broilers to process	Numeric in hours
StorageTempC[b]	Carcass storage temperature before reception	Numeric in °C
StorageTimeD[b]	Storage time before reception	Numeric in days

[a]Variables used for WCR-p
[b]Variables used for WCR-f

Table 4.2. Storage temperatures and average storage times in days for each farm

Farm	Storage temperature (°C)	Average storage time (Days)
A	4	1
B*	4	2
B*	-20	2
C	-20	11
D	-20	13
E	4	0.2
F	4	1
G	-20	30.3
H	-20	30.5
I	-20	8.75
J	4	1
K	4	0

*Farm B used different carcass storage methods

Table 4.3. Predictive performances of random forest models and the confusion matrix of the models

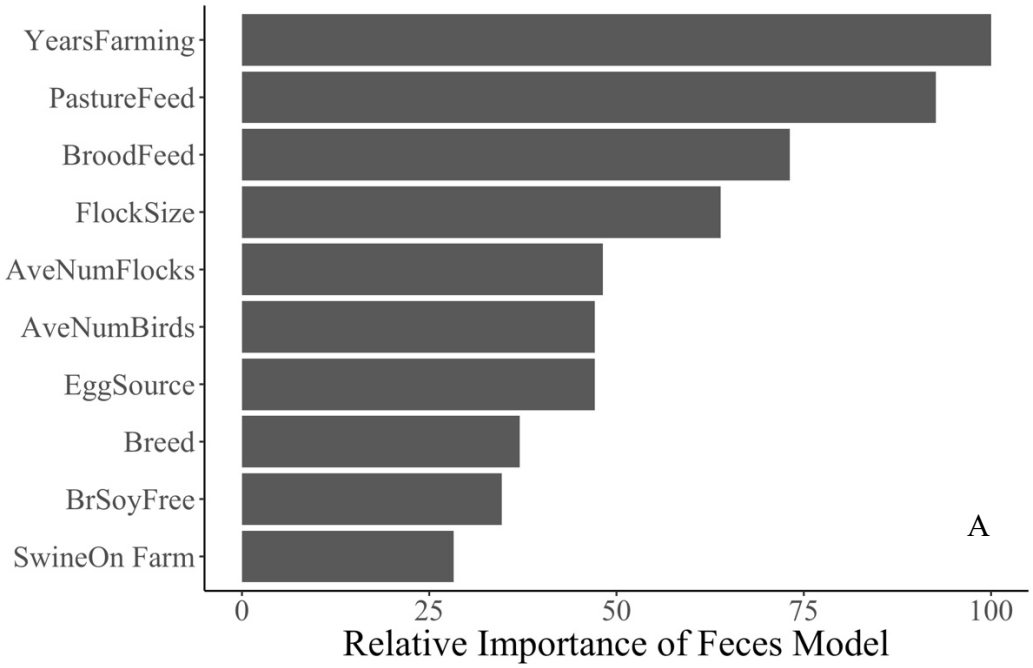
Models		Actual	Actual	Sensitivity	Specificity	AUC*
		+	-			
	Predicted					
Feces	+	24	17	0.89	0.84	0.93
	-	3	91			
Soil	+	15	26	0.94	0.76	0.92
	-	1	84			
WCR-p	+	12	7	0.86	0.74	0.87
	-	2	20			
WCR-f	+	6	8	0.86	0.77	0.83
	-	1	27			

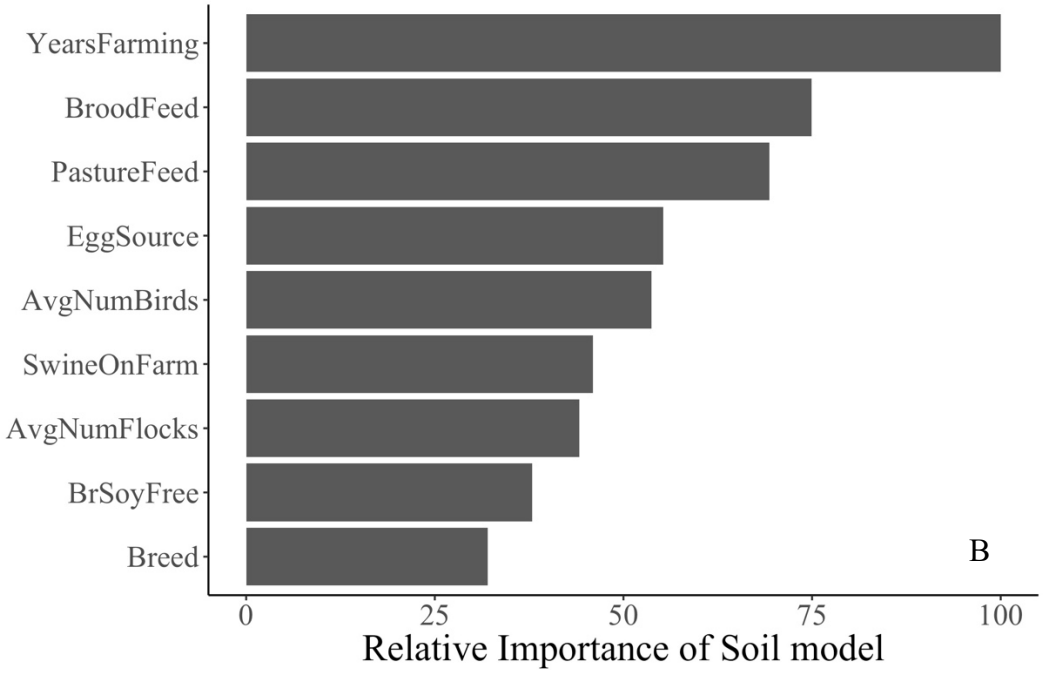
WCR-p is the whole carcass rinse, post processing samples
WCR-f is the whole carcass rinse, final product samples
*AUC is the area under receiver operating characteristic curve
+: *Salmonella* isolated
-: *Salmonella* not isolated

