

ABSTRACT

Title of Dissertation:

EVALUATION AND VALIDATION OF
VARIOUS SAMPLING PLANS FOR THE
DETECTION OF PATHOGENIC OR
INDICATOR MICROORGANISMS ON PRE-
HARVEST LEAFY GREENS

Aixia Xu, Doctor of Philosophy, 2017

Dissertation directed by:

Robert L. Buchanan, Ph.D., Director and
Professor,
Department of Nutrition and Food Science,
Center for Food Safety and Security Systems

The consumption of leafy greens increased over the last few decades due to health concerns. However, leafy green vegetables are highly susceptible to microbial contamination. The pre-harvest sampling and testing are highly important to ensure safety of leafy greens. Z-pattern sampling scheme is extensively used currently. However, the scientific rationale and performance attributes of these sampling plans are unclear in relation to detection of both indicator microorganisms and pathogens. The overall goal of this study is to evaluate and validate various sampling plans for the detection of pathogenic bacteria on pre-harvest leafy greens and find the optimal sampling plan. Computer simulations and field trials were performed to compare the effectiveness of various sampling plans, including simple random sampling, stratified random sampling,

Z-pattern sampling, “samples of opportunity” sampling and iterative Bayesian sampling. Studies showed that Z-pattern sampling plan had larger variability than random sampling plan and stratified sampling plan when the contamination sites were randomly distributed, although the mean detection probabilities of these three sampling plans were the same. Samples of opportunity sampling performed better than random sampling plan and stratified sampling plan when the contamination sites were non-randomly distributed, such as flooded field, field with animal house nearby or field with power line above. And iterative Bayesian sampling was suggested when the number of samples is limited. A what-if sampling strategy would be made to get a more efficient detection of pathogenic bacteria for the industry and government farms. This study provides the scientific and mathematical rationale for various sampling plans and allows leafy green growers to make informed decisions regarding strategies for optimizing pre-harvest microbiological testing programs.

EVALUATION AND VALIDATION OF VARIOUS SAMPLING PLANS FOR
THE DETECTION OF PATHOGENIC OR INDICATOR MICROORGANISMS ON
PRE-HARVEST LEAFY GREENS

By

Aixia Xu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory committee:
Professor Robert L. Buchanan, Chair
Professor Patricia Millner
Professor Donald W. Schaffner
Professor Adel Shirmohammadi
Associate Professor Shirley A. Micallef
Associate Professor Abani K. Pradhan
Assistant Professor Rohan V. Tikekar

© Copyright by
Aixia Xu
2017

Acknowledgements

I would like to take the opportunity to expression my appreciation to the people who made this dissertation possible and made my graduate experience complete, meaningful and unforgettable.

My deepest gratitude is to my advisor, Dr. Robert Buchanan for his guidance, patience and understanding in the last five years. I appreciate him for taking the time and effort to guide me into the world of food science. He always helps me deal with any crisis situations with his patience and his experience. His enthusiasm for the work and positive attitude toward life influenced me gradually. I hope I would be still such enthusiastic and energetic like him when I was at his age.

I would also like to thank Dr. Patricia Millner from U.S. Department of Agriculture. She gave me a lot of precious suggestions on the field trials and made my field studies more efficient and meaningful. My sincere thanks also go to the members of my dissertation committee: Dr. Donald W. Schaffner, Dr. Adel Shirmohammadi, Dr. Shirley A. Micallef, Dr. Abani Pradhan, and Dr. Rohan V. Tikekar, whose constructive guidance was a key factor in the elaboration of this dissertation.

I am grateful to all the help from members of Dr. Buchanan's lab, Dr. Pradhan's lab, Dr. Tikekar's lab and Dr. Meng's lab: Dr. Yangyang Wang, Lucy Ruan, Dr. Abhinav Mishra, Dr. Hao Pang, Yinzhi Qu, Qiao Ding, Qingyang Wang and Xun Yang. Many thanks also go to Dr. Manan Sharma and Kate White from the U.S. Department of Agriculture, Mike Dwyer from Central Maryland Research &

Education Center (CMREC) and the faculty and staffs in the Department of Nutrition and Food Sciences of the University of Maryland. These people took part in my Ph.D. training and helped me to make my work move ahead smoothly. I am very lucky that I could work with them for all these years.

My many friends in the department have helped me throughout my studies. We support each other and share each other. I will cherish our friendship forever.

Most importantly, I would like to thank to my family for their love, support, encouragement and understanding. I would not be here without them. I would like to thank to my mother Cuiying Shen and my father Yuping Xu. Thanks to my wonderful husband Yuwei Cui, for his love, support and understanding.

Table of Contents

Chapter 1: Introduction	1
1.1 Increasing demand for fresh produce	1
1.2 Microbiological risks of fresh leafy greens	1
1.3 Microbiological testing for fresh produce	2
1.4 Introduction to sampling	3
Chapter 2. Literature Review	5
2.1 Hazards associated leafy green at pre-harvest	5
2.1.1 <i>Salmonella</i>	5
2.1.2 Shiga Toxin-producing <i>E. coli</i> (STEC).....	6
2.1.3 <i>Cryptosporidium</i>	7
2.1.4 Norovirus	8
2.1.5 <i>Shigella</i>	9
2.2 Sources of Pre-harvest Contamination	10
2.2.1 Irrigation water	10
2.2.2 Manure, green waste and compost	12
2.2.3 Animal and insect activity.....	14
2.2.4 Human activity.....	15
2.3 Sampling in food safety	15
2.3.1 Importance of food safety sampling	15
2.3.2 Introduction to sampling plans	16
2.3.3 Sampling strategies.....	17
2.3.3.1 Simple random sampling.....	17
2.3.3.2 Systematic sampling	17
2.3.3.3 Stratified random sampling.....	18
2.3.3.4 Z-pattern sampling.....	18
Chapter 3. Use of Simulation Modeling and Field Validation to Evaluate Three Sampling Plans for the Detection of Pathogenic Bacteria on Pre-harvest Leafy Greens	19
3.1 Abstract	19
3.2 Introduction	20
3.3 Materials and Methods	21
3.3.1 Simulation modeling	21
3.3.1.1 Description of Model Fields Used to Compare Sampling Plans.....	22
3.3.1.2 Mathematical Basis for Sampling Plan Efficiency.....	22
3.3.1.2.1 Random Sampling Plans	22
3.3.1.2.2 Stratified Random Sampling Plans	25
3.3.1.2.3 Z-pattern sampling plan.....	26
3.3.1.3 Comparison of Sampling Plans Using Simulation Modeling.....	28
3.3.2 Validation in the field.....	28
3.3.2.1 Inoculation	29
3.3.2.1.1 Preparation of dairy solids extracts.....	29
3.3.2.1.2 Strains and culture conditions used	29
3.3.2.1.3 Inoculum application	30
3.3.2.2 Sampling in the field.....	33
3.3.2.3 Sample processing	34
3.3.2.4 Data analysis.....	35

3.4 Results	35
3.4.1 Simulation	35
3.4.2 Validation	44
3.5 Discussion	49
3.6 Conclusions	53
Chapter 4: Evaluation of the potential for enhanced sampling effectiveness by assessment of field environments and consideration of likely sources of contamination:	54
Comparison of sampling plans based on random, stratified random, Z-pattern and “samples of opportunity” sampling	54
4.1 Abstract	54
4.2 Introduction	55
4.3 Materials and methods	58
4.3.1 Computer simulation modeling.....	58
4.3.1.1 Model of non-random contamination factors	58
4.3.1.2 Generation of simulated contaminated fields.....	61
4.3.1.3 Sampling plans.....	62
4.3.1.4 Evaluation of sampling plans	64
4.3.2 Validation field experiment.....	65
4.3.2.1 Sampling in the field.....	66
4.3.2.2 Sample processing	67
4.3.2.3 Data analysis.....	68
4.4 Results	69
4.4.1 Simulation	69
4.4.2 Validation	75
4.4.2.1 Flooded field.....	75
4.4.2.2 Field with animal house nearby.....	78
4.4.2.3 Field with power line above	82
4.4.2.4 Comparison of sampling plans across three repeated trials	86
4.5 Discussion	87
4.6 Conclusion	89
Chapter 5: <i>In silico</i> Evaluation of a Novel Iterative Bayesian Sampling Strategy for Efficient Detection of Pathogenic Bacteria in Pre-harvest Produce and Environments	91
5.1 Abstract	91
5.2 Introduction	92
5.3 Materials and Methods	94
5.3.1 Iterative sampling with Bayesian Global Optimization.....	94
5.3.2 Model of non-random contamination factors	98
5.3.3 Generation of simulated contaminated fields.	100
5.3.4 Sampling plans.....	101
5.3.5 Evaluation of sampling plans.....	103
5.4 Results	104
5.5 Discussion	111
Chapter 6: Summary and Future Studies	114
6.1 Summary	114
6.2 Future studies	117
References	118

List of Tables

Table 4.1 Summary of model parameters.....	62
Table 4.2 The p-values of ANOVA repeated measures analysis on average bacterial indicators from three fields.....	86
Table 4.3 The p-values of ANOVA repeated measures analysis on fraction of positive samples of five indicators out of 18 samples.....	87
Table 5.1 Summary of model parameters.....	102

List of Figures

- Figure 3.1** Example of samples drawn according to a random sampling plan ($N_{\text{sample}} = 30$). The solid lines represent 30 (5 X 6) plots and the dashed lines represent the subplots (9 subplots/plot) for a total of $9 \times 30 = 270$ subplots. Samples were randomly selected chosen from the 270 subplots.....23
- Figure 3.2** Example of samples drawn according to a stratified random sampling plan ($N_{\text{sample}} = 30$). The solid lines represent 30 (5 X 6) plots and the dashed lines represent the subplots (9 subplots/plot) for a total of $9 \times 30 = 270$ subplots. Each plot has one subplot that was randomly selected for sampling.26
- Figure 3.3** Example samples drawn according to the Z-pattern (see shaded plots) sampling plan ($N_{\text{sample}} = 30$). The solid lines represent 30 (5 X 6) plots and the dashed lines represent the subplots (9 subplots/plot) for a total of $9 \times 30 = 270$ subplots. Each plot within "Z" area has one or two subplots randomly selected for sampling. No samples were selected in the plots outside the Z-pattern.....27
- Figure 3.4** Aerial view of the experiment field (top) and the calculated grid for sampling. The black dots represent calibration points used to define the rough location of the experiment field. The red, green and blue dots represents sampling grid for three experiment fields (270 subplots per each experiment field).32
- Figure 3.5** Example of one subplot identified by strings.33

Figure 3.6 Computer generated sampling locations in the field. Samples for random, stratified random, and z-pattern sampling plan were shown in red, blue, and green respectively.34

Figure 3.7 Mean detection probability (probability of detecting at least one positive sample) as a function of the number of samples collected and the number of contamination sites. The dashed line represents theoretical calculations and the solid lines represent observed mean detection probability from the random sampling plan simulation. Detection probability increases as a function of the number of contamination sites (x-axis) and the number of samples (shown as different symbols). Each data point represents the mean detection probability from 100 simulations. For each simulation, a randomly contaminated field and estimated the detection probability was determined by repeating the random sampling plan 100 times (100 iterations/simulation).....36

Figure 3.8 Mean detection probability as a function of number of contamination sites. The theoretical estimate captures the average detection probability for three different sampling plans ($N_{\text{sample}} = 30$). For other three simulation results, each column represents the detection probability from 100 simulations. The simulation generated a randomly contaminated field and estimated the detection probability by repeating the three sampling plans 100 times (100 iterations/simulation). The error bars represent ± 1 standard deviation37

Figure 3.9 Log-log plot of mean detection probability as a function of total number of subplots per field for the three sampling plans.38

Figure 3.10 Distribution of detection probabilities for the three sampling plans for four combinations of contamination sites ($N_c=2$ or 6) and numbers of samples ($N_{\text{sample}}=10$ or 30). The vertical dashed line represents theoretical estimate of the mean detection probabilities.....41

Figure 3.11 Graphical user interface of a Matlab application that allows comparisons of the three sampling plans. The user can configure the fields and specify the sampling plans on the left. The right side shows displays the distribution of detection probabilities for different sampling plans.42

Figure 3.12 Mean detection probability of three sampling plans as a function of the number of contamination sites (x-axis) with number of samplers (N_{sampler}). The assumption is that each sampler collected samples independently. The detection probabilities were estimated using one simulation with 100 iterations for each sampler. The results for each sampler were combined and detection probability calculated by the number of iterations where at least one positive sample was indicated by any of the samplers.43

Figure 3.13 Number of positive samples and standard deviation of positive samples for different sampling plans in two trials.....45

Figure 3.14 Experimental data from first field sampling based on GPS. (see text for details).....46

Figure 3.15 Experimental data from second field sampling based on post & string (see text for details).....48

Figure 3.16 Distance to the nearest contamination sites for negative and positive samples.49

Figure 4.1 Three types of non-random contamination in 18m (x axes) by 15m (y axes) fields. (A). Stationary line contamination. (B). Stationary point contamination. (C). Directional line contamination. The color represents relative contamination likelihood. The red dots represent contamination sites. The direction of contamination deposition from their sources is highlighted with black lines or arrows. The color depiction represents relative likelihood of contamination of the field from the contamination source.....60

Figure 4.2 Example samples drawn according to different sampling plans. The high risk areas identified by the SOO plan are shown with squares that has gray boundaries. In this case we assume SOO correctly identifies contamination sources.64

Figure 4.3 Example of soil berm built to contain floodwater.66

Figure 4.4 Detection Probability of four sampling plans: random (red), stratified random (green), z-pattern (yellow) and sampling of opportunity plan (blue). Values from bars with different letters are significantly different based on an ANOVA and Tukey’s post hoc test ($p < 0.01$). Error bars represent standard deviation of detection probability across 500 simulated contaminated fields. .70

Figure 4.5 Detection probability of different sampling plans as a function of number of samples. The relative effectiveness of different sampling plans is the same as in Fig. 4.4.71

Figure 4.6 Detection probability as a function of number of contaminations in the field.72

Figure 4.7 Detection probability of the SOO sampling plan while the high risk area cannot be accurately determined. Each colored line represent simulations when a fraction of the plots in the high risk area are randomly replaced with plots that lie outside of the high risk area. The overlap parameter denotes fraction of the plots that lies in the high risk area. The detection probability is plotted as a function of number of samples collected from the high risk area in the SOO plan (total number of samples=18).74

Figure 4.8 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in flooded field from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the first trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$). 76

Figure 4.9 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in flooded field from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the second trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$).77

Figure 4.10 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in flooded field from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the third trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$)..78

Figure 4.11 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with animal house nearby from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the first trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$).79

Figure 4.12 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with animal house nearby from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the second trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$).80

Figure 4.13 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with animal house nearby from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the third trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$).81

Figure 4.14 Example samples drawn from the field with animal house nearby according to three sampling plans (18 samples for each). Solid symbols represent positive samples of *E.coli* (37°C) indicator and empty symbols represent negative samples of *E.coli* (37°C) indicator based on 270 subplots in the field.82

Figure 4.15 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with power line above from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the first trial). Red bar and “*” indicate that the means were significantly different ($p < 0.05$) based on Fisher’s exact test and ANOVA test, respectively.83

Figure 4.16 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with power line above from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the second trial). Red bar and “*” indicate that the means were significantly different ($p < 0.05$) based on Fisher’s exact test and ANOVA test, respectively.84

Figure 4.17 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with power line above from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the third trial). Red bar and “*” indicate that the means were significantly different ($p < 0.05$) based on Fisher’s exact test and ANOVA test, respectively.85

Figure 5.1 Example samples drawn according to different sampling plans. A. The samples drawn according to the random (cross), stratified random (triangle) and BGO (square) are shown together with contaminations and likelihood of contaminations. The order of the BGO samples are labeled with numbers. B. Estimated distribution of contamination likelihood in the field from the BGO

sampling plan. C. Standard deviation of the estimated likelihood based on the BGO sampling plan. 105

Figure 5.2 Performance of three sampling plans with different types of contamination patterns. (A) Detection probability of the contaminated site for different sampling plans. (B) Average contamination likelihood of samples collected from different sampling plans. (C) The maximum contamination likelihood among the 18 samples collected from different sampling plans. The error bars represents standard deviation across different fields with the same types of contaminations. Values from bars with different letters are significantly different based on an ANOVA and Tukey's *post hoc* test ($P < 0.05$).
 107

Figure 5.3 Performance of three sampling plans with different number of samples collected. The detection probability is plotted as a function of number of samples. Note that the first 6 samples of the BGO plan are drawn randomly. The difference between BGO and other sampling plans is significantly when number of samples is higher than 10. 108

Figure 5.4 Detection probability of the BGO plan as a function of number of samples per iteration. The detection probability of BGO is calculated with increasing number of samples per iteration (decreasing number of iterations) on fields with point contaminations. The number of samples collected in the initial random sampling period is 6 for $N_{\text{sample}}^{\text{iter}} \leq 6$ and 9 for $N_{\text{sample}}^{\text{iter}} = 9$. The total number of iterations is 12, 6, 4, 3, 2, 1 for $N_{\text{sample}}^{\text{iter}} = 1, 2, 3, 4, 6, 9$ respectively.

The detection probability for random and stratified-random plans are shown as blue and red dashed lines respectively..... 109

Figure 5.5 Robustness of the BGO plan to noise in contamination likelihood measurement. Mean detection probability is plotted as a function of the variance of the multiplicative Gamma noise. The mean detection probabilities for random and stratified-random plans are shown as blue and red dashed lines respectively. This simulation is performed with point contamination..... 110

List of Abbreviations

ANOVA	Analysis Of Variance
APC	Aerobic Count Plates
BGO	Bayesian Global Optimization
CDC	Centers for Disease Control and Prevention
CMREC	Central Maryland Research & Education Center
DCM	Dairy Cow Manure
DCMSE	Dairy Cow Manure Solids Extract
EC37	<i>E. coli</i> 37°C
EC44	Thermo tolerant <i>E. coli</i>
GAP	Good Agricultural Practices
GHP	Good Handling Practices
HACCP	Hazard Analysis Critical Control Point
HUS	Hemolytic Uremic Syndrome
LGMA	Leafy Greens Marketing Agreement
MOE	Metric Optimization Engine
MACR	MacConkey Agar with 80µg/mL rifampicin
PAA	Peroxyacetic acid
PCR	Polymerase Chain Reaction
PVC	Premature Ventricular Contraction
PW	Peptone Water

QCM	Quartz Crystal Microbalance
RMMs	Rapid Microbial Methods
RifR	Rifampicin-resistant
SCRI	Specialty Crop Research Initiative
SD	Standard Deviation
STEC	Shiga Toxin-producing E. coli
SOO	Samples of Opportunity
SPR	Surface Plasmon Resonance
TC37	Total Coliforms
TC44	Thermo tolerant Coliforms
TSBR	Tryptic Soy Broth with 80 µg/ml Rifampicin
USDA	United States Department of Agriculture

Chapter 1: Introduction

1.1 Increasing demand for fresh produce

Fresh fruit and vegetables are an important part of our daily diet. The demand for fresh produce has increased continuously in the United States [1] and other countries in the world in the past few decades. This is largely due to consumer awareness about the linkages between diet and health, as well as rising incomes [1]. In 2014, the United States produced about 129 billion pounds of commercial vegetables, and the average vegetable and pulses consumption reached 385 pounds per person [2]. Despite this seemingly large number, after adjustment for losses and conversion to cups of produce consumed daily, each American consumed only 1.6 cups per day on average, which is well below the daily recommendations of 2.5 cups from the *2010 Dietary Guidelines for Americans* [2]. Therefore, the demand for fresh produce is likely to increase further in the coming years as consumers gain a better knowledge of diet and health issues fueled by the availability of information, especially through social media.

1.2 Microbiological risks of fresh leafy greens

While fresh fruit and vegetables are an indispensable component of a healthy diet, contamination by foodborne pathogens due to improper handling can lead to serious diseases and even death. Approximately 48 million people are affected by food-related diseases every year in the United States. A recent report from the U.S. Centers of Disease Control reviewed foodborne illness between 1998 and 2008. It reported that almost half (46%) of foodborne illnesses and outbreaks between 1998

and 2008 were attributable to fresh produce, with leafy greens causing about one fifth (22%) of the foodborne illness [3].

There are several reasons why cut leafy greens have higher risks for foodborne illness. Leafy greens typically grow near the soil's surface and have large surface-to-volume ratios, thus are more likely to be affected by pre-harvest contamination [4]. It is also known during post-harvest activities, pathogens can grow quickly on cut salad greens and reach high levels of contaminations in short periods of time if the leafy greens are not kept at cold enough temperatures [5]. Due to these reasons, there have been numerous leafy green associated foodborne illness outbreaks. For example, an *Escherichia coli* outbreak linked to raw spinach caused 3 deaths and 199 illness, including 102 hospitalizations [6].

1.3 Microbiological testing for fresh produce

Microbiological testing of foods for specific pathogens or indicator microorganisms is used extensively to help ensure the safety of foods. Particularly with foods that receive a minimum of processing, microbiological testing can be a critical component of an overall food safety system. When sampling plans and methodology are properly designed and performed, microbiological testing can provide important information about the risk of fresh produce. However, if poorly designed or executed improperly, testing can provide inaccurate or misleading information that may create either unwarranted concerns or false reassurances about the safety of produce.

Pathogen contaminations can occur at every stage of production from the field to the table. Although microbiological testing is important along the whole supply

chain, the probability of detecting the pathogens is very low when limited number of samples are distributed across multiple stages. Instead, the HACCP (Hazard Analysis Critical Control Point) approach to risk reduction requires preventive interventions at critical points in a process where the risk of contamination is high [7]. In the case of fresh leafy green vegetables, it is preferred to perform product testing in the field prior to shipment to the processing facility, where knowledge of the environment and likely sources of contamination can potentially increase the statistical power of the testing regime [8].

1.4 Introduction to sampling

In statistics, sampling is concerned with the selection of a subset of individuals to estimate characteristics of the whole population. The only way to ensure 100% safety of a field of fresh produce is to test every part of it, yet this is not practical as it would leave no product to sell . Instead, microbiological testing is performed on a subset of the food product. Considering the large batch sizes associated with food products, the low frequency and levels of pathogens in the product, and the heterogeneity of pathogen distribution in the food product, it is critical to design appropriate sampling plans to improve the statistical power of the microbiological testing [9].

A variety of sampling methods can be employed, either individually or in combination. For example, a simple random sampling plan selects all subsets of a batch of food product with equal probability. A systemic sampling plan selects every *k*th elements in the batch. A stratified random sampling plan first divides the whole batch of food into distinct, independent strata, and then randomly select equal

numbers of samples from each strata. Choice of sampling plans depends on nature and quality of the food batch likely sources of contamination and cost or operational concerns related to testing. However, there is a lack of systemic comparisons among sampling plans for assessing the food safety attributes of pre-harvest fresh produce.

Chapter 2. Literature Review

Over 250 types of pathogens and toxins can be transmitted by food, with 31 of them classified as major foodborne pathogens [10]. According to a report from the Centers for Disease Control and Prevention (CDC) of foodborne illness, about 1 in 5 foodborne illnesses were associated with leafy green vegetables consumption, more than any other type of food [3]. Microbiological testing is considered an integral component of food safety control systems from leafy greens to verify that food safety controls are maintaining contamination levels within acceptable levels. Sampling is a required and essential first step. However, selection of an appropriate sampling plan depends on knowledge of both microbiological hazard and sources of contamination. To fully understand the efficacy of sampling plans, the knowledge of the behavior of the microbiological hazard and the sources of contamination are required. This becomes particularly important when considering “samples of opportunity” approaches to sampling. In this section, the common microbial hazards associated with fresh produce, contamination sources and sampling for food safety are reviewed.

2.1 Hazards associated leafy green at pre-harvest

2.1.1 *Salmonella*

Salmonella enterica subspecies *enterica* is a gram-negative, rod-shaped bacterium commonly found in the gastrointestinal tract of both exothermic and

endothermic animals, including humans. It is a member of the Enterobacteriaceae. *Salmonella* can be divided into serotypes based on antigens that the organism presents. Scientists have classified *S. enterica* into over 2,500 serotypes. *Salmonella* is the most commonly diagnosed and reported foodborne illness associated with fresh produce, causing 15.19 cases of illness per 100,000 people in the U.S annually [11]. Despite some recent progress in reducing *Salmonella* infections, the infection rate is still well above the national goal for 2020: 4 cases per 100,000 people.

Salmonella has remarkable adaptability and high tolerance for environmental stress such as UV radiation [12, 13]. *Salmonella* are widely distributed in nature and survive well in a variety of foods, such as poultry, eggs, dairy products and fresh produce [14]. Furthermore, *Salmonella* can persist in the environment for extended periods, and cause infections after the ingestion of low doses, e.g., 10-100 cells [15]. Moreover, *Salmonella* can be carried in the intestines of domestic and wild mammals, birds, and reptiles. It is also present in the feces of pets, such as cats, dogs, hamsters, and guinea pigs, and humans can serve as asymptomatic carriers. These properties make it hard to control *Salmonella* contamination.

2.1.2 Shiga Toxin-producing *E. coli* (STEC)

Most strains of *E. coli* are benign inhabitants of the gastrointestinal tract of endothermic animals. However, strains that produce Shiga toxins, originally discovered in 1977 [16], can cause serious illness in people, especially children and elderly people [17]. Infections with Shiga toxin-producing *E. coli* can occur through consumption of undercooked ground beef, unpasteurized milk, cheese, and juice; contaminated raw fruits and vegetables [18]; water contaminated with animal feces;

or by direct contact with farm animals or their environment. A 2006 outbreak linked to bagged baby spinach caused more than 200 people to become ill and at least 30 to develop hemolytic uremic syndrome (HUS), a serious and potentially fatal kidney pathology associated with Shiga-toxin producing *E. coli* infections [19].

Surveys in the United States and Canada indicate wide spread distribution of *E. coli* O157:H7 in cattle operations [20]. A recent study identified the proximity to cattle feedlot as a risk factor for *E. coli* contamination of leafy greens[21]. *E. coli* O157:H7 may be present in animal manures and slurries, particularly cattle derived material [22], and can contaminate fresh produce during manure application. Wildlife, such as deer, may also be carriers for *E. coli* O157:H7 [23].

2.1.3 *Cryptosporidium*

Cryptosporidium is a parasitic protozoa that can cause gastrointestinal illness with diarrhea in humans and animals. Water, including both drinking water and recreational water is the most common way to spread the parasite. *Cryptosporidium* is a leading cause of waterborne disease among humans in the United States [24]. Many *Cryptosporidium* associated outbreaks were reported in the United States and other countries around the world [25]–[27].

A number of studies have reported contamination of leafy green vegetables with *Cryptosporidium*. Dixon et al. reports the presence of *Cryptosporidium* on 5.9% of ready-to-eat packaged salads and leafy greens samples purchased at retail in Canada [28]. Maikai et al. collected 200 fresh vegetable samples in Zaria metropolis, Kaduna State, Nigeria and found 35% (70/200) samples were positive for *Cryptosporidium*, among which lettuce had the highest contamination rate (40%)

[29]. Rai et al. reported detection of *Cryptosporidium* in unprocessed food (unpasteurized milk and meat samples) in India [30].

The parasite can survive outside the body for long periods of time, and is very tolerant to chlorine disinfection. In a recent study, Chandra et al. (2014) evaluated six different wash solutions for their effectiveness in removing *Cryptosporidium parvum* from basil. At high contamination condition (1,000 oocysts/25 g basil), the protozoa's oocysts could be recovered from all samples regardless of wash solutions. However, at low contamination conditions (100 oocysts/25 g), the recovery rates were in the range of 18.5% to 70.4% [31].

2.1.4 Norovirus

Norovirus is a highly contagious human virus that can be transmitted by fecal contamination of food or water, by touching contaminated surfaces and through person-to-person contact. Norovirus is the most common cause of viral gastroenteritis in humans worldwide. The virus has been reported to cause 267 million infections and over 200,000 deaths each year [32], mostly in infants and the elderly in developing countries [33, 34].

The consumption of contaminated uncooked food such as leafy green vegetables and fruits has been identified as a common source of norovirus outbreaks. In a recent study, Baert et al [35] analyzed a large number of samples of leafy greens, fresh fruits and other types of fresh produce in Belgium, Canada and France. Norovirus was detected by real-time RT-PCR in 28.2%, 33.3% and 50% of leafy green samples tested in Canada, Belgium and France, respectively.

A number of recent studies assessed the effects and efficacy of different washing methods on reducing human norovirus on leafy green. Bae et al. [36] used artificially contaminated vegetables to evaluate the efficacy of washing treatments in the removal of norovirus. The authors found that wash treatments achieved a 0.69 to 1.29 log reduction in norovirus levels from the surfaces of the vegetables. Baert et al [37] compared the efficiency of sodium hypochlorite (NaOCl) and peroxyacetic acid (PAA) to reduce a murine norovirus (a surrogate for human norovirus) at two inoculation levels with different organic loads [37]. They found that 200 mg/L NaOCl or 250 mg/L PAA was needed to obtain an additional 1-log reduction of murine norovirus on shredded iceberg lettuce, whereas only 250 mg/L PAA achieved this for normal bacterial pathogens. Therefore, PAA and NaOCl are useful in preventing cross-contamination during the washing process, but does not directly cause a significant reduction of the number of pathogens present on lettuce. The authors further found that the effectiveness of NaOCl, but not PAA was greatly influenced by the presence of organic material.

2.1.5 *Shigella*

Shigella is a gram negative, rod-shaped bacterium closely related to *Salmonella* and *E. coli*. It is the causative agent of shigellosis in humans and other primates, but not in other animals. *Shigella* is a leading bacterial cause of diarrhea worldwide. Every year, there are an estimated 500,000 cases of shigellosis in the United States, or 4.82 cases per 100,000 individuals [38].

Numerous outbreaks of *Shigella* have been linked to consumption of fresh leafy green vegetables. For example, a *Shigella sonnei* outbreak with 46 cases

occurred in Norway during October 2011 [39]. Epidemiological evidence and trace back investigations linked the outbreak to the consumption of imported fresh basil.

Several recent studies have evaluated different methodologies for removing *Shigella* from fresh produce. Two studies showed that X-ray radiation significantly reduced *Shigella* and other pathogens on leafy green surfaces [40, 41]. Another study reported that a combined treatment with malic acid and ozone reduced pathogen populations by 4.4 log in radish and 4.8 log in mungbean sprouts [42]. There is also evidence that organic acids, such as acetic acid, citric acid and lactic acid can be effective in reducing *Shigella* contamination [43].

2.2 Sources of Pre-harvest Contamination

Many factors affect the frequency and extent of contamination of fresh produce during cultivation of produce. Pre-harvest contamination can occur via direct contact with contaminated manure and irrigation water, as well as wild animals, insects and field workers. In this section, common sources of pre-harvest contamination are reviewed.

2.2.1 Irrigation water

Irrigation water is considered an important vehicle for contaminating leafy greens with foodborne pathogens. Irrigation water can be contaminated by pathogens from animal fecal deposits or contact with contaminated surface runoff [44]. Once contaminated, pathogens can survive in water for extended periods [45]. Indeed, a number of *E. coli* O157:H7 outbreaks have been linked with contaminated

water. Irrigation water is a significant potential source of pathogens during pre-harvest conditions [46 – 48].

There is substantial variation in the quality and safety of irrigation water, depending on its source. In the United States, surface water and ground water are commonly used to irrigate a variety of crops. It is well established that human pathogens can contaminate and persist in surface and ground water [49, 50]. A recent study collected a total of 123 samples at 18 sites across New York State and detected generic *E. coli*, a microbial indicator of fecal contamination in 33% of the samples and *Salmonella* in 43% [51]. Fecal indicator organisms have also been reported in well water used for drinking and irrigation [52]. A recent study in southern Brazil detected *E. coli* O157:H7 in irrigation and wash waters [53].

A number of studies have assessed the effect of irrigation method on transmission and persistence of human pathogens on leafy green produce. Solomon et al. [54, 55] demonstrated the transmission of *E. coli* O157:H7 through spray and surface irrigation. The pathogen was found to persist for 20 days following spray irrigation with contaminated water. A study of the survival dynamics of *E. coli* after introduction into the phyllosphere and soil of spinach via spray irrigation suggested a 6-day period between the last irrigation and harvest would minimize the risks of *E. coli* survival in the spinach phyllosphere (but not in the soil) [56]. One study showed that *E. coli* O157:H7 contamination in soil persisted for more than 5 months after application of contaminated compost or irrigation water, with the effects of irrigation water and manure being similar [57]. Another study showed that *Salmonella*-containing manure compost and irrigation water could contaminate soil

and root vegetables for several months [58]. Conversely, a recent study found that under growth chamber conditions, *E. coli* O157:H7 populations in irrigation water that complied with the Leafy Greens Marketing Agreement (LGMA) water standards will not persist for more than 24 h when applied to the foliar surfaces of spinach plants [59].

In addition to contaminating the surface of produce, irrigation water may also lead to pathogen internalization. Erickson et al. (2010) showed that after applying contaminated irrigation water that contained *E. coli* O157:H7 for long periods of time (48-69 days), the pathogen was detected both on the surface and internally on spinach and lettuce [60]. The internalization of murine norovirus into hydroponically grown Romaine lettuce was observed when the virus was present in the water source at high levels [61].

2.2.2 Manure, green waste and compost

Animal manure, compost, and green waste are key ingredients for fertilizer and soil amendments in both organic and conventional farming. However, such biologically-derived materials have the potential for carrying pathogenic microorganisms. These biomaterials are potential sources for pathogens either through their use as soil amendments or through accidental cross-contamination. The United States Department of Agriculture (USDA)-Agricultural Marketing Service incorporated the Produce Good Agricultural Practices (GAP) harmonized food safety standard into its GAP & Good Handling Practices (GHP) audit program in 2011 [62]. It is recommended that the time between application of manure to produce production areas and harvest should be maximized [63].

Many studies have reported the presence of human pathogens in animal manure and green waste. Jay-Russel et al. [64] isolated *Salmonella* from horses, wild turkeys and an edible home garden fertilized with raw horse manure from a rural farm in costal Northern California. A number of factors impact the growth of pathogenic bacteria in manure compost (e.g., moisture content, strain variation, level of background microflora, inoculum, duration of composting, and temperature and related thermal factors [65 – 67]. Avery et al. [68] reviewed the prevalence and survival of pathogens in green waste compost. They concluded that zoonotic bacteria such as verotoxigenic *E. coli* and *S. enterica* are unlikely to survive an effective composting process, whereas spore-forming microorganisms are more resistant to composting and equally ubiquitous in the environment.

Human pathogens in manure or compost may increase the risk of fresh produce contamination through attachment and internalization. Wei et al. [61] observed gold-labeled murine norovirus on the lettuce surface, inside open cuts, and occasionally within stomata after lettuce was exposed to the pathogen. The pathogens entry into open cuts and stomata is especially risky as they are likely to be protected from subsequent sanitization treatments [61]. Mootian et al. [69] studied the transfer of low numbers of *E. coli* O157:H7 from soil, manure-amended soil and water to growing lettuce plants. They found that approximately 30% of the mature plants initially irrigated with or grown in contaminated soil (including manure amended soil) for 15 days were positive for *E. coli* O157:H7 [69]. Ongeng et al. [70] demonstrated that under tropical conditions the likelihood of surface contamination and internalization of *E. coli* O157:H7 and *S. Typhimurium* in

cabbage leaf tissues at harvest depended on the inoculum concentration and the time of manure application.