Table 4.4 Seasonal effect on *Salmonella* prevalence in different sample types in the pastured poultry farms

Season	Month	Positive/Total (%) samples for *Salmonella*			
		Soil sample	Fecal sample	WCR-p sample	WCR-f sample
Spring	March-May	20/186 (11%)	27/193 (14%)	5/35 (14%)	0/15 (0%)
Summer	June-August	53/381 (19%)	70/378 (19%)	38/130 (29%)	15/125 (12%)
Fall	September-November	16/145 (11%)	21/145 (14%)	22/60 (37%)	17/69 (25%)
Winter	December-February	0/10 (0%)	1/10 (10%)	2/10 (20%)	8/20 (40%)

WCR-p is the whole carcass rinse, post processing samples
WCR-f is the whole carcass rinse, final product samples


YearsFarming
PastureFeed
BroodFeed
FlockSize
AveNumFlocks
AveNumBirds
EggSource
Breed
BrSoyFree
SwineOn Farm
A
0
25
50
75
100
Relative Importance of Feces Model


YearsFarming
BroodFeed
PastureFeed
EggSource
AvgNumBirds
SwineOnFarm
AvgNumFlocks
BrSoyFree
Breed
B
0
25
50
75
100
Relative Importance of Soil model

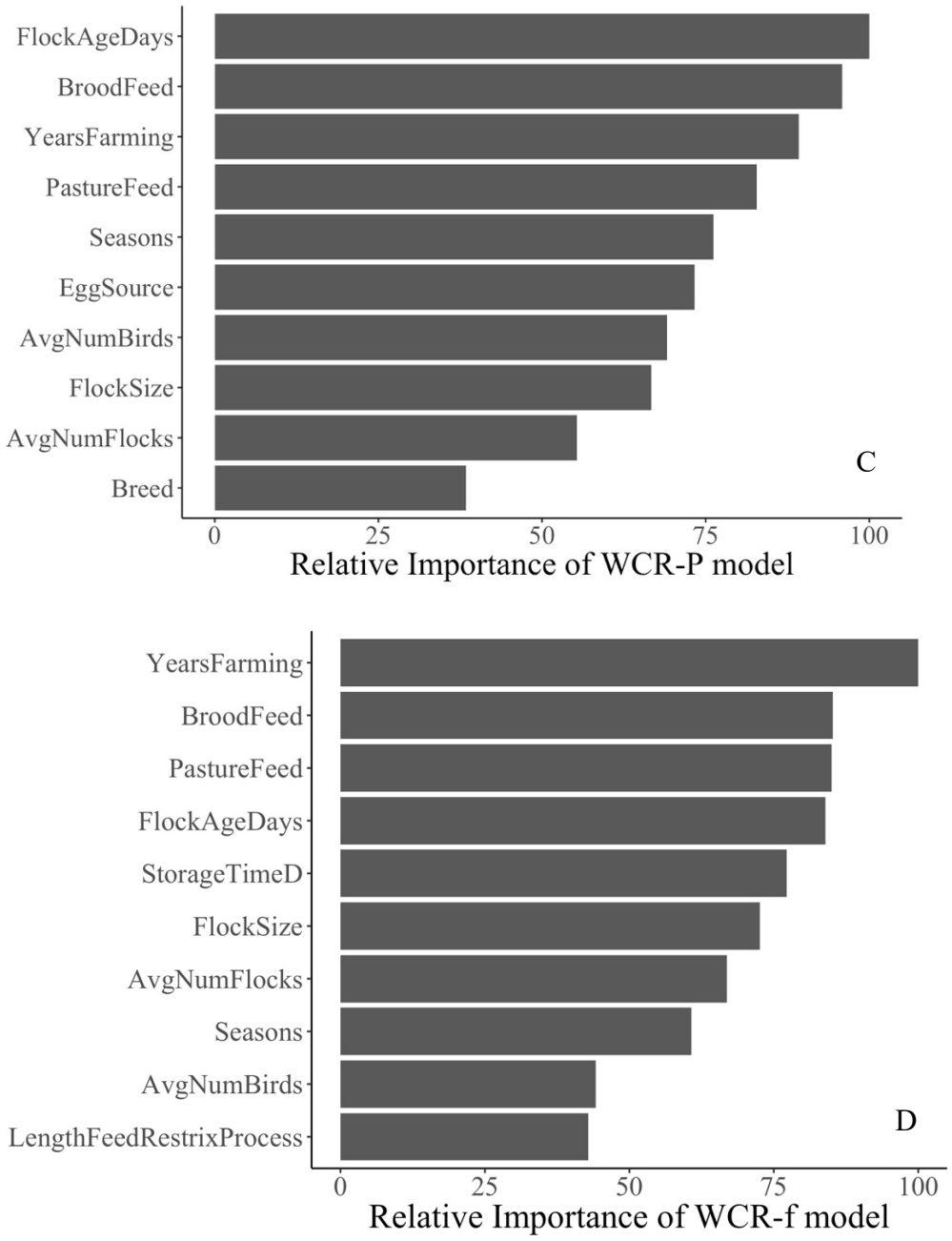


Figure 4.1. Relative importance plots for feces (A), soil (B), WCR-p (C), and WCR-f (D) samples. Relative importance plot ranks the predicting variables in importance score order.

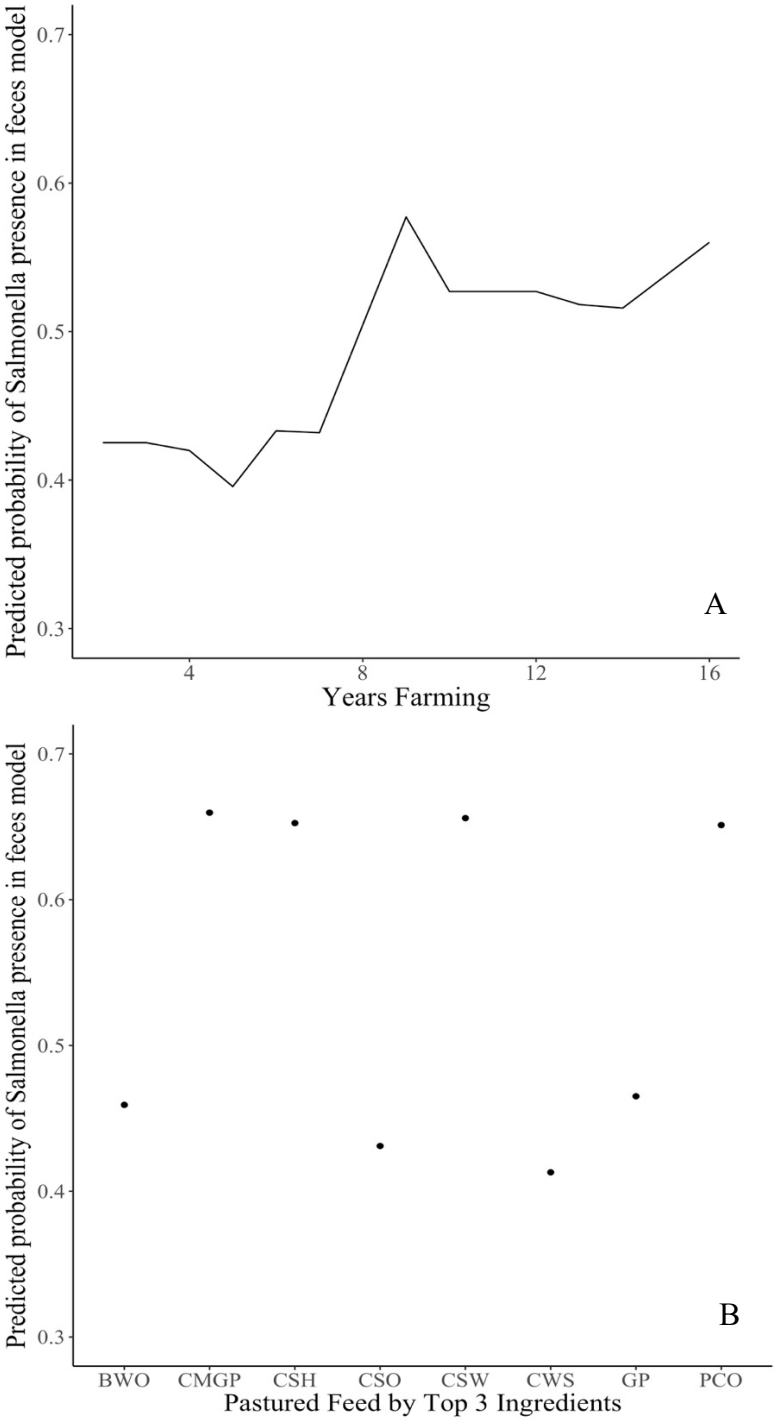


Figure 4.2. Partial dependency plots for the top two variables for feces model. The number of years farming was the most important predicting variable (A), and the Pasture feed (B) was the second important variable, and the top 3 ingredients are listed:
B (Barly), C (corn), GP (grain product), S (soy), H (sorghum), M (cotton seed mill), P (pea), O (oats) W (wheat).