2.2.3 Animal and insect activity

Intrusion of wild animal and cattle presents another route of pathogen contamination. Animals may carry foodborne pathogens and contaminate crops directly via fecal deposition or indirectly through fecal contamination of soil or irrigation water [71]. For example, the *E. coli* O157:H7 strain associated with the highly publicized 2006 spinach outbreak was isolated from domestic cattle and feral swine from adjacent rangeland. It has also been reported that migratory birds can carry *Salmonella bongori* [72]. Fecal contamination of crops by animals is now considered a significant risk factors at the pre-harvest stage [69, 71, 73].

Insects can also be a potential source of contamination. Talley et al. [74] studied the association of *E. coli* O157:H7 with filth flies captured in leafy green fields adjacent to cattle-occupied rangeland habitats. A subset of the filth flies was found positive for *E. coli* O157:H7. The authors also demonstrated that flies are capable of contaminating leafy greens under laboratory conditions. Another group of researchers assessed whether physical damage to the lettuce leaves could cause internalization of *E. coli* O157:H7 [75]. However, they found no internalization occurred due to exposure to insects [75].

2.2.4 Human activity

Activities and personal hygiene of field workers are potential sources of pre-harvest contamination. A recent study reported that generic *E. coli* was significantly reduced after field workers were trained to use portable toilets and hand-washing stations [76]. Annual worker training is a required part of Good Agricultural Practices. Appropriate worker supervision can reduce the risk of contamination.

2.3 Sampling in food safety

2.3.1 Importance of food safety sampling

Sampling and subsequent microbiological testing are widely used by government and industry as part of their programs to ensure safe food products. While it is widely recognized that testing by itself cannot ensure food safety, it is equally recognized that testing is an important part of most food safety systems. The major limitation associated with sampling of foods is that the only way to absolutely ensure that no serving of food contained a pathogenic microorganism would be test every serving. Since the microbiological testing of foods is a destructive process, this would leave no food for consumption. It is therefore the common practice to select a fraction of the food product for analysis, assuming the samples collected are representative.

Sampling can occur at many stages in the farm-to-fork chain. For example, the industry uses sampling to verify the microbiological quality of raw material and ingredients. The government use sampling to verify safety and quality of imported and domestic food products [77]. It is also common to sample fresh produce just

prior to harvest [4, 78]. End-product testing is sometimes required when there are no critical control points in the production process (e.g., raw or minimally processed ready-to-eat foods). Furthermore, microbiological sampling is one of key tools used to verify on an ongoing basis that a food safety system is working as intended. Using a realistic sampling scheme, it is possible to test for absence of a pathogen in a batch of food to a specified level of confidence, but this can lead to large type II errors when not carried out correctly [9, 79].

2.3.2 Introduction to sampling plans

Samples are drawn from a batch of food product or from the field according to sampling plans. A sampling plan defines the number of sample units to be taken (n); the analytical unit size; the analytical reference method, the microbiological limit (m) that separates good quality from non-acceptable or defective quality; and the maximum allowable number of sample units (c) yielding a positive test result (which is usually set to zero for pathogens) [80].

There are two types of sampling plans in statistics, variables plans and attributes plans. Variable plans are intended for quality characteristics measured in a continuous scale, and require the knowledge of the statistical distribution (e.g., Gaussian distribution). Attribute plans test against a single criterion or attribute (e.g., the presence of *E. coli* O157:H7 in the sample).

2.3.3 Sampling strategies

Within the framework identified above, a variety of sampling strategies can be employed to specify how samples are drawn from a population. Several commonly used sampling strategies pertinent to pre-harvest or post-harvest testing of leafy greens are reviewed below.

2.3.3.1 Simple random sampling

Simple random sample is a sampling strategy where in such that each part of the food produce has an equal chance of being selected [81]. This sampling strategy eliminates bias. In practice, the food product lot is divided into small batches/locations of the same size. Each batch/location is assigned a number and then computer generated random numbers are used to select the batch/location to be tested.

2.3.3.2 Systematic sampling

Systematic sampling (also known as interval sampling) arranges the population under consideration according to some ordering scheme, and then selects elements at a regular interval through that ordered list. Systematic sampling generally involves a random start and then proceeds with the selection of every k th element from then onwards. Jongenburger et al. (2011b) detailed the effects of systematic sampling on the probability of detecting positive sample unit. The study indicated that systematic sampling is preferred over random sampling to detect localized contamination fraction [82]. Similar results have also been reported by previous studies [83, 84]. However, a potential drawback of systematic sampling is

the potential to miss systematic, reoccurring contamination events, such as a contaminated filler head or dripping condensation [79].

2.3.3.3 Stratified random sampling

Stratified random sampling is a design where the target population is divided into non-overlapping parts or sub-regions called strata. A simple random sampling plan is applied within each stratum. This often improves the representativeness of the sample by reducing sampling error. Because stratified random sampling combines the qualities of systematic and random sampling, this strategy is often preferred [79]. Potential disadvantages of the stratified random sampling strategy include difficulty in selecting relevant stratification and higher implementation costs.

2.3.3.4 Z-pattern sampling

A “zig-zag” or “Z” traversal sampling pattern is commonly used for sampling with squared-shaped or long narrow rectangular fields. This sampling plan has been used for field sampling such as insect sampling [85] and soil sampling [86]. An advantage for the Z-pattern sampling plan is that it is more efficient to collect the same number of samples. The total distance travelled with a Z-pattern sampling strategy is generally shorter than simple random sampling or stratified random sampling.

Chapter 3. Use of Simulation Modeling and Field Validation to Evaluate Three Sampling Plans for the Detection of Pathogenic Bacteria on Pre-harvest Leafy Greens

3.1 Abstract

Recent outbreaks of foodborne disease associated with leafy greens have led to increased pre-harvest testing for pathogens and indicator microorganisms. However, the scientific and statistical rationale and performance attributes for pre-harvest sampling plans are not well understood. The performance of three pre-harvest sampling plans, random, stratified random, and Z-pattern sampling, was evaluated by consideration of their mathematical derivations, computer simulations and field validation. As anticipated, the simulation modeling showed that the probability of detecting at least one positive sample increased by increasing the number of contamination sites in the field and the number of samples analyzed. Consideration of probabilistic basis of the sampling plans indicated that the three sampling plans mean detection rates were similar. However, use of simulation modeling to assess the uncertainty associated with the three sampling plans indicated that the inherent variability of the Z-pattern sampling plan was substantially greater than the other two sampling plans. This uncertainty is most dramatic when the number of contamination sites and number of samples analyzed

were small. A simple tool was developed in Matlab that allows the user to evaluate the effectiveness of these three sampling plans. A limited validation study also observed that Z-pattern sampling had higher variability than other two sampling plans. The results of this study indicate that while the mean result obtained with all three sampling plans is similar, the performance of the random or stratified random sampling plans are less variable, particularly when the number of contamination sites or number of samples analyzed are small.

3.2 Introduction

As one of the tools for managing microbial food safety risks associated with leafy greens, leafy green producers and their clients are increasingly requiring pre-harvest microbiological testing just prior to harvest as part of their risk management programs. Such testing may target indicator microorganisms (e.g., coliforms, generic *E. coli*) or specific pathogens (e.g., *S. enterica*) as a means of verifying that good agricultural practices and other pre-harvest mitigation steps have been effectively implemented. Thus, it is not surprising that during the past decade there has been a substantial research effort in the development of methods for the detection of various microorganisms associated with leafy greens. However, as important for effective microbiological testing programs is an understanding of the sampling plans used to implement such microbiological testing systems. This requires knowledge of the statistical principles that are the foundation of microbiological sampling and means for measuring the uncertainty associated with selected sampling schemes. Considering the large batch sizes associated with pre-harvest leafy greens, the low frequency and levels of pathogens or indicator

microorganisms, and the heterogeneity of pathogen distribution in leafy greens fields, it is critical that appropriate sampling plans are selected to optimize the statistical power of the microbiological testing and appropriately interpret the results [9]. However, despite the fact that pre-harvest testing of leafy greens is a common practice, there are few systematic comparisons of different sampling plans. Thus, the objectives of this segment was to review the mathematical principles underlying three commonly used sampling plans and use simulation modeling and field validation to assess the relative uncertainty associated with these sampling plans based on random contaminations.

3.3 Materials and Methods

3.3.1 Simulation modeling

Three commonly used sampling plans, random sampling, stratified random sampling, and Z-pattern sampling, were compared by (i) evaluating their mathematical basis and (ii) characterizing their relative performance using computer simulation modeling. This combination of approaches allowed estimation of both mean detection probabilities and the relative uncertainty of detection likelihood.

A variety of sampling plans can be employed, either individually or in combination. Choice of pre-harvest sampling plans depends on nature and quality of the leafy greens, the prevalence and likely sources of contamination, the degree of confidence required, and the cost and related operational concerns associated with microbiological testing (e.g., time needed to complete the microbiological analyses).

The three sampling plans selected were based on their common use in leafy green testing.

3.3.1.1 Description of Model Fields Used to Compare Sampling Plans

For the purposes of this study, we considered a generic, rectangular leafy green field that was divided into $N_x \times N_y$ plots. Each plot, in turn, was subdivided into $S_x \times S_y$ subplots. The total number of subplots is $N_{subplot} = N_x N_y S_x S_y$. We assumed that the field was contaminated N_c times ($N_c > 0$), with each contamination event randomly contaminating a subplot. For example, Fig. 3.1-3.3 depict the actual field layout used in subsequent blinded field trials (see below), where we used a 5×6 plot array (30 plots), with each plot having 9 subplots ($N_x = 5, N_y = 6, S_x = 3, S_y = 3$). The x's within the field represent the 30 subplots that were selected for sampling ($N_{sample} = 30$).

3.3.1.2 Mathematical Basis for Sampling Plan Efficiency

3.3.1.2.1 Random Sampling Plans

A random sampling plan selects N_{sample} subplots randomly from the total of $N_{subplot}$ subplots. An example of a set of 30 random samples is shown in Figure 3.1.

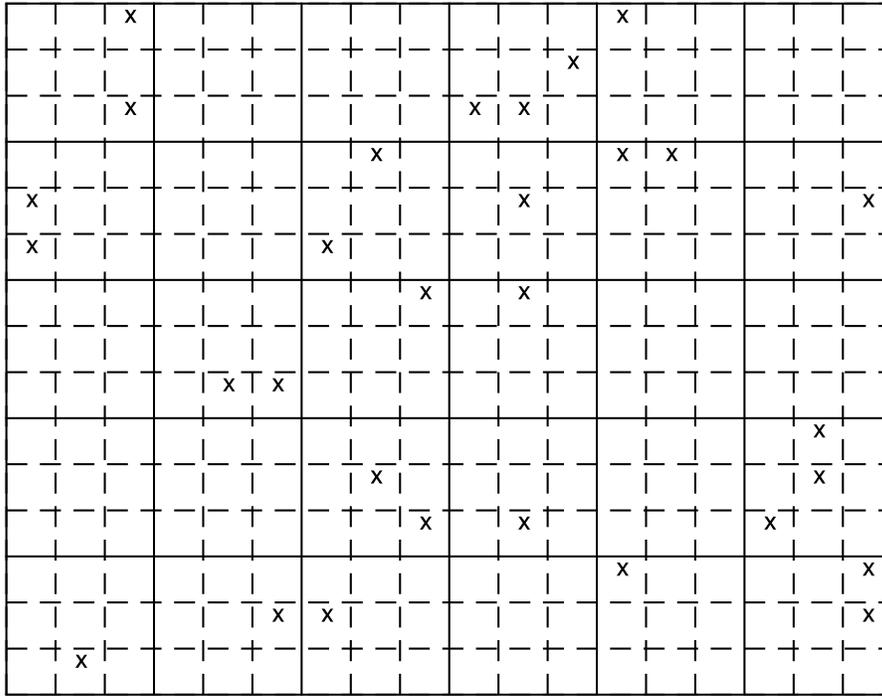


Figure 3.1 Example of samples drawn according to a random sampling plan ($N_{sample} = 30$). The solid lines represent 30 (5 X 6) plots and the dashed lines represent the subplots (9 subplots/plot) for a total of 9 X 30 = 270 subplots. Samples were randomly selected chosen from the 270 subplots.

If the sampling is without replacement, this means that exactly N_{sample} unique samples were selected. The probability that none of the contaminated subplots is detected is,

$$P_{miss} = \left(1 - \frac{N_{sample}}{N_{subplot}}\right)^{N_c} \quad (3.1)$$

Where:

N_{sample} = the number of samples taken

$N_{subplot}$ = the number of subplots in the field

N_c = the number of subplots that are contaminated

P_{miss} = the probability that no positive sample is detected

The detection probability is $1 - P_{miss}$.

If the sampling is with replacement, the probability that a subplot is contaminated (P_c) is,

$$P_c = 1 - \left(1 - \frac{1}{N_{subplot}}\right)^{N_c} \quad (3.2)$$

The probability that a subplot is selected by the random sampling plan is,

$$P_s = 1 - \left(1 - \frac{1}{N_{subplot}}\right)^{N_{sample}} \quad (3.3)$$

Since the contamination is independent of the sampling procedure, the probability that a subplot is both contaminated and selected by the random sampling plan is $P_c P_s$. The probability that the random sampling plan fails to detect any contaminated sample is then,

$$P_{miss} = (1 - P_c P_s)^{N_{subplot}} \quad (3.4)$$

The random sampling with replacement always has a higher miss probability than the random sampling without replacement, because the number of unique samples drawn from the population will be smaller or equal to N_{sample} for random sampling with replacement.

3.3.1.2.2 Stratified Random Sampling Plans

A stratified random sampling plan randomly selects one sample from each plot. An example set of samples drawn according to the stratified random sampling plan is shown in Fig. 3.2. The number of unique samples would normally be exactly $N_{sample} = N_x N_y$ (though one could consider multiple samples per plot, e.g., 2 samples per plot for a total N_{sample} of 60). For each contamination site, the probability that the contamination site is not selected is $(1 - 1/S_x S_y)$. The probability that none of the contaminated subplots is detected is then,

$$P_{miss} = \left(1 - \frac{1}{S_x S_y}\right)^{N_c} \quad (3.5)$$

Note that because $N_{subplot} = N_x N_y S_x S_y$ and $N_{sample} = N_x N_y$, $\frac{1}{S_x S_y} = \frac{N_{sample}}{N_{subplot}}$.

The mean detection probability of the stratified random sampling plan is identical to that of the random sampling plan (without replacement).

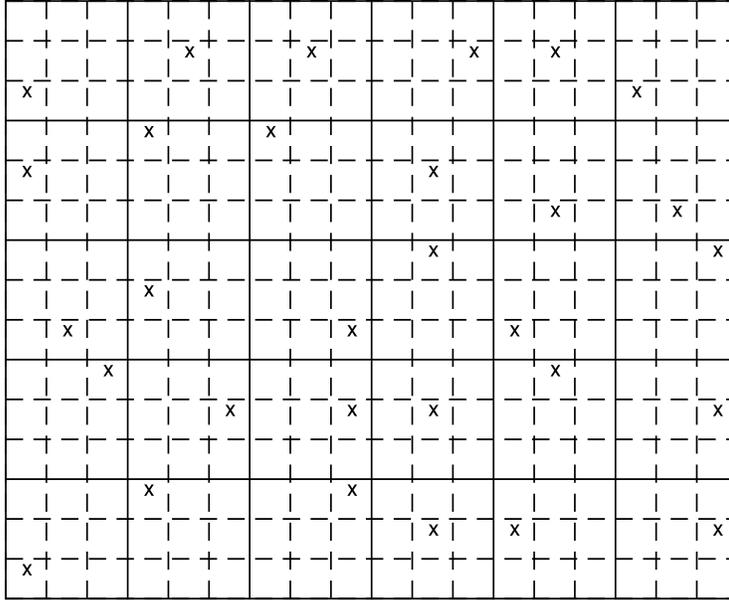


Figure 3.2 Example of samples drawn according to a stratified random sampling plan ($N_{\text{sample}} = 30$). The solid lines represent 30 (5 X 6) plots and the dashed lines represent the subplots (9 subplots/plot) for a total of $9 \times 30 = 270$ subplots. Each plot has one subplot that was randomly selected for sampling.

3.3.1.2.3 Z-pattern sampling plan

The Z-pattern sampling plan only samples from plots lying at two-opposing edges and in one of the diagonal lines (Fig. 3.3). The number of subplots that lay within the selected "Z" region is N_z . In this sampling plan, $N_z = 144$ (Fig. 3.3). For each contamination site, the probability that a contaminated subplot lies in the "Z" region is

$$p = \frac{N_z}{N_{\text{subplot}}} \quad (3.6)$$

The probability that k contamination sites lie in the "Z" region is

$$P_k = \binom{N_c}{k} p^k (1 - p)^{N_c - k}$$

Given that there are k contamination sites in the "Z" region, the probability that none of the contaminations is detected is

$$P_{miss}^k = \left(1 - \frac{N_{sample}}{N_z}\right)^k \quad (3.7)$$

The overall probability of failed detection is

$$P_{miss} = \sum_{k=0}^{N_c} P_k P_{miss}^k \quad (3.8)$$

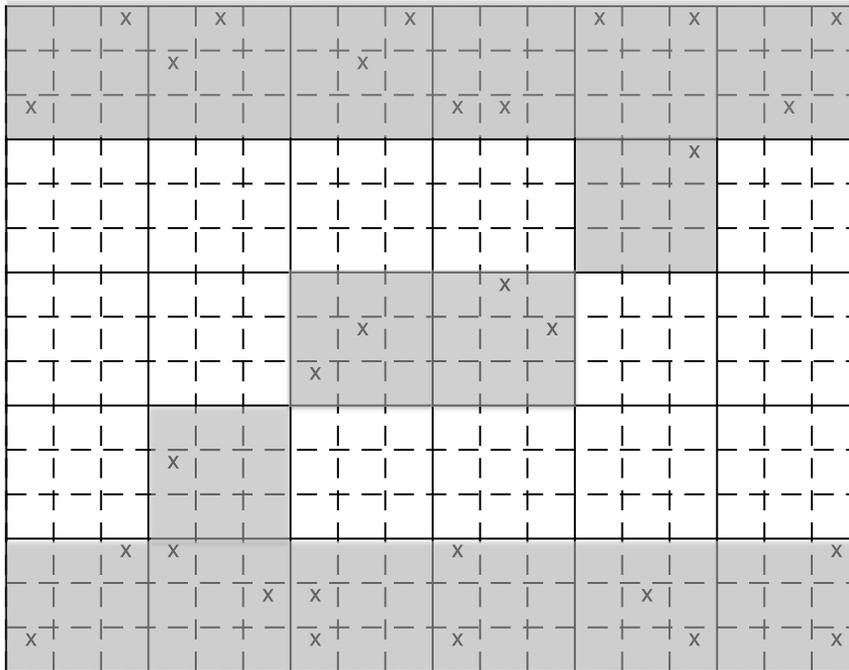


Figure 3.3 Example samples drawn according to the Z-pattern (see shaded plots) sampling plan ($N_{sample} = 30$). The solid lines represent 30 (5 X 6) plots and the dashed lines represent the subplots (9 subplots/plot) for a total of $9 \times 30 = 270$ subplots. Each plot within "Z" area has one or two subplots randomly selected for sampling. No samples were selected in the plots outside the Z-pattern.