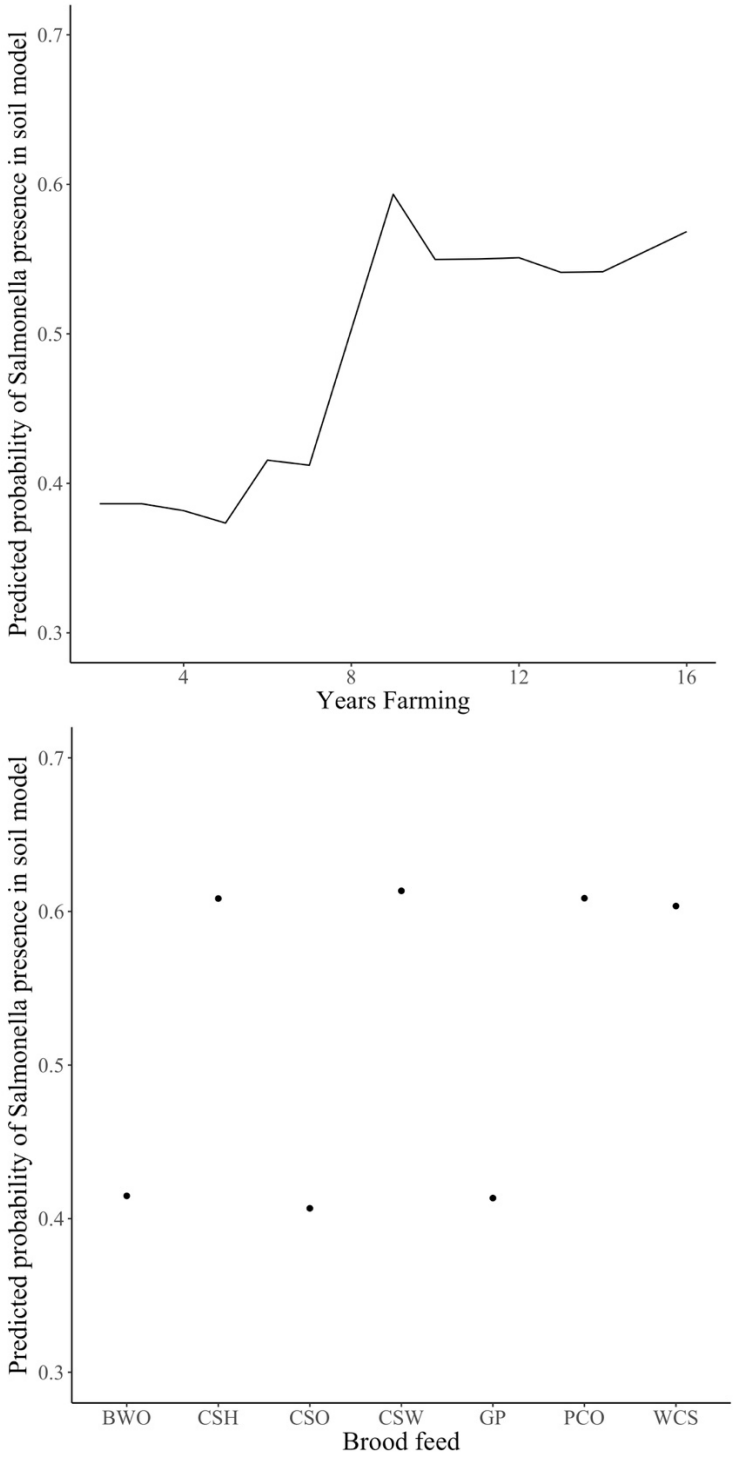


Figure 4.3. Partial dependency plots for the top 2 variables for soil model. Years farming (A) was most important variable identified. Brood feed (B) was the second important variable, and the top 3 ingredients are listed:
B (Barley), C (corn), GP (grain product), S (soy), H (sorghum), M (cotton seed mill), P (pea), O (oats) W (wheat).

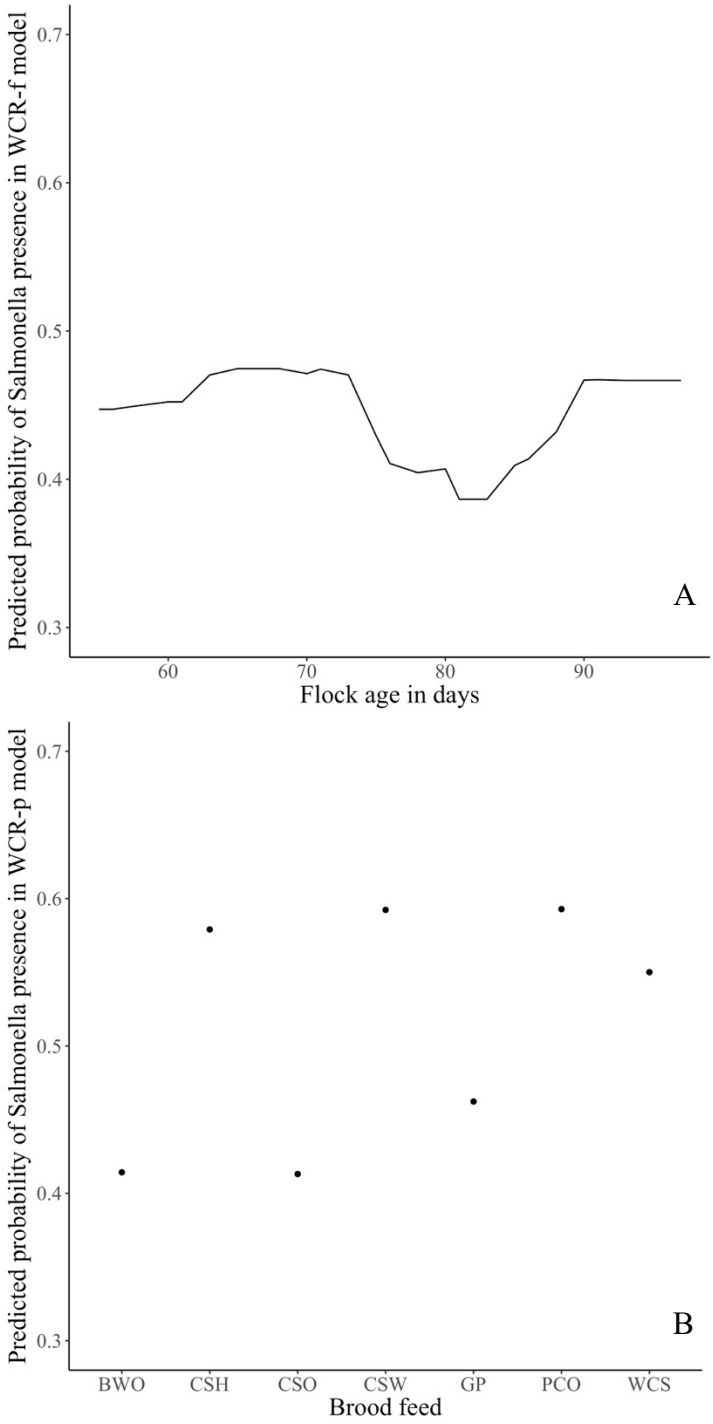


Figure 4.5. Partial dependency plots for the top 2 variables for WCR-p model. Flock age in days (A) was identified as the most important variable. Brood feed (B) was the second important variable, and the top 3 ingredients are listed:
B (Barley), C (corn), GP (grain product), S (soy), H (sorghum), M (cotton seed mill), P (pea), O (oats) W (wheat).

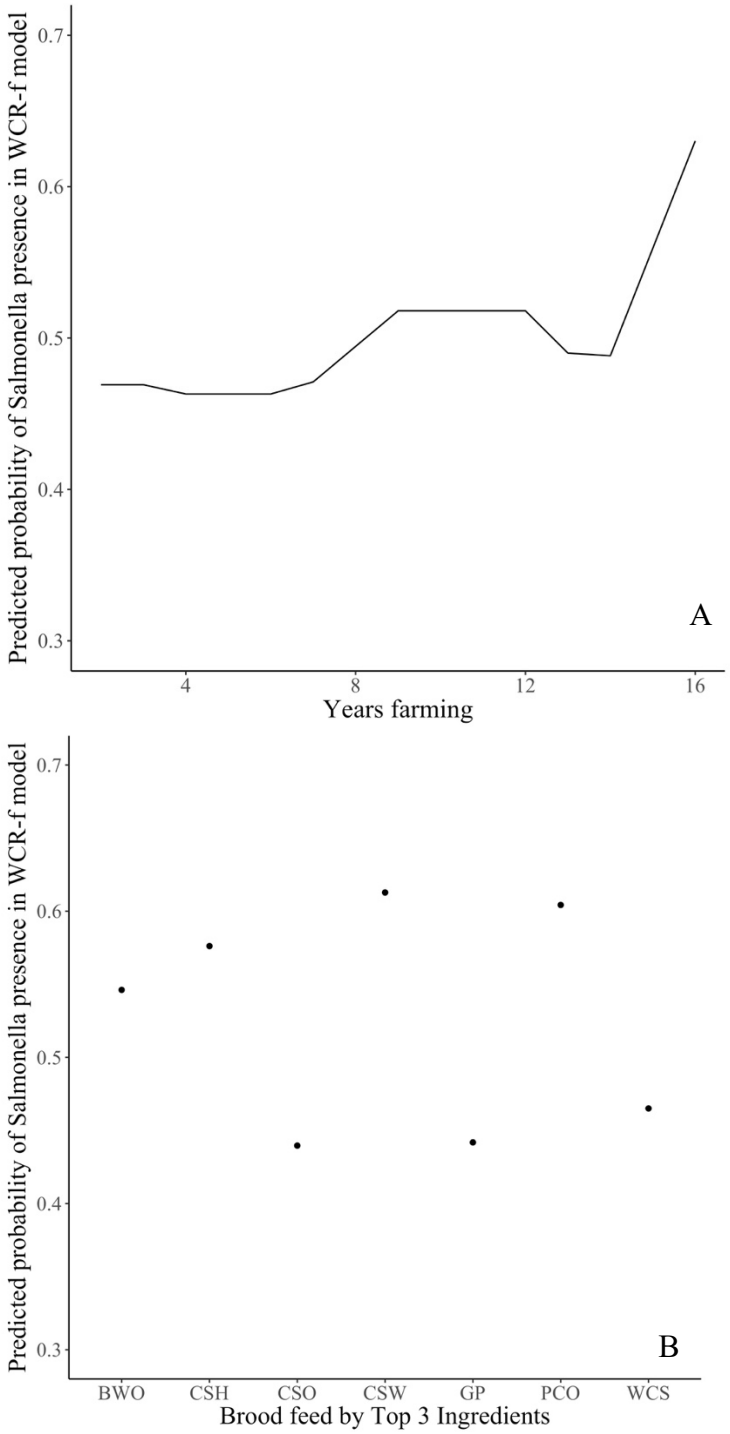
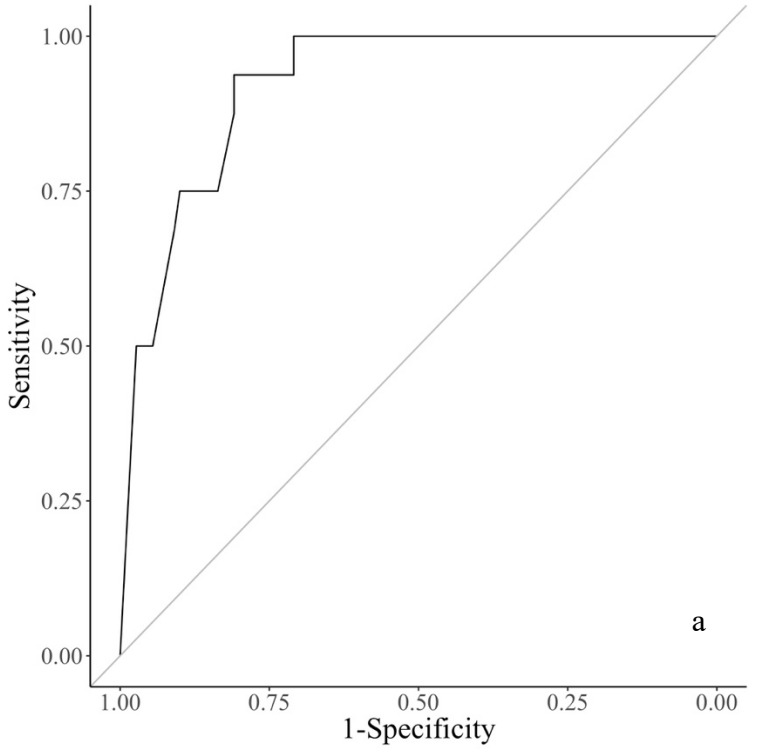
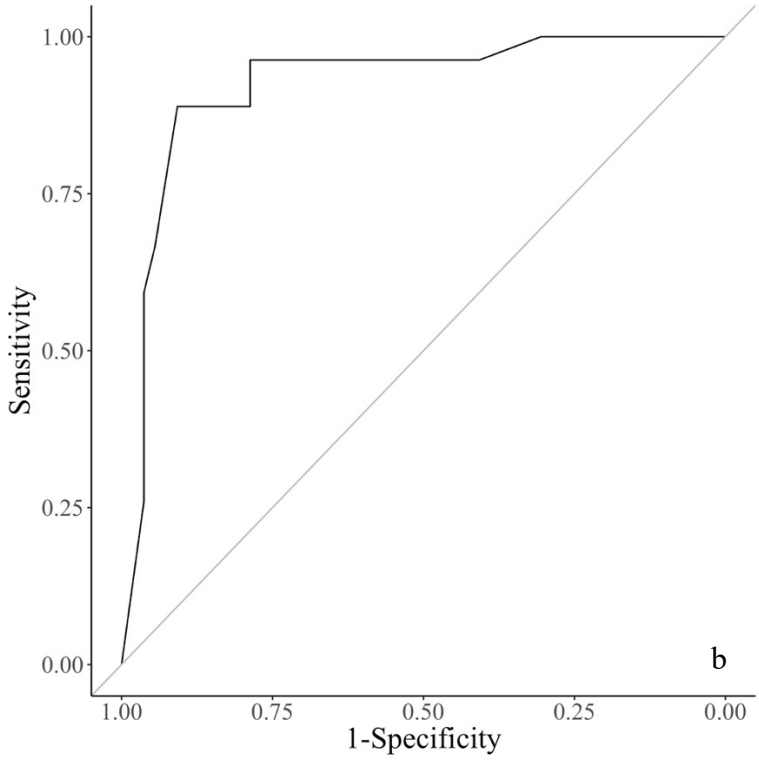