3.3.1.3 Comparison of Sampling Plans Using Simulation Modeling

A simulation model was developed in Matlab as a means of validating and better characterizing the three sampling plans. At the beginning of each simulation, a random number between zero and one was generated for each subplot, we then contaminated subplots with the largest N_c random numbers. This generated a randomly contaminated field was generated with N_c subplots contaminated. The field was then sampled using the three types of sampling plans, each with 100 iterations. For each iteration the sample locations were generated according to the different sampling plans independently. A detection probability was estimated based on the number of iterations that a sampling plan successfully detected at least one of the contaminated sites. The simulation was then repeated 100 times, each with a different contaminated field, to estimate the distribution of detection probabilities for each sampling plan. It is important to repeat the simulation multiple times as one sampling plan may be good at detecting certain contamination patterns but not others. This was particularly true with the Z-pattern sampling plan where specific plots are not tested if they fall outside the Z-pattern. Thus, the simulations examined a total of 10,000 combinations of random contamination site assignments and random sampling site assignments based on the number of contamination sites and the number of sampling sites specified.

3.3.2 Validation in the field

Initial “blinded” field trials were conducted at the Beltsville Agricultural Research Center, Beltsville, MD in October 2015 and June 2016, using a field of soybeans and a field of lettuce, respectively. The experimental fields were

inoculated with a known indicator microorganism by one group, with the field subsequently sampled by another group which had no knowledge of where the field was inoculated.

3.3.2.1 Inoculation

3.3.2.1.1 Preparation of dairy solids extracts

Fresh dairy cow manure (DCM) solids were collected from Beltsville Agricultural Research Center. The DCM solids (100 g) was diluted 1:10 into deionized water in a 2-L sterile plastic beaker, stirred manually and with a stir bar for 5 min before the slurry was filtered through two layers of sterile cheese cloth in a Buchner funnel, and collected in 4-L flask. The extract was transferred to 9-L or 20-L carboys, where an equal volume of deionized water was added prior to autoclaving for 1 h at 121°C. The resulting sterile DCM solids extract (DCMSE) was stored at 4°C until used.

3.3.2.1.2 Strains and culture conditions used

A rifampicin-resistant (Rif^R), non-pathogenic *E. coli* strain (TVS 355) was provided by Dr. Trevor Suslow at the University of California Davis. The strain was cultured from frozen stocks stored at -80°C on MacConkey agar (Neogen, Lansing, MI) supplemented with 80 mg/ml rifampicin (Sigma Aldrich, St. Louis, MO) (MACR) and incubated at 42°C for 24 h. Three to five colonies of the *E. coli* strain were inoculated to 200 ml of tryptic soy broth (Neogen) with 80 µg/ml rifampicin (TSBR) and incubated at 37°C for 24 h. A 200-ml portion of each 24-h culture was added to separate 7-L carboys of DCMSE and incubated at 37°C for 48 h. Populations of the

cultured strain were determined after the incubation period by plating serial dilutions in sterile 0.1% peptone water (Becton Dickson, Sparks, MD) onto MacConkey Agar (Becton Dickinson, Sparks, MD) with 80 µg/mL rifampicin (MACR). Plates and 3M™ Petrifilm™ *E. coli*/Coliform Count Plates were incubated at 37°C for 18-24 h. *E. coli* population densities were enumerated and recorded. The carboys were mixed by manual shaking, and were poured into a 13-L backpack sprayer (H.D. Hudson Manufacturing Company, Chicago, IL) immediately prior to spray inoculation of the experimental field.

3.3.2.1.3 Inoculum application

Inoculation was carried out by backpack application onto leaves. *Escherichia coli* levels in the inoculum were ~ 2.3×10^7 CFU/ml. Inoculations were performed on a sunny day. The field is divided into 5×6 plots so that there are a total of 30 plots. Each plot is subdivided into 3×3 subplots. The total number of subplots is $N_{subplot} = 270$. The field was contaminated 10 locations. The subplots to be contaminated were randomly selected using a random number generator. The plot and subplot numbers were recorded, ranging from 1-30 and 1-9, respectively. The GPS coordinates of each location (latitude and longitude) of each subplot center was calculated in a computer in advance (Fig. 3.4). For the first trial, a GPS (Bad Elf GNSS Surveyor (BE-GPS-3300)) was used to locate each of the ten contamination plot/subplots to be inoculated. In the second trial, a more traditional “stakes and strings” approach was used to identify the 270 subplots (Fig. 3.5).

After arriving at the correct location of a plot/subplot selected by the random number generator for contamination, a one-meter² polyvinylchloride (PVC) pipe

was used to identify the inoculation area. The backpack sprayer was used to apply the inoculum within the area. An aliquot of 417 ml (25 s by sprayer) of inoculum were measured out and applied to each 1-meter² subplot. Care was taken to minimize spraying any inoculum outside the designated subplot. The team applying the inoculum recorded the exact positions within the subplot that was inoculated but did not share that information with the testing team until after all analyses were completed.

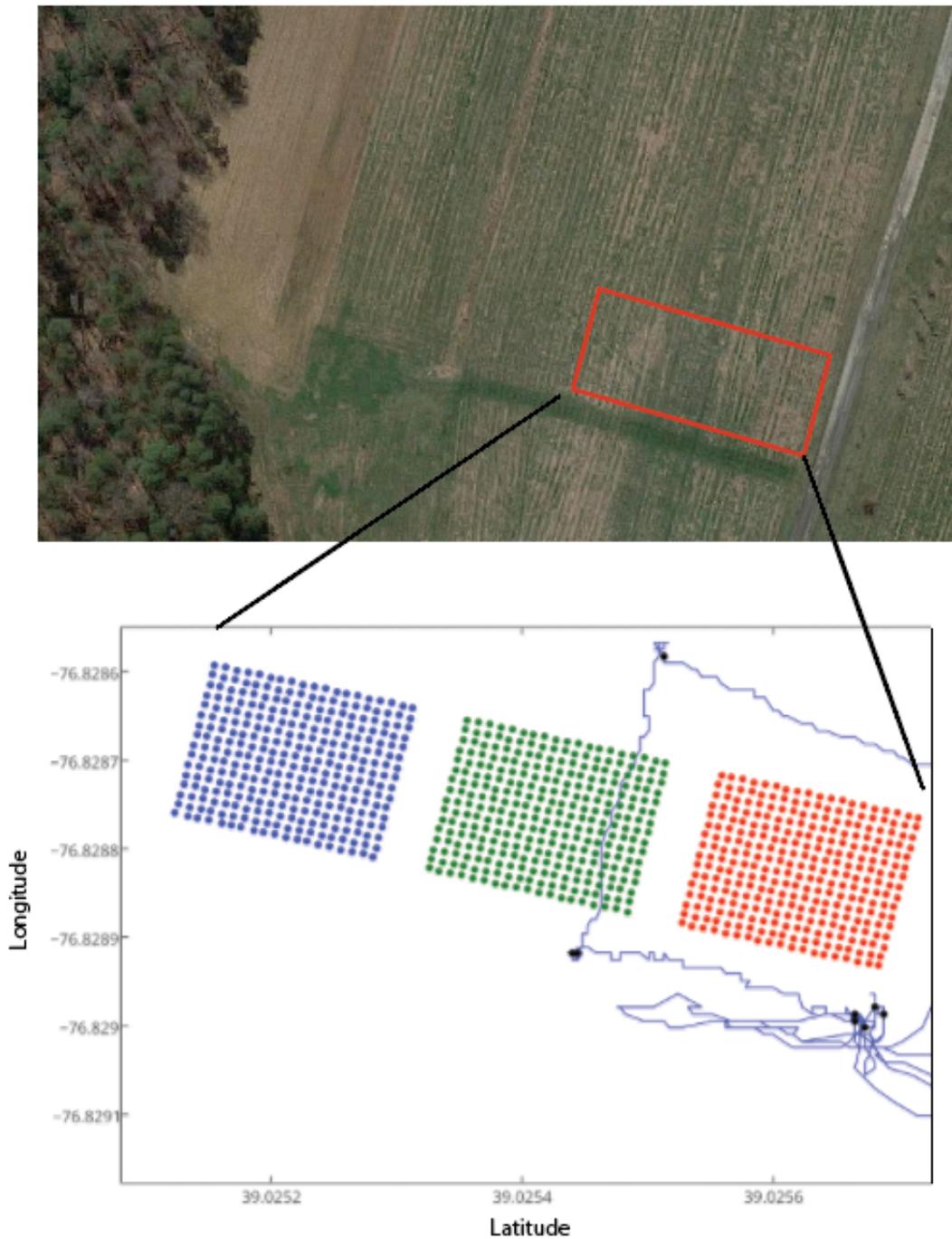


Figure 3.4 Aerial view of the experiment field (top) and the calculated grid for sampling. The black dots represent calibration points used to define the rough location of the experiment field. The red, green and blue dots represents sampling grid for three experiment fields (270 subplots per each experiment field).



Figure 3.5 Example of one subplot identified by strings.

3.3.2.2 Sampling in the field

The sampling team went to the field the day after inoculation. The sampling team took a total of 30 samples for each of the three sampling plans. The sampling sites were selected using the random number generator. For random sampling plan, a new set of random numbers of plot (1-30) and subplot (1-9) were generated by the random number generator. For stratified random sampling plan, a sample was taken from each of the 30 plots. A pair of random numbers was generated to determine the subplot to be sampled within each plot. For Z-pattern sampling plan, the plots used in this sampling plan are fixed. Plots 1, 2, 3, 4, 5, 6, 15, 16, 25, 26, 27, 28, 29, and 30 were sampled twice. Plots 11 and 20 were only sampled once. The computer generated sample locations are shown in Figure 3.6. According to the plot

and subplot numbers, GPS locations of the subplots to be sampled recorded. The sampling team members proceeded to designated GPS location and conducted the sampling.

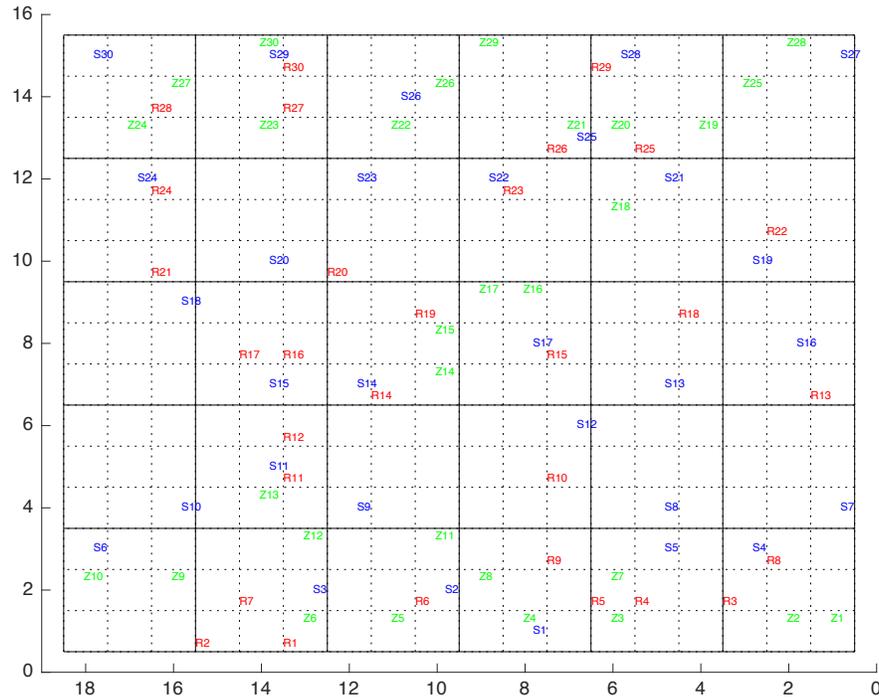


Figure 3.6 Computer generated sampling locations in the field. Samples for random, stratified random, and z-pattern sampling plan were shown in red, blue, and green respectively.

3.3.2.3 Sample processing

At each of the 30 locations for the three sampling plans, leaf samples of approximately 25 g (10 leaves) were collected, transferred to a sample bag, placed in a cooler, and then transported to the laboratory. Samples were processed within 24 h of collection. For each sample, 25 g of leaves were weighed from each sample into a sterile Whirlpak bag (Nasco, Jackson,WI). A 50-ml aliquot of sterile 0.1% peptone water (PW) was added to each bag and stomached in a laboratory stomacher (Seward, Stomacher 400 circulator, U.K.) for 2 min at 250 rpm at room

temperature. A 1-ml aliquot was then pipetted onto 3M™ Petrifilm™ *E. coli*/Coliform Count Plates and MacConkey Agar plates (Becton Dickinson, Sparks, MD) with 80 µg/mL rifampicin (MACR). The presence/absence results were recorded after a 24 h and 48 h incubation. [87].

3.3.2.4 Data analysis

Data analysis was performed with Matlab R2015b (*Mathworks*). The number of positive samples for each sampling plan was calculated. The spatial relationship between positive samples and contamination sites was analyzed visually and statistically.

3.4 Results

3.4.1 Simulation

The theoretical analysis of the mean detection probability was validated by computer simulation. The performance of sampling plans was characterized by running a series of simulations with different numbers of contamination sites and different numbers of samples. The mean detection probabilities as a function of the number of contamination sites from both simulation and probability theory are depicted in Fig. 3.7. As expected, the detection probability is higher with increasing numbers of contaminated subplots and increasing number of samples analyzed. The simulation results agree well with the theoretical derivations.

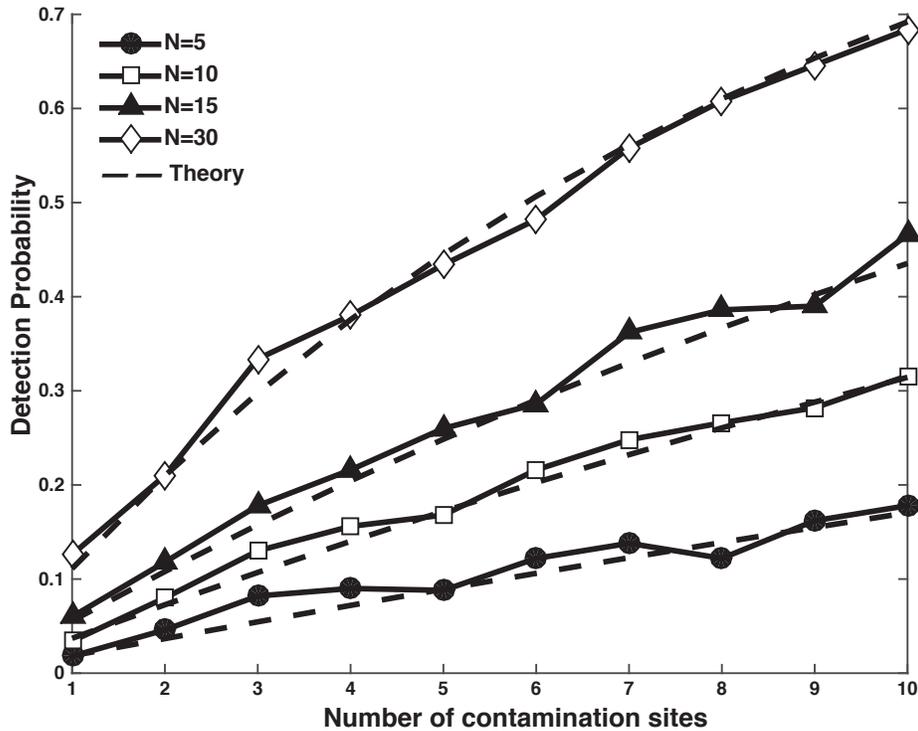


Figure 3.7 Mean detection probability (probability of detecting at least one positive sample) as a function of the number of samples collected and the number of contamination sites. The dashed line represents theoretical calculations and the solid lines represent observed mean detection probability from the random sampling plan simulation. Detection probability increases as a function of the number of contamination sites (x-axis) and the number of samples (shown as different symbols). Each data point represents the mean detection probability from 100 simulations. For each simulation, a randomly contaminated field and estimated the detection probability was determined by repeating the random sampling plan 100 times (100 iterations/simulation).

The characteristics of the three sampling plans were evaluated by examining the distribution of the mean detection probabilities among the 100 simulations (each with 100 iterations) for the three sampling plans (Fig. 3.8). No significant differences in mean detection probabilities were observed among the three sampling plans, and they all agree well with the theoretical predictions when a sufficient number of simulations (and iterations) were performed. However, there

were differences in the standard deviations associated with the mean detection probabilities observed among the three sampling plans, i.e., the Z-pattern testing consistently had larger standard deviations.