Figure 4.6. Partial dependency plots for top 2 variables for WCR-f model. Years farming (A) was most important variable identified. Brood feed (B) was the second important variable, and the top 3 ingredients are listed:
B (Barley), C (corn), GP (grain product), S (soy), H (sorghum), M (cotton seed mill), P (pea), O (oats) W (wheat).


1.00
0.75
0.50
0.25
0.00
Sensitivity
1.00
0.75
0.50
0.25
0.00
1-Specificity
a


1.00
0.75
0.50
0.25
0.00
Sensitivity
1.00
0.75
0.50
0.25
0.00
1-Specificity
b

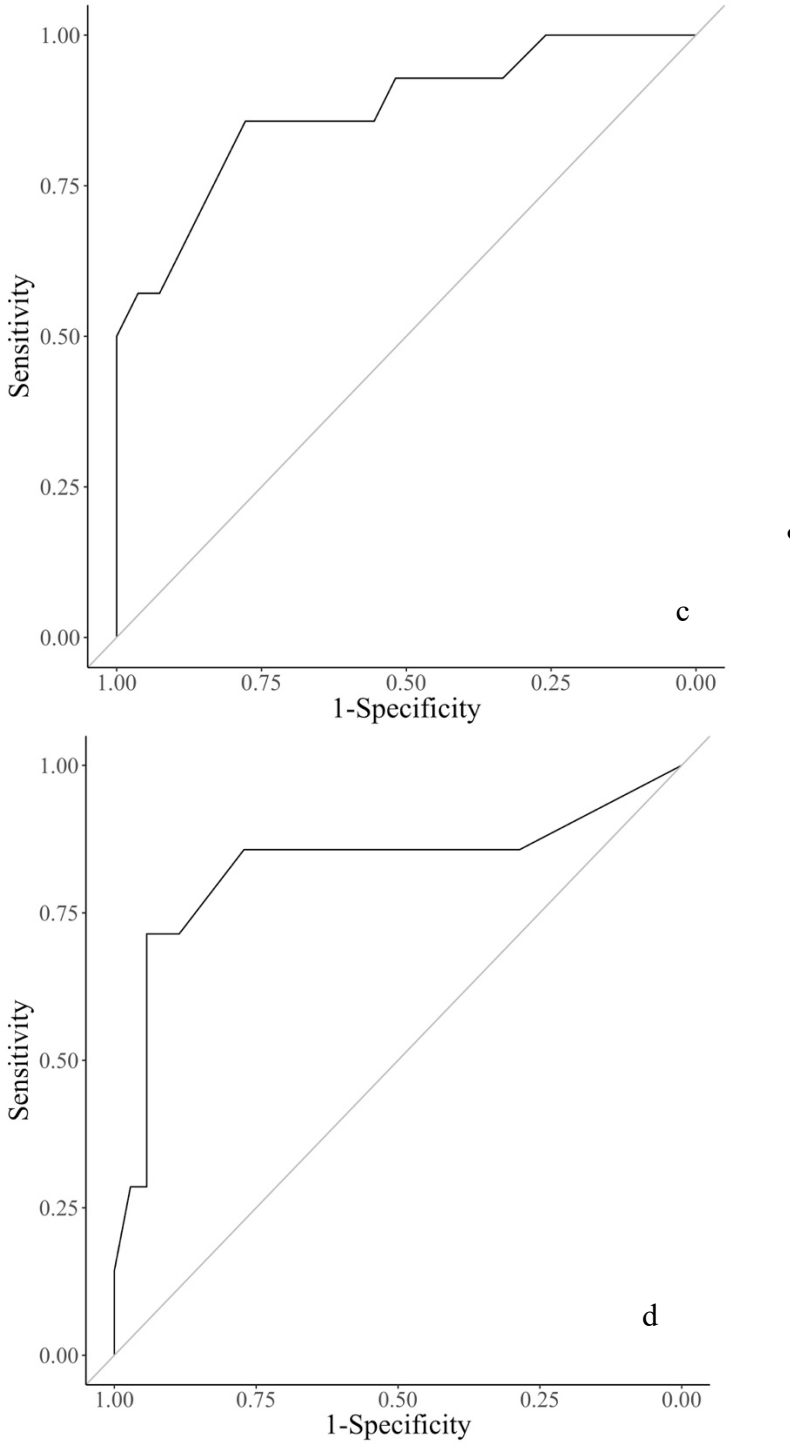


Figure 4.7. Receiver operating characteristic (ROC) curves for random forest models: (a) feces model (b) soil model (c) WCR-p model, and (d) WCR-f model.