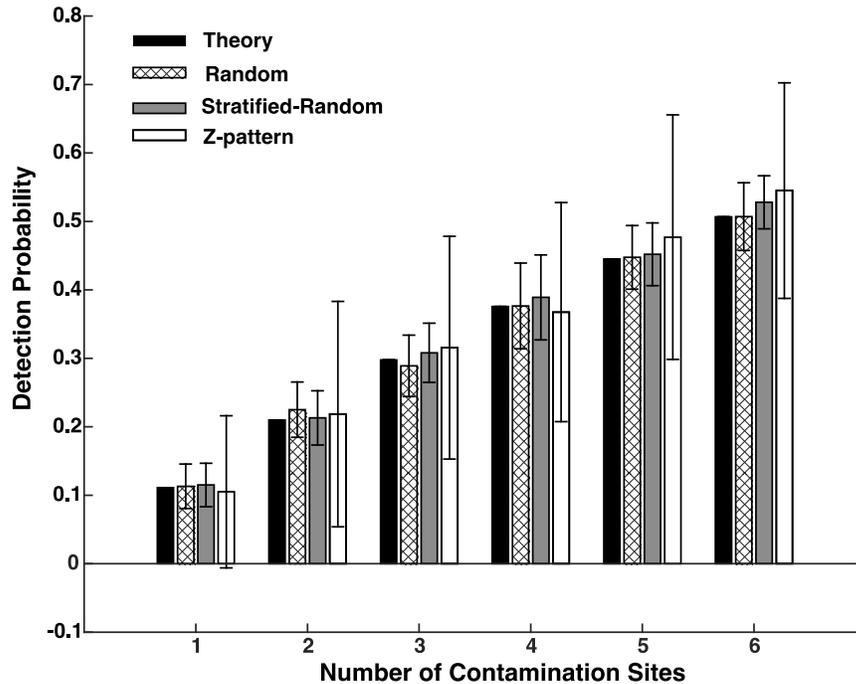


Figure 3.8 Mean detection probability as a function of number of contamination sites. The theoretical estimate captures the average detection probability for three different sampling plans ($N_{\text{sample}} = 30$). For other three simulation results, each column represents the detection probability from 100 simulations. The simulation generated a randomly contaminated field and estimated the detection probability by repeating the three sampling plans 100 times (100 iterations/simulation). The error bars represent ± 1 standard deviation

How the detection probability depends on the number of subplots in each plot was examined. In the original simulations, each plot is divided into 3X3 subplots. In this evaluation the number of subplots per plot was varied from 2X2 to 10X10 (Fig. 3.9). The number of subplots is inversely related with the size of each sample. A 10X10 subplot scheme would mean that the sample size is 100 times smaller than a 1X1 subplot scheme. The mean detection probability decreases

rapidly as a function of total number of subplots. This relationship is roughly linear on a log-log plot (Fig 3.9). There was no difference among the three sampling plans in terms of the mean probability of detection. It should be noted that if the size of the subplots had been kept constant and the size of the plots changing accordingly, similar changes in mean detection probabilities would be anticipated.

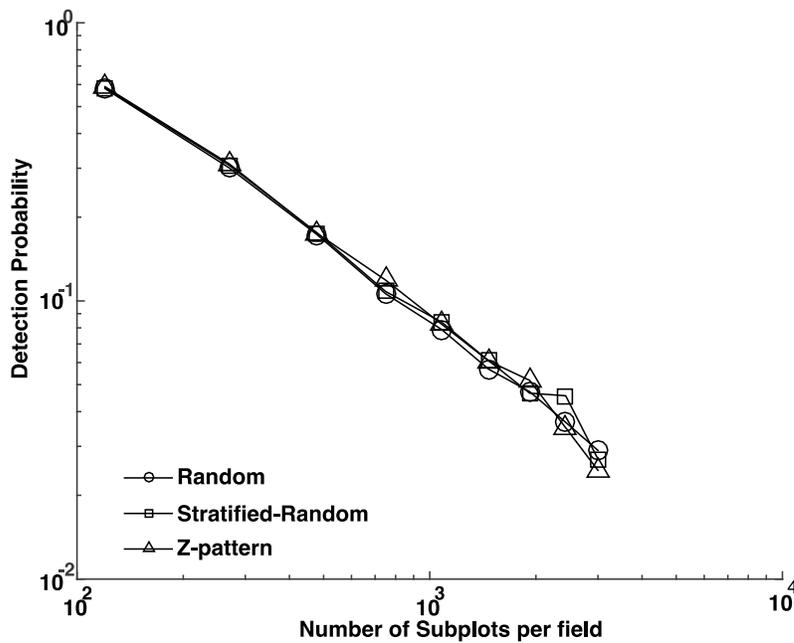
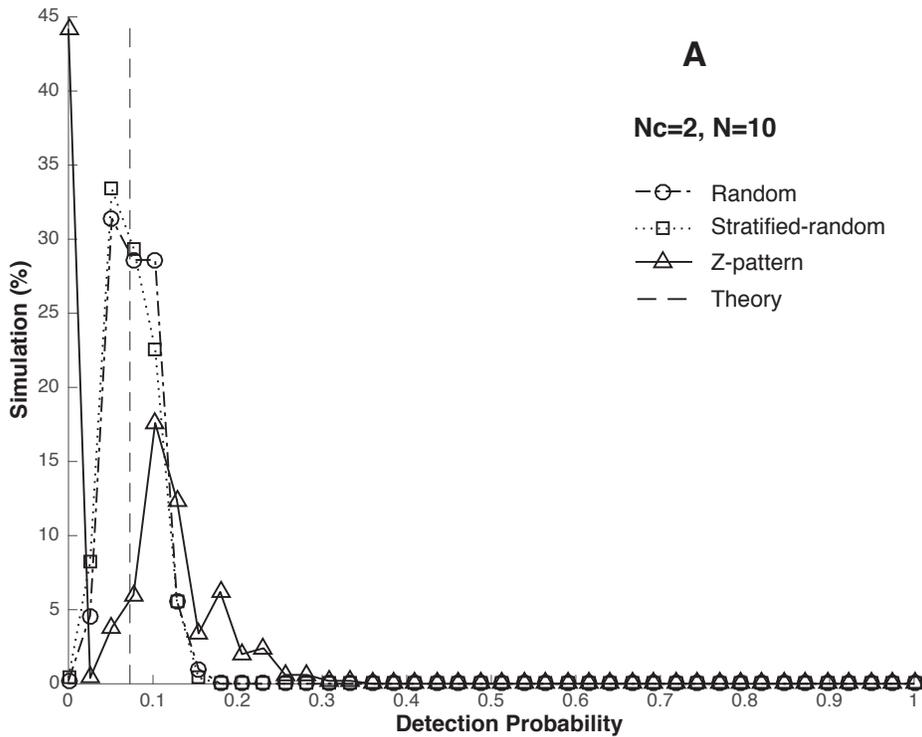
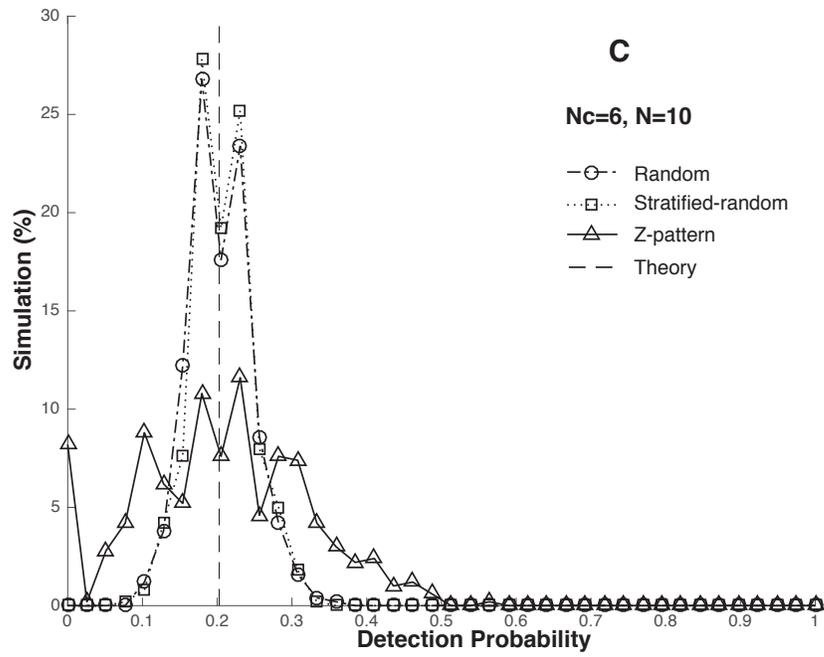
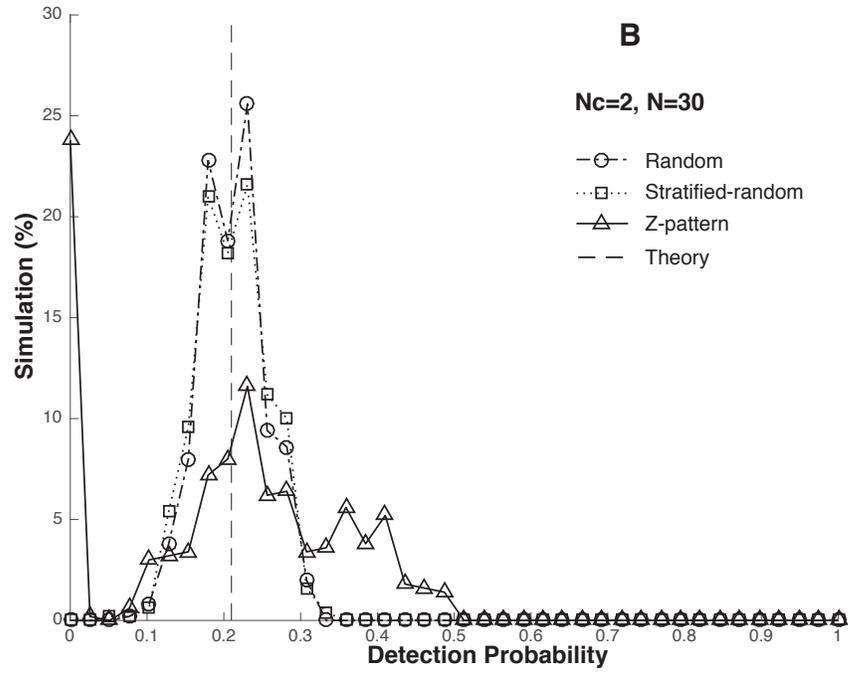


Figure 3.9 Log-log plot of mean detection probability as a function of total number of subplots per field for the three sampling plans.

As indicated above, while the mean detection probabilities among the three sampling plans were the same, differences in the standard deviations of the detection probabilities among the simulations were noted (see error bars in Figure 3.8). The distribution of mean detection probabilities across the 100 iterations per simulation was further evaluated as a function of contamination sites and number of samples taken. Figures 3.10 (A-D) provide examples of the impact of the number of contamination sites (2 vs. 6) and samples taken (10 vs. 30). The detection

probability of the Z-pattern sampling plan consistently had wider dispersion around the mean. This difference is most dramatic when the number of contamination sites was small, where the Z-pattern sampling plan had zero detection probability for a larger fraction of the simulations. This is due, in part, to the Z-pattern sampling plan only including a fraction of the field's subplots. If the contaminated sites lies in the Z-pattern, it would have better detection probability than the other sampling plans, but if the contamination site is not in a sampled plot, the sampling plan would fail to detect the contamination completely, regardless of sample numbers.





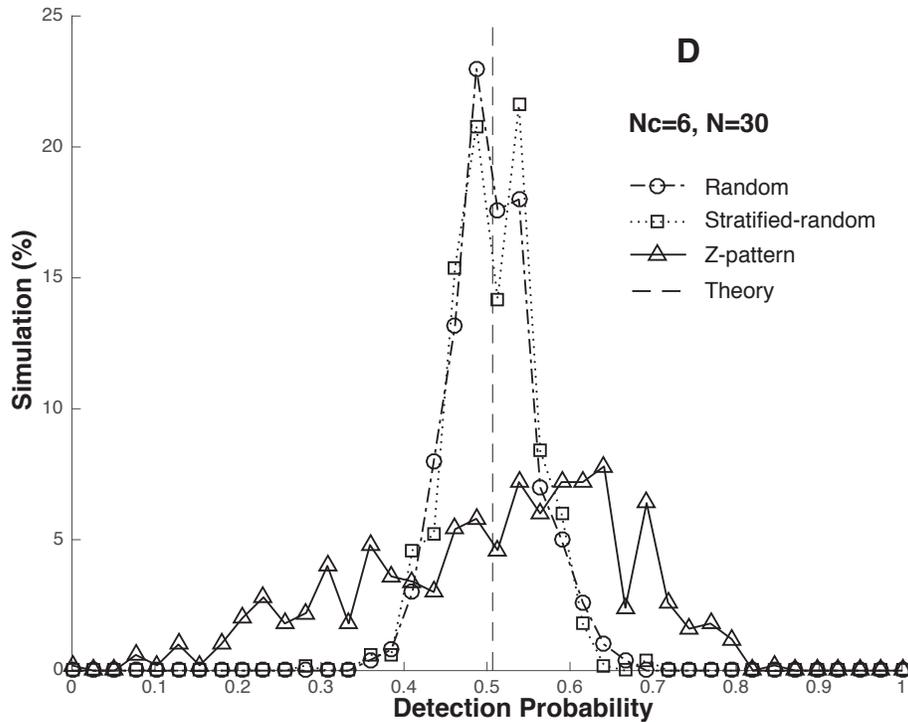


Figure 3.10 Distribution of detection probabilities for the three sampling plans for four combinations of contamination sites ($N_c=2$ or 6) and numbers of samples ($N_{\text{sample}}=10$ or 30). The vertical dashed line represents theoretical estimate of the mean detection probabilities.

A user friendly program was developed using Matlab that allows the dispersion of detection probabilities of the three sampling plans to be visually displayed (Fig. 3.11). The user enters values for “field size,” “subplot size,” “sample size (area),” “number of samples,” “sampling plan,” and estimated number of “contamination sites.” When initiated by the user (“run simulation button”), the program runs simulations similar to those used in the current study. This program is available upon request and will be posted on the UM Specialty Crop Research Initiative (SCRI) website.

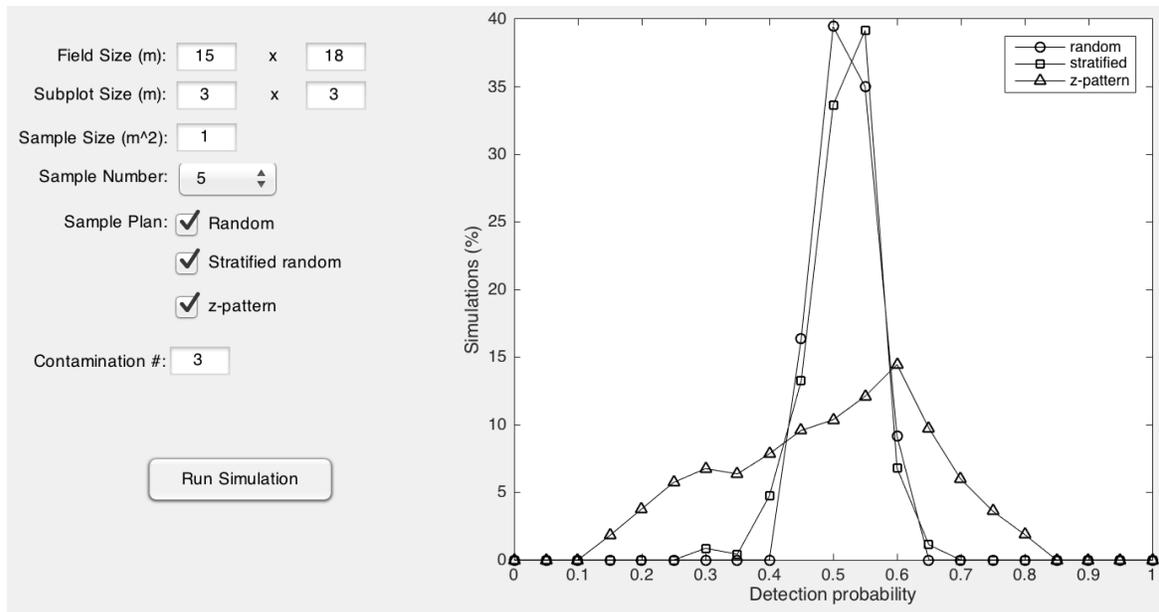


Figure 3.11 Graphical user interface of a Matlab application that allows comparisons of the three sampling plans. The user can configure the fields and specify the sampling plans on the left. The right side shows displays the distribution of detection probabilities for different sampling plans.

Simulation modeling was also used to consider when there are multiple “samplers” or when a field is sampled multiple times (assuming that new contamination sites have not been introduced between sampling times). The impact of multiple samplers or sampling times are then combined to calculate overall mean detection probability (i.e., the probability that at least one sampler detecting at least one positive sample). An example of the impact of multiple samplers (3-6) each taking 30 samples in conjunction with 1-6 contamination sites was evaluated via a single simulation each with 100 iterations (Fig. 3.12). As was anticipated, increasing the number of samplers, and thus the total number of samples, increased the mean detection probability. As noted before, increasing the number of contamination sites also increased the mean detection probability. The results with the random

and stratified random sampling plans were similar whereas the Z-pattern sampling plan had lower mean detection probabilities. The differential between the Z-pattern and the other sampling plans increased as a function of the number of samplers, particularly when the number of contamination sites was low.