CHAPTER 5

CONCLUSIONS

Salmonella presence in pastured poultry farms is a serious food safety concern. Meteorological factors and farm management factors affect the prevalence of *Salmonella* due to the unique setting of pastured poultry farms. The random forest method was used to generate the models in this study. The important factors that contributed to the presence of *Salmonella* include average maximum windspeed, average temperature, average humidity, pasture and brood feed, years farming, and flock age. The models created can provide useful information in goals to reduce the *Salmonella* prevalence in pastured poultry farms.

Random forest models and other machine learning techniques can be applied to numerous food safety situations. Modeling the presence of other pathogens such as *Campylobacter* could provide further understanding food safety risks in pastured poultry farms. Another future study could include the comparison of risks among conventional, organic, and pastured poultry farms.

APPENDICES

Appendix A

R programs for random forest analysis

The codes provided are the R script for the soil sample. The analyses were conducted in the same manner in all other sample types including feces for meteorological data, and feces, soil, WCR-p and WCR-f for farm practices data.

```
meteo.soil<- read.csv("meteo.soil.csv")
library(randomForest)
set.seed(111)
id<-sample(2,nrow(meteo.soil), prob = c(0.7,0.3), replace=TRUE)
meteo.train.soil <- meteo.soil[id==1,]
meteo.test.soil <- meteo.soil[id==2,]
is.salmonella <- meteo.soil$salmonella....
summary(is.salmonella)
str(meteo.train.soil)
bestmtry <- tuneRF(meteo.train.soil, meteo.train.soil$salmonella...., stepFactor = 1.2,
improve = 0.01,
                trace = TRUE, plot = TRUE)
meteo.forest.soil <- randomForest(meteo.train.soil$salmonella....~.,
                                  mtry=10, data = meteo.train.soil)
meteo.forest.soil
```

```
meteo.forest.soil$importance
varImpPlot(meteo.forest.soil, sort=T, n.var = 10, main="Top 10 Variable Importance")
predict.meteo2 <- predict(meteo.forest.soil, newdata= meteo.test.soil, type= "class")
predict.meteo2.prob <- predict(meteo.forest.soil, newdata= meteo.test.soil, type= "prob")
head(predict.meteo2)
head(predict.meteo2.prob)
model.soil2<- randomForest(meteo.train.soil$salmonella.... ~ ., data=meteo.train.soil,
                ntree=1000, mtry=10, importance=T)
plot(model.soil2)
library(caret)
set.seed(123)
p3<- predict(model.soil2, meteo.test.soil)
confusionMatrix(p3,meteo.test.soil$salmonella....)
library(pROC)
predict.with.prob2<- predict(model2, meteo.test.soil, type= 'prob')
auc<- auc(meteo.test.soil$salmonella....,predict.with.prob2[,2])
auc
plot(roc(meteo.test.soil$salmonella....,predict.with.prob2[,2]))
partialPlot(meteo.forest,meteo.train, AVEWS6)
partialPlot(meteo.forest,meteo.train, AVEWS0)
library(ROSE)
set.seed(134)
over.soil<- ovun.sample(salmonella....~., data= meteo.train.soil, method="over",
```

```
                N=886)$data
table(over.soil$salmonella....)
summary(over)
rfover.soil<-randomForest(salmonella....~., data=over.soil)
confusionMatrix(predict(rfover.soil,meteo.test.soil),
                meteo.test.soil$salmonella....,
                positive = '+')
library(pROC)
set.seed(133)
predict.over.soil<- predict(rfover.soil, meteo.test.soil, type= 'prob')
auc.soil<- auc(meteo.test.soil$salmonella....,predict.over.soil[,2])
auc.soil
```

Appendix B

Predictor variables used in development of meteorological random forest models

Variable	Description (unit)
MaxWindSpeed1	Average maximum wind speed between the day of sample collection and 1 day prior (m/s)
MaxWindSpeed2	Average maximum wind speed between the day of sample collection and 2 days prior (m/s)
MaxWindSpeed3	Average maximum wind speed between the day of sample collection and 3 days prior (m/s)
MaxWindSpeed4	Average maximum wind speed between the day of sample collection and 4 days prior (m/s)
MaxWindSpeed5	Average maximum wind speed between the day of sample collection and 5 days prior (m/s)
MaxWindSpeed6	Average maximum wind speed between the day of sample collection and 6 days prior (m/s)
MaxWindSpeed7	Average maximum wind speed between the day of sample collection and 7 days prior (m/s)
MinWindSpeed1	Average minimum wind speed between the day of sample collection and 1 day prior (m/s)
MinWindSpeed2	Average minimum wind speed between the day of sample collection and 2 days prior (m/s)
MinWindSpeed3	Average minimum wind speed between the day of sample collection and 3 days prior (m/s)
MinWindSpeed4	Average minimum wind speed between the day of sample collection and 4 days prior (m/s)
MinWindSpeed5	Average minimum wind speed between the day of sample collection and 5 days prior (m/s)
MinWindSpeed6	Average minimum wind speed between the day of sample collection and 6 days prior (m/s)
MinWindSpeed7	Average minimum wind speed between the day of sample collection and 7 days prior (m/s)