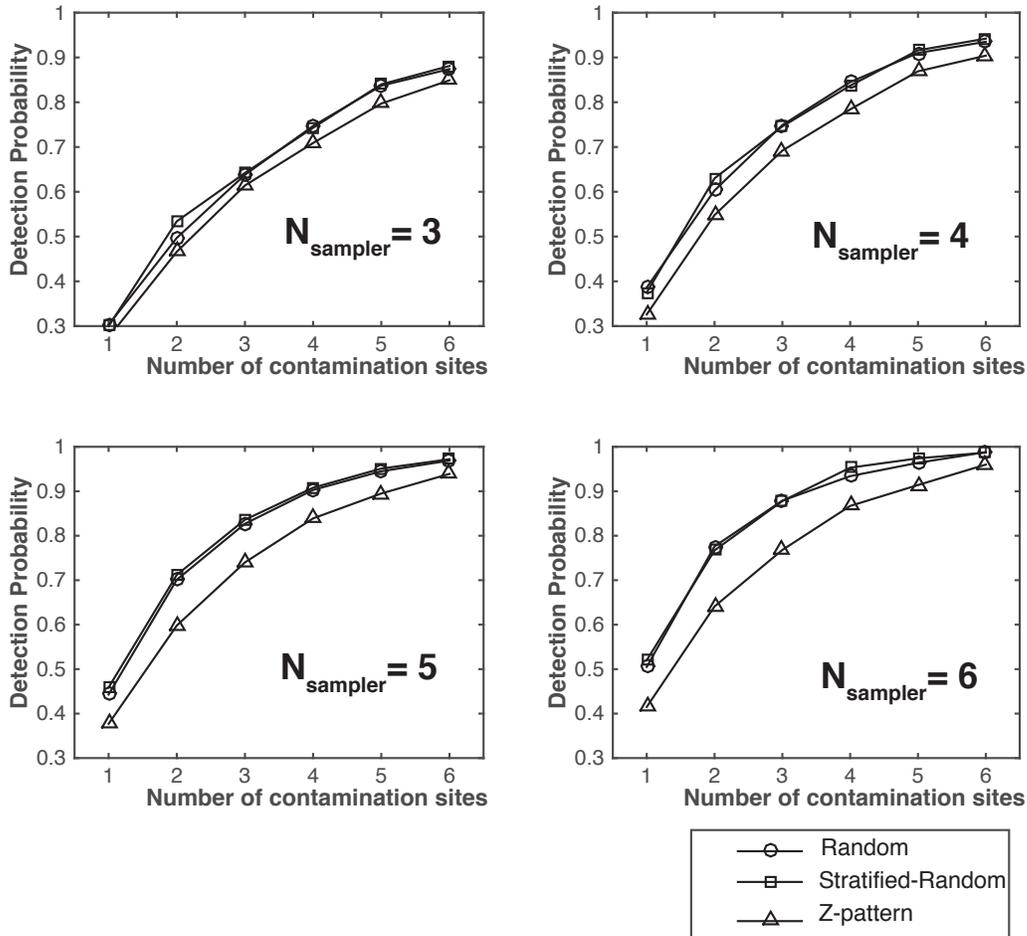


Figure 3.12 Mean detection probability of three sampling plans as a function of the number of contamination sites (x-axis) with number of samplers (N_{sampler}). The assumption is that each sampler collected samples independently. The detection probabilities were estimated using one simulation with 100 iterations for each sampler. The results for each sampler were combined and detection probability calculated by the number of iterations where at least one positive sample was indicated by any of the samplers.

3.4.2 Validation

The calculated numbers of positive samples for each sampling plan are depicted in Figure 3.13. The total numbers of positive samples for each sampling plan were similar when summed over all three experimental fields. In the first validation experiment: 9 for random sampling, 7 for stratified random sampling and 9 for Z-pattern sampling. There is considerable amount of variation on the number of positive samples across fields. The z-pattern sampling plan had highest variation (standard deviation (SD) of positive sample number = 2.6). Stratified sampling plan had lowest variation (SD of positive sample number = 1.2). Random sampling had the medium variation (SD of positive sample number = 2.0). In the second validation experiment, 4 for random sampling, 3 for stratified random sampling and 4 for Z-pattern sampling. Again, there was considerable amount of variation on the number of positive samples across fields. The Z-pattern sampling plan had highest variation (SD of positive sample number = 1.5). Random sampling plan had lowest variation (SD of positive sample number = 0.58). Stratified sampling had the medium variation (SD of positive sample number = 1.0). Z-pattern sampling has the highest variability in both experiments.

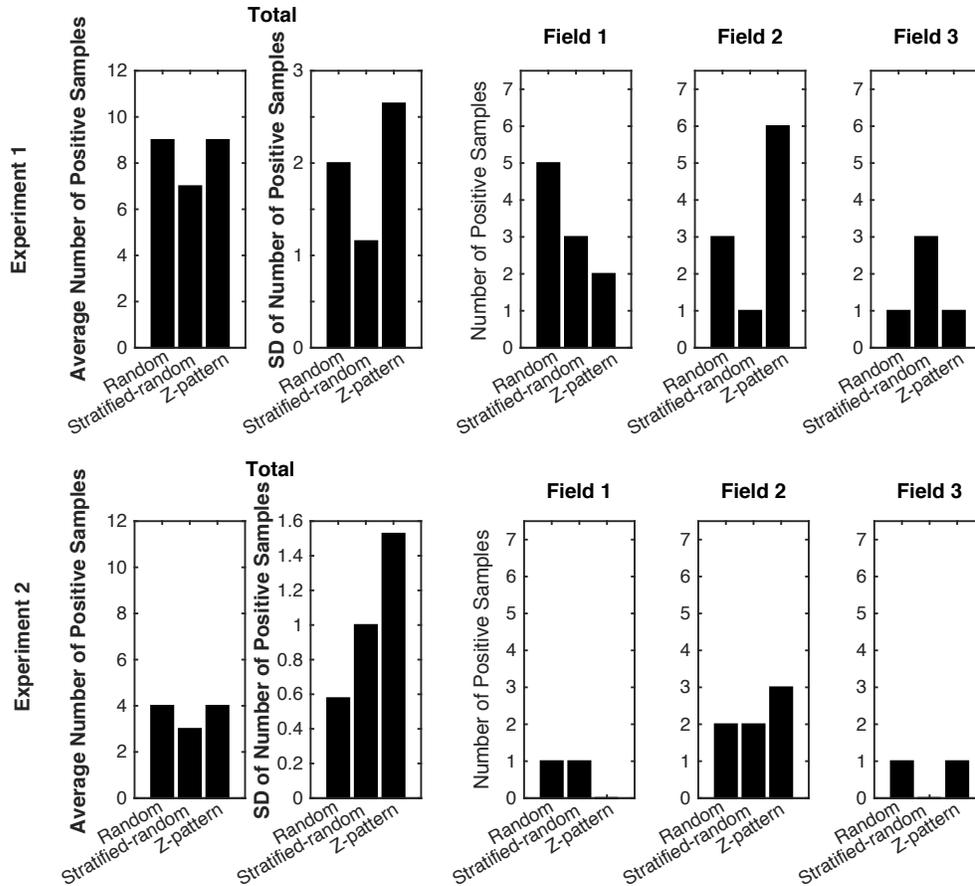


Figure 3.13 Number of positive samples and standard deviation of positive samples for different sampling plans in two trials.

The spatial relationship of the positive samples relative to the contaminated sites was also evaluated. The raw experimental data from the field experiments are shown in Fig. 3.14 and Fig. 3.15. For each field, there were 10 contamination sites (shown as black cross), and 30 samples were drawn according to each of the three sampling plans for each field. Samples drawn based on the random, stratified random, and z-pattern sampling plan were colored as red, green and blue respectively. The positive samples were labeled as solid diamonds, and the negative samples were labeled as empty diamonds.

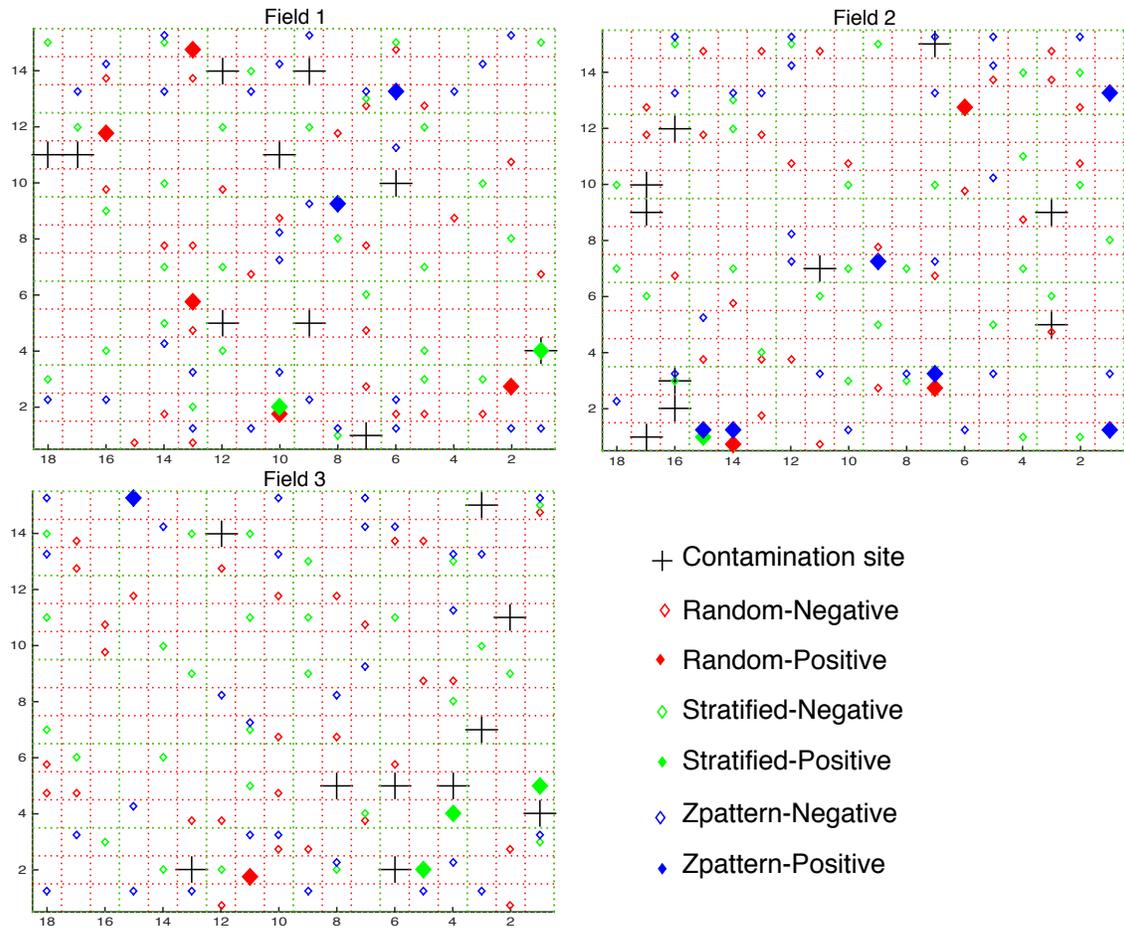
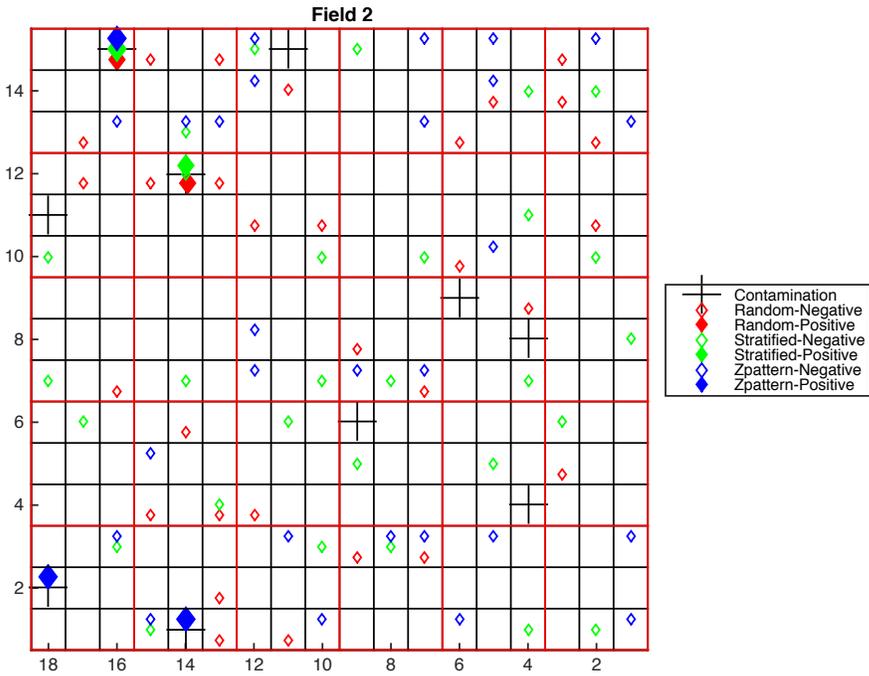
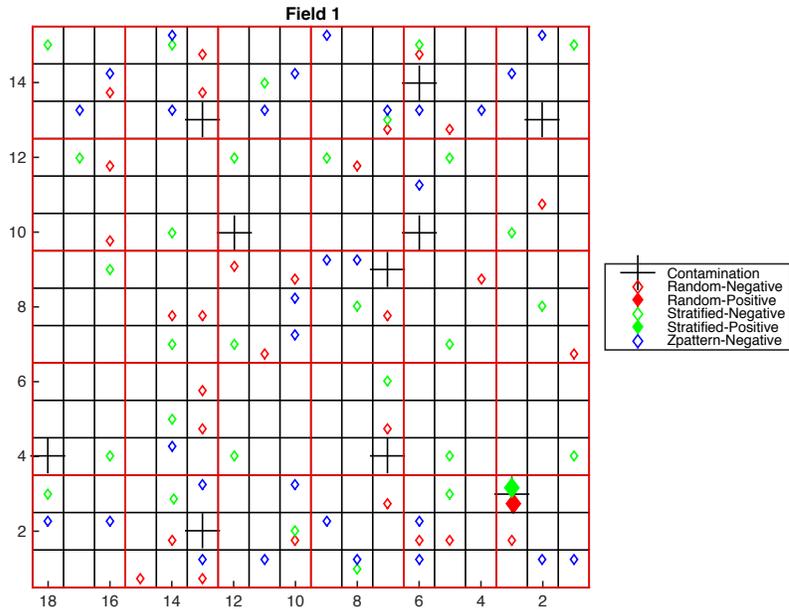


Figure 3.14 Experimental data from first field sampling based on GPS. (see text for details).



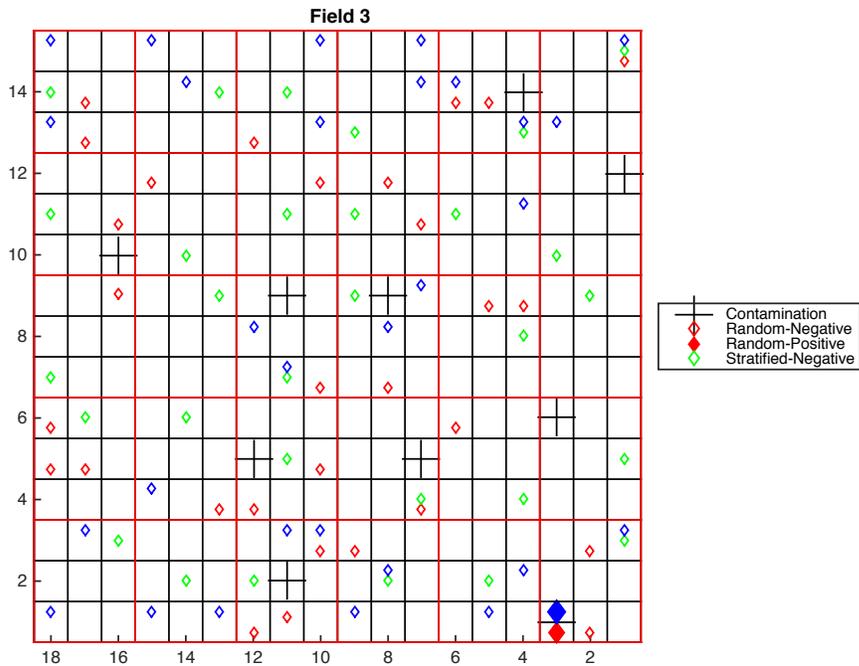


Figure 3.15 Experimental data from second field sampling based on post & string (see text for details).

There is no clear match between the contamination sites and the positive samples for first verification experiment, but an excellent match for the second experiment. The mismatch during the first verification study was likely due to the limited of spatial resolution of the GPS device we used for the experiment. However, when the distance between a sample and a contamination site were quantified, positive samples are closer to the contamination sites than the negative sample on average. The trend is consistent for all three of the experimental fields.

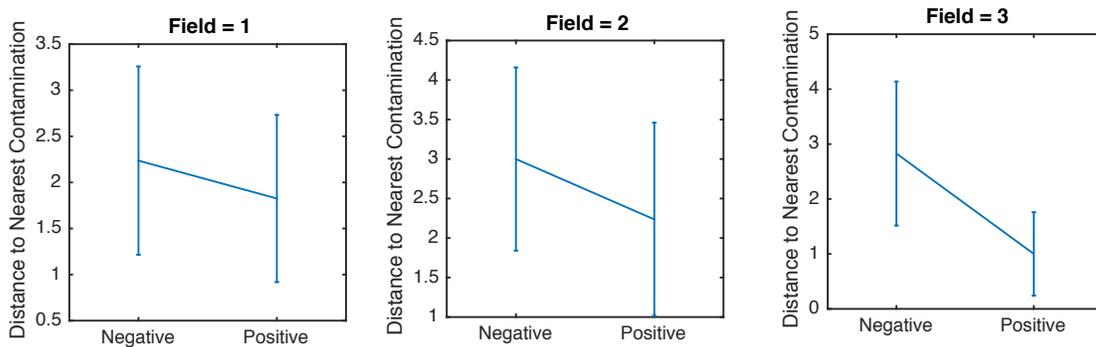


Figure 3.16 Distance to the nearest contamination sites for negative and positive samples.

3.5 Discussion

The goal of this study was to compare the effectiveness of three widely used sampling plans for the detection of bacteria in pre-harvest leafy green fields. Our results help quantify the impact that the number of contamination sites and the number of samples analyzed have on the likelihood of positively detecting a targeted pathogen or indicator microorganism. Further, the results demonstrate that basing decisions related to choice of sampling plan cannot rely on mean detection probability solely but also must address the uncertainty associated with those values. The use of relatively simple simulation models proved to be highly effective for estimating that uncertainty.

Quantifying the probabilities provides a more objective means of estimating the relative effectiveness sampling plans and clearly points out the limitations when small numbers of samples are used to evaluate fields that are sparsely contaminated.

Small numbers of contamination sites in combination with small number of samples ($N \leq 5$) typically resulted in mean detection probabilities below 0.2 (Fig. 3.7).

Similarly, the chance of detecting a target bacterium is very small when the sample size/area becomes increasingly small (Fig. 3.9). The use of simulation modeling provides new insights into the characterization of the performance of different sampling plans by allowing the variability and uncertainty of the sampling plans to be quantified along with the mean detection probability. This is dramatically demonstrated when current simulation modeling of Z-pattern sampling plan yielded the same mean detection probabilities, but the variability was effectively over and under estimating the degree of contamination, leading to more erratic results.

Some insect sampling [85] and soil sampling [86] projects found Z-pattern sampling to be more efficient for insect control and soil composition analyses. The obvious advantage of Z-pattern sampling is the increased ease of collection of samples. The total distance travelled with a Z-pattern sampling strategy would be typically shorter than either a random or stratified random sampling plan. This is undoubtedly one of the reasons for the adoption of Z-pattern sampling plans when there was an increased need to collect samples for pre-harvest assessment of leafy greens for microbiological contaminants. The current study clearly indicates that the mean detection probabilities of the three sampling plans are similar (Fig. 3.7, 3.8, 3.9). However, the current study also clearly establishes that the convenience of Z-pattern testing must be balanced against its increased variability, particularly if acquiring a small number of samples when the number of contamination sites is likely to be low. The results of the current evaluation clearly indicate that the

consistency of random or stratified random sampling plans is superior to Z-pattern plans.

Field validation experiments were also conducted in this study. It is very important to apply the computer simulations to the real world to approve the model really working. Since we only applied into 6 fields (6 iterations), compared to 10,000 iterations in the simulation, we cannot conclude that there is no difference in mean detection probability among three sampling plans. But the sum of positive samples of three sampling plans from 6 iterations is similar and the variance of Z-pattern tends to the biggest. To identify each subplot, a GPS was used in the first trial, due to the accuracy limitation of the GPS device, we could not get accurate location resolution of less than 1meter. This is a likely source of error in the initial validation trial. So, strings and stakes were used to identify the subplots and it performed very well.

The current study is based on the assumption that contamination is the result of a random event(s). However, there are a number of scenarios where non-random events could be root case of contamination of leafy green fields. In many instances, microbial contamination could be non-random. Studying the contamination of powdered infant powder, Jongenburger et al.[88] reported that stratified sampling is preferred over random sampling for detection of localized contamination. Similar results have also been reported by other investigators [83, 84]. However, a potential drawback of stratified sampling is the potential to miss a systematic, reoccurring contamination event [79]. In such instances sampling performance could be enhanced by moving to stratified random sampling. It is also worth noting the

performance of random and randomized stratified sampling plans are similar when the number of samples ≥ 30 [89], [90].

The current evaluation was based on the underlying assumption that the samplers have no *a priori* knowledge of potential sources of contamination. However, knowledge of the leafy green production environment and likely sources of contamination could potentially increase dramatically the statistical power of testing regimes [91]. It is feasible that assessment of a leafy green field by a sampler could take advantage of knowledge of potential contamination sources to improve the statistical basis for pre-harvest testing. For example, the presence of overhead wires, adjacent animal facilities, areas of periodic flooding, and prevailing winds are all factors that can lead to increased potential for non-homogeneous contamination of leafy green fields. This suggests that pre-harvest sampling could be enhanced by more systematically and quantitatively studying sources of contamination of pre-harvest produce to focus sampling on the conditions that are likely to increase contamination risks. This would allow use of hybrid sampling plans which combine systematic sampling with “samples of opportunity (SOO).” Such plans would involve collecting a portion of the samples via a standardized sampling plan, while a second portion of the samples would increase the extent of sampling in areas which the sample collector observed factors that are associated with an increased potential for contamination. A study comparing the relative effectiveness of sampling plans that combine a statistically based sampling plan with such with such “samples of opportunity” is provided in the next chapter.

3.6 Conclusions

This study has shown that the detection probabilities for random and stratified sampling plan are more stable than those for Z-pattern sampling. Similarly random and stratified random sampling plans performed more consistently than Z-pattern sampling plans during limited in-field validation trials. In the field experiments, the study also shown that Z-pattern sampling had higher variability than the other two samplings, which is consistent to the model result. The study demonstrates the utility of simulation modeling for evaluating the performance of different sampling plans used in pre-harvest testing.

Chapter 4: Evaluation of the potential for enhanced sampling effectiveness by assessment of field environments and consideration of likely sources of contamination:

Comparison of sampling plans based on random, stratified random, Z-pattern and “samples of opportunity” sampling

4.1 Abstract

Traditional sampling plans assume sample collectors have no knowledge related to the history or origins of a food, including information on potential contamination sources. Knowledge of factors that could lead to non-random contamination could potentially increase the effectiveness of pre-harvest sampling programs. The goal of this segment was to use mathematical modeling and field validation to determine the impact of including a portion of the samples based on the sampler’s knowledge of risk factors. The performance characteristics of sampling plans that include such “samples of opportunity” (SOO) were compared to that of traditional pre-harvest sampling plans. Computer simulations were

performed to compare the relative effectiveness of random, stratified-random, and Z-pattern vs. SOO sampling. The SOO sampling reserved two thirds of samples to be taken from identified high-risk areas within a field. These evaluations assumed the contamination in the field was non-random, with three contamination scenarios being evaluated: point contamination (animal house nearby), line contamination (power line above the field), and directional contamination (field partially exposed to floodwaters). The simulation modeling tool allowed a large number of field contamination scenarios to be generated and evaluated systemically. The initial scenarios used 6 contamination sites in each field, with 18 samples being subsequent taken by each sampling plan. The detection probability for a non-randomly contaminated pre-harvest field (5X6 plots with 9 subplots per plot (total of 270 subplots)) was 0.30 ± 0.11 , 0.32 ± 0.11 , 0.32 ± 0.17 for random, stratified-random, and z-pattern sampling plans, respectively, whereas the SOO sampling plan had a detection probability of 0.61 ± 0.25 . The mean detection probability of SOO was 96% higher than other sampling plans ($p < 0.001$). However, if the assumption of contamination source is incorrect, detection probability of SOO drops to 0.33 ± 0.23 , which is not significantly different than the other sampling plans. This study provides mathematical approach for evaluating the effectiveness four pre-harvest sampling plans, and suggest that utilizing knowledge of likely contamination sources in the field can be effectively incorporated into sampling plans to improve sampling effectiveness.

4.2 Introduction

Effective microbiology testing requires not only a sound understanding of

detection and identification techniques but also requires a firm grasp of the statistical principles that are the foundation of microbiological sampling. Traditional sampling plans are based on the underlying assumption that the sampler has no knowledge of the degree or location of contamination with a lot or field. A major impact of the assumption that contamination is randomly distributed is that where samples are taken in a field will not affect the performance of the sampling plan. However, if the contamination were not randomly distributed, then knowledge of the geographical distribution would be expected to substantially enhance the likelihood of detection. This would be particularly true when contamination in commercial fields is infrequent and at low levels.

Knowledge of the spatial distribution of contamination in a field could enhance sampling, testing and the food safety risk management decisions that are informed by the acquired data. However, typically there is little knowledge of how microorganisms are actually distributed in a field, hence the default assumption of a random or homogeneous distribution. However, produce safety research during the past decade has identified and characterized many of the risk factors that could lead to localized contamination, and the heterogeneous distribution of microbial hazards during primary production [92]. In such cases, samples collected from different parts of the field would have different probabilities for contamination and thus likelihood of detecting pathogens or indicatory microorganisms.

There are few studies on sampling plans for the detection of microbial food safety hazards in the field, but there have been some sampling plan studies for food safety concerns with other food products. Several model studies in recent years

have shown that sampling strategy becomes important when contamination is heterogeneously distributed [90 – 92]. It has been suggested that stratified sampling was more effective for detecting localized contamination, particularly when the number of samples is limited [96]. A model study with powdered infant formula demonstrated that stratified random sampling was better than random sampling when $n < 30$ [88]. However, no studies have researched how sampling plans could be enhanced for detecting pathogens in pre-harvest leafy greens, nor has the use of knowledge of sources of contamination been used to enhance sampling strategies.

As a means of addressing this deficiency, the current study was undertaken to consider three contamination scenarios that would lead to non-random distribution in a field setting. Computer simulation modeling were used to determine if targeting a percentage of samples in areas identified as posing increase likelihood of contamination would increase the frequency of detecting pathogens or indicator microorganisms of concern. Thus, the sampling plan consisted of a portion of the samples being targeted “samples of opportunity” consistent with the contamination scenario, with the remainder of the samples selected by random assignment of subplots to be sampled. These knowledge-informed hybrid sampling plans were compared against random, stratified random, and Z-pattern sampling with respect to their ability to detect a non-random contamination by a pathogen. The hypothetical scenarios considered were based on risk factors for pathogens such as *S. enterica* and *E. coli* that are commonly encountered during leafy greens cultivation [95, 96]. The overall objective was to gain quantitative insights into the potential

impact of harnessing expert knowledge to address the challenges to microbiological testing when dealing with non-random distributions of microbiological hazards associated with pre-harvest leafy greens.

4.3 Materials and methods

4.3.1 Computer simulation modeling

4.3.1.1 Model of non-random contamination factors

A field divided into $N_x \times N_y$ plots of equal sizes similar to the fields in Chapter 3 was used to develop a model for studying the sampling of non-random contamination. Each plot, in turn, was subdivided into $S_x \times S_y$ subplots. The total number of subplots is $N_{subplot} = N_x N_y S_x S_y$. Like chapter 3, a 5×6 plot array (30 plots), with each plot having 9 subplots ($N_x = 5, N_y = 6, S_x = 3, S_y = 3$) was used. Each subplot represents the smallest unit that would be sampled. The likelihood that a plot will be contaminated in the designated area from the source is assumed to be dependent on its spatial relationship with a contamination sources. Three types of contamination were evaluated: (1) stationary line contamination (e.g., birds roosting on a power line), (2) stationary point contamination (e.g., windblown manure dust coming from a chicken or animal house), and (3) directional line contamination (e.g., contamination due to flooding).

The likelihood that a plot will be contaminated in the designated area from the stationary line contamination source was modeled as:

$$I_{line}(x, y) = \exp\left\{-\frac{[\cos \theta (y - y_0) - \sin \theta (x - x_0)]^2}{2\sigma_{line}^2}\right\} \quad (4.1)$$

where θ is the orientation of the line, (x_0, y_0) is a point in the field that the line goes through, σ_{line} is the spatial spread of the contamination around the line. An example of line contamination is depicted in Fig. 4.1A.

The likelihood that a plot will be contaminated in the designated area from the stationary point contamination factor is modeled as a two dimensional Gaussian function

$$I_{point}(x, y) = \exp\left\{-\left[\frac{(x - x_0)^2}{2\sigma_{point}^2} + \frac{(y - y_0)^2}{2\sigma_{point}^2}\right]\right\} \quad (4.2)$$

where (x_0, y_0) is the location of the contamination source, σ_{point} controls the spatial spread of the stationary point contamination. An example a stationary point contamination is shown in Fig. 4.2B.

The contamination likelihood of the directional line contamination factor is modeled as a planar equation,

$$I_{directional}(x, y) = \exp\left\{k \left[\cos \theta \frac{(x - x_0)}{N_x} + \sin \theta \frac{(y - y_0)}{N_y} - 1 \right]\right\} \quad (4.3)$$

where θ is the direction of contamination, (x_0, y_0) is the corner of the field that is closest to the contamination source, k controls how fast the contamination decays along the direction of contamination. An example of directional line contamination is shown in Fig. 4.3C.

The overall contamination likelihood is the sum of the three types of contamination likelihood.

$$I(x, y) = I_{line}(x, y) + I_{point}(x, y) + I_{directional}(x, y) \quad (4.4)$$

The contamination probability of a plot is proportional to the risk factor,

$$P(x, y) = \frac{I(x, y)}{\sum_{(x,y)} I(x, y)} \times N_c \quad (4.5)$$

where N_c is the number of contaminations in the field. In the simulations we generate contamination sites without replacement to avoid contaminating the same plot multiple times.

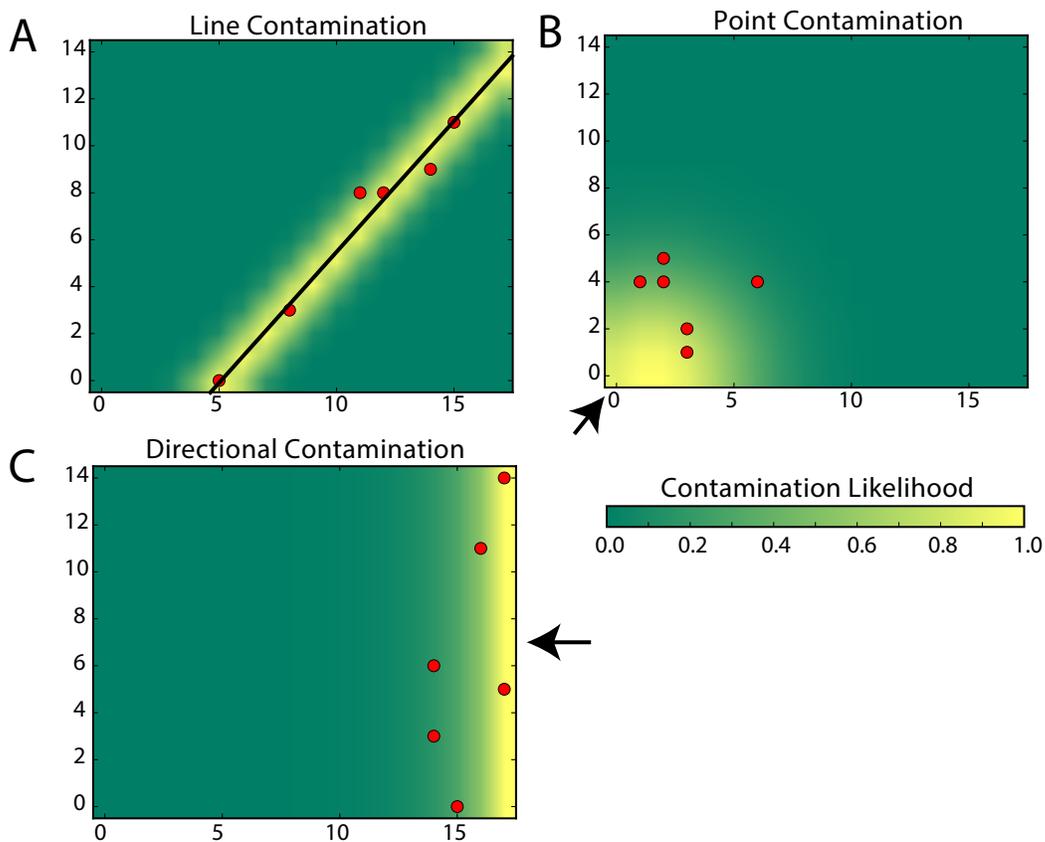


Figure 4.1 Three types of non-random contamination in 18m (x axes) by 15m (y axes) fields. (A). Stationary line contamination. (B). Stationary point contamination. (C). Directional line contamination. The color represents relative contamination likelihood. The red dots represent contamination sites. The direction of

contamination deposition from their sources is highlighted with black lines or arrows. The color depiction represents relative likelihood of contamination of the field from the contamination source.

4.3.1.2 Generation of simulated contaminated fields.

A large number of fields with non-random contamination are generated to systemically evaluate different sampling plans. Each simulated field has 18×15 plots, each a 1.0 m^2 (the same as used in Chapter 3). The default parameters for the equations are provided in Table 4.1.

For each simulated field with stationary line contamination, the line orientation θ was randomly selected in the range of $[0, 180^\circ]$. A point was randomly selected in the center part ($4 \leq x \leq 15, 4 \leq y \leq 12$) of the field as (x_0, y_0) to make sure that the stationary line contamination goes through the center of the field. The spatial spread of the stationary line contamination σ_{line} was chosen to be 1.

For each simulated field contamination from a stationary point source, a randomly selected location on the boundary of the field was selected as the point location (x_0, y_0) . The spatial spread of the stationary point contamination was assumed to be $\sigma_{point} = 3$.

For each simulated field with directional line contamination, the direction of contamination was selected among $[0^\circ, 90^\circ, 180^\circ, 270^\circ]$ with equal probabilities, and (x_0, y_0) is selected to be $[(0, 0), (N_x, 0), (N_x, N_y), (0, N_y)]$, respectively. The parameter k is set to be 10 for directional line contaminations.

Table 4.1 Summary of model parameters

Parameter	Description	Default value
N_x	Number of plots along the x-axis	18
N_y	Number of plots along the y-axis	15
σ_{line}	Spatial spread of stationary line contamination	1
σ_{point}	Spatial spread of stationary point contamination	3
k	Decay rate of directional line contamination	10
N_{sample}	Number of samples	18
N_c	Number of contaminated plots in the field	6

4.3.1.3 Sampling plans

We considered four sampling plans in this study: (1) random, (2) stratified random, (3) Z-pattern, and (4) SOO. The same number of distinct samples was collected each of the different plans. The number of samples per each sampling plan was designated as N_{sample} . An example of samples drawn according to different sampling plans is shown in Fig. 4.3.