AverageWindSpeed1	Average wind speed between the day of sample collection and 1 days prior (m/s)
AverageWindSpeed2	Average wind speed between the day of sample collection and 2 days prior (m/s)
AverageWindSpeed3	Average wind speed between the day of sample collection and 3 days prior (m/s)
AverageWindSpeed4	Average wind speed between the day of sample collection and 4 days prior (m/s)
AverageWindSpeed5	Average wind speed between the day of sample collection and 5 days prior (m/s)
AverageWindSpeed6	Average wind speed between the day of sample collection and 6 days prior (m/s)
AverageWindSpeed7	Average wind speed between the day of sample collection and 7 days prior (m/s)
MaxGustSpeed3	Average maximum gust speed between the day of sample collection and 3 days prior (m/s)
MaxGustSpeed4	Average maximum gust speed between the day of sample collection and 4 days prior (m/s)
MaxGustSpeed5	Average maximum gust speed between the day of sample collection and 5 days prior (m/s)
MaxGustSpeed6	Average maximum gust speed between the day of sample collection and 6 days prior (m/s)
MaxGustSpeed7	Average maximum gust speed between the day of sample collection and 7 days prior (m/s)
MinHumidity1	Average minimum humidity between the day of sample collection and 1 day prior (%)
MinHumidity2	Average minimum humidity between the day of sample collection and 2 days prior (%)
MinHumidity3	Average minimum humidity between the day of sample collection and 3 days prior (%)
MinHumidity4	Average minimum humidity between the day of sample collection and 4 days prior (%)
MinHumidity5	Average minimum humidity between the day of sample collection and 5 days prior (%)
MinHumidity6	Average minimum humidity between the day of sample collection and 6 days prior (%)

MinHumidity7	Average minimum humidity between the day of sample collection and 7 days prior (%)
MaxHumidity1	Average maximum humidity between the day of sample collection and 1 day prior (%)
MaxHumidity2	Average maximum humidity between the day of sample collection and 2 days prior (%)
MaxHumidity3	Average maximum humidity between the day of sample collection and 3 days prior (%)
MaxHumidity4	Average maximum humidity between the day of sample collection and 4 days prior (%)
MaxHumidity5	Average maximum humidity between the day of sample collection and 5 days prior (%)
MaxHumidity6	Average maximum humidity between the day of sample collection and 6 days prior (%)
MaxHumidity7	Average maximum humidity between the day of sample collection and 7 days prior (%)
AverageHumidity1	Average humidity between the day of sample collection and 1 day prior (%)
AverageHumidity2	Average humidity between the day of sample collection and 2 days prior (%)
AverageHumidity3	Average humidity between the day of sample collection and 3 days prior (%)
AverageHumidity4	Average humidity between the day of sample collection and 4 days prior (%)
AverageHumidity5	Average humidity between the day of sample collection and 5 days prior (%)
AverageHumidity6	Average humidity between the day of sample collection and 6 days prior (%)
AverageHumidity7	Average humidity between the day of sample collection and 7 days prior (%)
MinTemperature1	Average minimum temperature between the day of sample collection and 1 day prior (°C)
MinTemperature2	Average minimum temperature between the day of sample collection and 2 days prior (°C)
MinTemperature3	Average minimum temperature between the day of sample collection and 3 days prior (°C)

MinTemperature4	Average minimum temperature between the day of sample collection and 4 days prior (°C)
MinTemperature5	Average minimum temperature between the day of sample collection and 5 days prior (°C)
MinTemperature6	Average minimum temperature between the day of sample collection and 6 days prior (°C)
MinTemperature7	Average minimum temperature between the day of sample collection and 7 days prior (°C)
MaxTemperature1	Average maximum temperature between the day of sample collection and 1 day prior (°C)
MaxTemperature2	Average maximum temperature between the day of sample collection and 2 days prior (°C)
MaxTemperature3	Average maximum temperature between the day of sample collection and 3 days prior (°C)
MaxTemperature4	Average maximum temperature between the day of sample collection and 4 days prior (°C)
MaxTemperature5	Average maximum temperature between the day of sample collection and 5 days prior (°C)
MaxTemperature6	Average maximum temperature between the day of sample collection and 6 days prior (°C)
MaxTemperature7	Average maximum temperature between the day of sample collection and 7 days prior (°C)
AverageTemperature1	Average temperature between the day of sample collection and 1 day prior (°C)
AverageTemperature2	Average temperature between the day of sample collection and 2 days prior (°C)
AverageTemperature3	Average temperature between the day of sample collection and 3 days prior (°C)
AverageTemperature4	Average temperature between the day of sample collection and 4 days prior (°C)
AverageTemperature5	Average temperature between the day of sample collection and 5 days prior (°C)
AverageTemperature6	Average temperature between the day of sample collection and 6 days prior (°C)
AverageTemperature7	Average temperature between the day of sample collection and 7 days prior (°C)

Precipitation1	Average amount of precipitation between the day of sample collection and 1 day prior (mm)
Precipitation2	Average amount of precipitation between the day of sample collection and 2 days prior (mm)
Precipitation3	Average amount of precipitation between the day of sample collection and 3 days prior (mm)
Precipitation4	Average amount of precipitation between the day of sample collection and 4 days prior (mm)
Precipitation5	Average amount of precipitation between the day of sample collection and 5 days prior (mm)
Precipitation6	Average amount of precipitation between the day of sample collection and 6 days prior (mm)
Precipitation7	Average amount of precipitation between the day of sample collection and 7 days prior (mm)
Seasons	Season of the sampling date, 4 levels: spring, summer, fall, and winter