The random sample plan “collected” N_{sample} samples randomly without replacement from all plots (see Chapter 3). The stratified random sample plan first divides the field into 5×6 primary plots, with each primary plot containing 3×3 subplots (see Chapter 3). Since we used $N_{sample} \leq 30$ for all simulations, there was

at most one sample collected from each primary plot. If $N_{sample}=30$ there is exactly one sample collected from each primary plot.

The Z-pattern sampling plan divided the field into 5×6 primary plots similar to the stratified sampling plan. Unlike the stratified sampling plan, the Z-pattern model only collects samples from plots lying at two-opposing edges and one of the diagonal lines (see Chapter 3). The "Z" region consist of 16 primary plots for a total of 144 plots. Note that Z-pattern sampling plan never collects any sample from the other 126 plots.

The SOO plan first identifies a potential contamination source, and then divides the field into high risk and low risk areas based on spatial layout of the designated contamination source. It then collects two thirds of the samples from the high risk area and one third of the samples from the low risk. In the simulation, it assumes the SOO plan identifies the type of contamination. For stationary point contamination, high risk area contains plots that are less than 5 m away from the contamination source. For stationary line contamination, the high risk area contains plots that are < 2 m away from the line. For directional line contamination, the high risk area was plots that are < 4 m from the contamination leading edge.

In one series of simulations, scenarios where sampler incorrectly identifies a high risk area for consideration by a SOO sampling plan. In those scenarios the actual high risk area was first identified using the actual locations of the contamination. The high risk area of the SOO plan was then chosen to have a variable amount of overlap with the real high risk area. Specifically, a random subset

of the real high risk area was included in the designated SOO high risk area. The rest of the SOO designated high risk area is chosen from the low risk portions of the field.

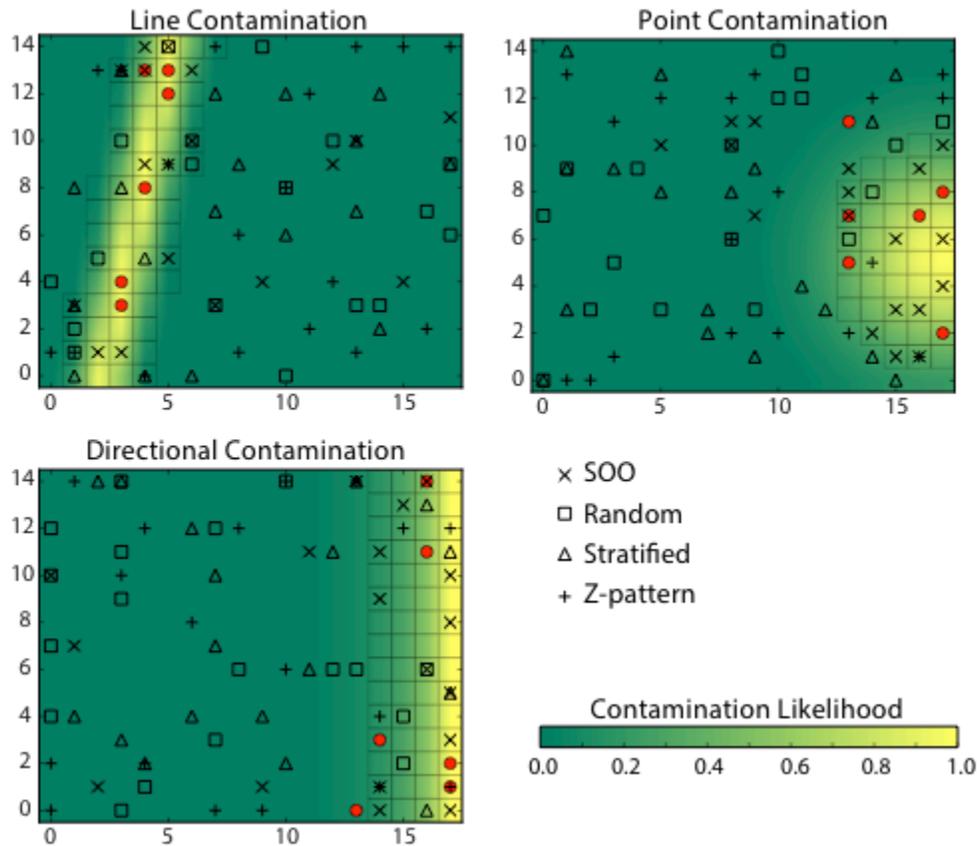


Figure 4.2 Example samples drawn according to different sampling plans. The high risk areas identified by the SOO plan are shown with squares that has gray boundaries. In this case we assume SOO correctly identifies contamination sources.

4.3.1.4 Evaluation of sampling plans

The four sampling plans were compared using computer simulations. At the beginning of each simulation, a contaminated field was generated with one of the three contamination sources (Eq. 4.1-4.3). Given the simulated contaminated likelihood, N_c plots were selected as contaminated plots based on (Eq. 4.5). The field was then sampled using the four types of sampling plans, each with 100 iterations.

For each iteration, the sample locations were generated according to different sampling plans independently. A detection probability was estimated based on the number of iterations that a sampling plan successfully detected at least one of the contaminated sites. The simulation was then repeated 100 times (total of 10,000 iterations), each with a different contaminated field, to estimate the distribution of detection probabilities for each sampling plan. As noted in Chapter 3, it is important to repeat the simulation multiple times since sufficiently characterize the sampling plans since individual sampling plans may be good at detecting certain contamination patterns but not others.

4.3.2 Validation field experiment

A limited validation study was performed for each of the three contamination scenarios (i.e., line contamination (overhead power line), point source (adjacent Alpacas facility), and directional (flooding)). The validation studies were limited to consideration of SOO compared to random and stratified random sampling plans. Z-pattern sampling was not evaluated during these validation trials.

Field Site and Plot Design

The field experiments were conducted at the University of Maryland Central Maryland Research & Education Center (CMREC), Clarksville, MD and on a commercial farm in Maryland. Three fields were included: a field close to an animal facility (commercial farm), a field with an overhead power line (commercial farm) and a field that was purposefully flooded (CMREC). The experiments at each field were conducted three times during the summer of 2016 and winter and early spring of 2017.

The field used for the flooding trials was a plot (+5% grade). Prior to flooding, a “U” shape soil berm was hand-built at the bottom end to maintain the floodwater (Figure 4.3). Approximately 2000 gallons of the lagoon water used for the treatment of bovine waste was pumped to the bottom of the field. The edge of this water was marked as the edge of the flood. The *E. coli* and total coliform concentration in the lagoon water was ~ 2 log CFU/mL. The total aerobic bacteria concentration was ~ 7 log CFU/mL.



Figure 4.3 Example of soil berm built to contain floodwater.

A rectangular leafy green field was divided into 5×6 primary plots. Each plot, in turn, was subdivided into 3×3 1.0 m^2 subplots. The total number of subplots was 270.

4.3.2.1 Sampling in the field

A total of 18 samples for each of the three sampling plans were collected. The sampling sites were selected using the random number generator. For random

sampling plan, eighteen subplots were randomly sampled from 270 subplots. For stratified random sampling plan, the field was divided into 18 big plots with fifteen 1m² subplots. Then, one sample was randomly taken from each big plot. For samples of opportunity sampling plan, the research team performed a rapid assessment of the field based on a standardized set of criteria that could be contributing to contamination (overhead power line, proximity to an animal facility, partially flooded field). After the assessment, the sampling team took samplings by their judgments. One-third of the 18 samples were taken in a random manner and the remaining 2/3 samples were collected from areas within the field where were recognized as the risky area by the investigator. These locations were recorded on the field grid map. The location and nature of these factors were identified on the grid map.

4.3.2.2 Sample processing

At each of the 18 locations for the three sampling plans, a soil sample of approximately 50 g was collected, transferred to a sample bag, placed in a cooler, and then transported to the laboratory. Samples were processed within 24 h of collection. For each sample, 10g of soil were weighed from each sample into a sterile Whirlpak bag (Nasco, Jackson,WI). 40 ml 0.1% peptone water (PW) were added to each bag and then shaken using Orbital shaker(Forma Scientific, Marietta, OH) for 2 min at room temperature. Serial 10-fold dilutions were prepared in 0.1% PW. The levels of total coliforms and *E. coli* in all samples were enumerated by duplicate 1-ml samples of appropriate dilutions plated onto 3M™ Petrifilm™ *E. coli*/coliform count plates (Cat No. 6414) and incubated at 37±0.5°C and 44±0.5°C,

as per manufacturer's instructions. Red colonies with gas bubbles observed at 24 h were counted as coliforms and blue colonies with gas bubbles observed at 48 h were counted as *E. coli* colonies, according to standard TC/*E. coli* Petrifilm enumeration methods. Appropriate dilutions were also plated directly onto 3M™ Petrifilm™ Aerobic Count Plates (APC) (Cat No. 6406), incubated at 37±0.5°C, and observed for red colonies after 48 h for enumeration of aerobic mesophilic bacteria [99].

4.3.2.3 Data analysis

Various statistical tests to test the difference among sampling plans. First, the fraction of positive samples for each type of indicator microorganism was assessed for significant differences among sampling plans using Fisher's exact test[100]. This is performed for each individual experiment. Second, analysis of variance (ANOVA) was used to determine if there were significant differences in the counts of indicator microorganisms (log cfu/g) for the three sampling plans. This was also performed separately for each individual plan. Third, all three experimental data sets were used to test whether three sampling plans performed different for detecting indicator microorganisms. Since the microbial counts could have also been dependent on the season, Repeated Measures ANOVA was used to adjust for seasonality[101]. Finally, the positive and negative samples from each sampling plan was plotted on a two-dimensional map to visualize the spatial distribution of positive samples. The data analysis was performed in Matlab R2015b (The Mathworks).

4.4 Results

4.4.1 Simulation

The efficiency of different sampling plans was evaluated by computer simulations. The performance of each sampling plan was characterized by the probability of detecting at least one contamination site in the simulated field. We considered three types of non-random contamination: stationary point contamination, directional line contamination and stationary line contamination. The mean detection probability of four sampling plans with these three types of contamination is shown in Fig. 4.4. Stratified-random sampling plan was a little better than random sampling plans ($p=0.045$). The SOO plan had the best detection probability for all types of contamination ($p<0.01$). This is because most of the samples collected in the SOO plan are from regions with higher contamination probability. The Z-pattern sampling plan had significantly better mean detection probability than random and stratified sampling plan for stationary point and directional line contamination ($p<0.01$), but was significantly worse sampling plan for stationary line contamination ($p<0.01$). This is because the stationary point contamination and directional line contamination have highest contamination probability near the boundaries of the field, whereas stationary line contamination tends to have most contamination sites near the center of the field (see section 4.3.1.2).

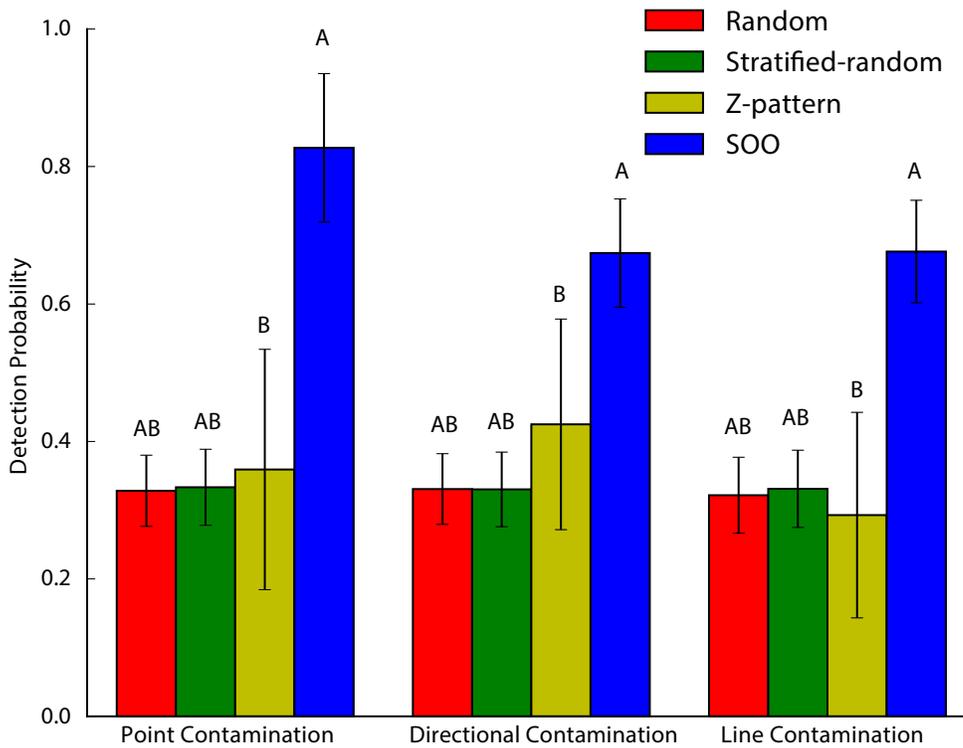


Figure 4.4 Detection Probability of four sampling plans: random (red), stratified random (green), z-pattern (yellow) and sampling of opportunity plan (blue). Values from bars with different letters are significantly different based on an ANOVA and Tukey’s post hoc test ($p < 0.01$). Error bars represent standard deviation of detection probability across 500 simulated contaminated fields.

Whether the performance of sampling plans depends on number of samples was analyzed. The mean detection probability was calculated with 5, 10, 18 or 30 samples. The mean detection rate for all sampling plans increased as a function of number of samples (Fig. 4.5). The relative effectiveness across sampling plans remained the same regardless of the number of samples.

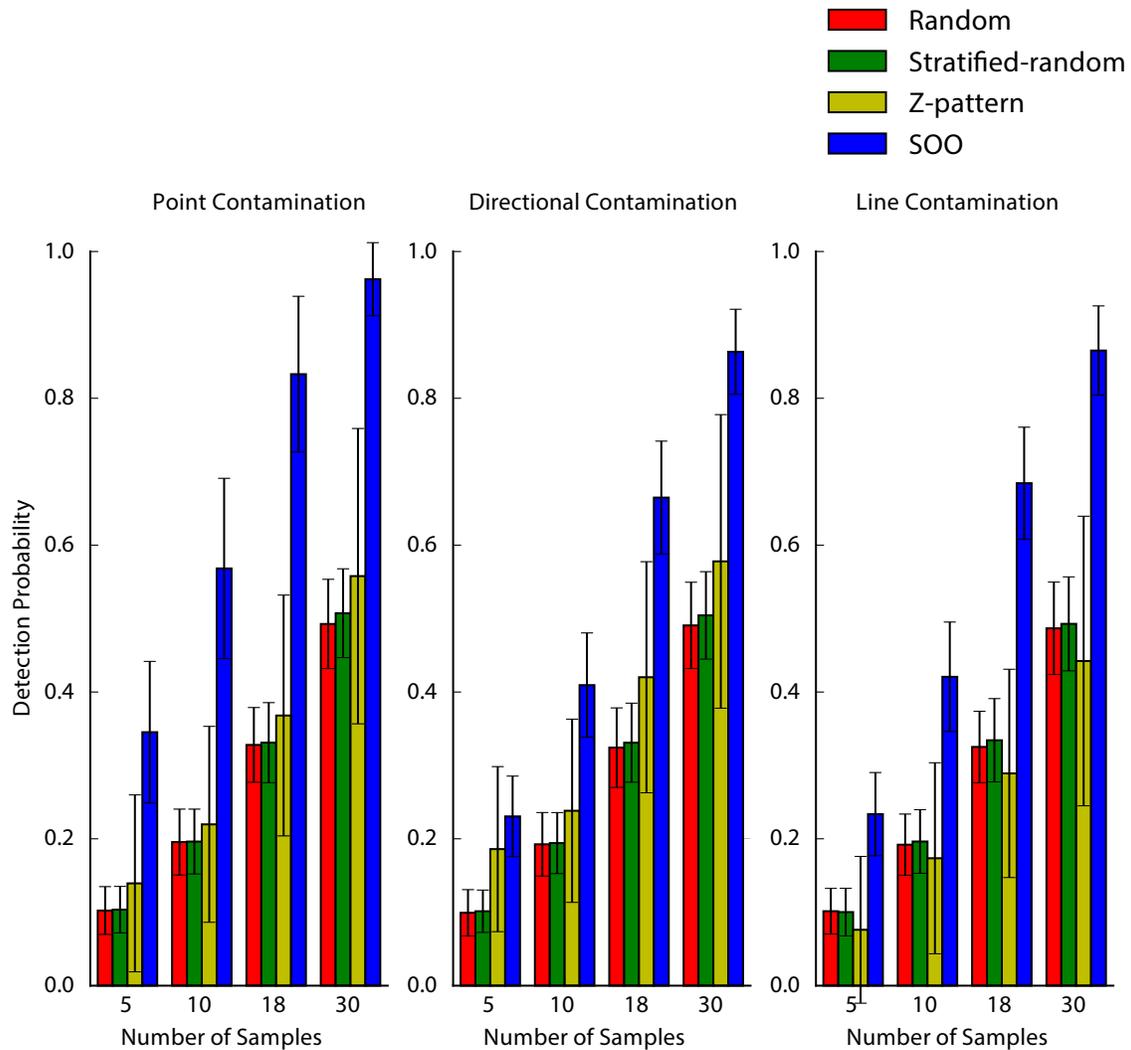


Figure 4.5 Detection probability of different sampling plans as a function of number of samples. The relative effectiveness of different sampling plans is the same as in Fig. 4.4.

The effect of number of contamination sites on the performance of the sampling plans was evaluated (Fig. 4.6). The mean detection probability for all sampling plans increased as a function of number of contaminations. The relative effectiveness across sampling plans remained the same regardless of the number of samples. The difference between SOO and other sampling plans is most dramatic when the number of contamination sites was small. The SOO yielded a two-fold to

three-fold improvement in mean detection probability when the number of contamination sites was three. This ratio of improvement decrease as the number of contamination sites increase above 3.

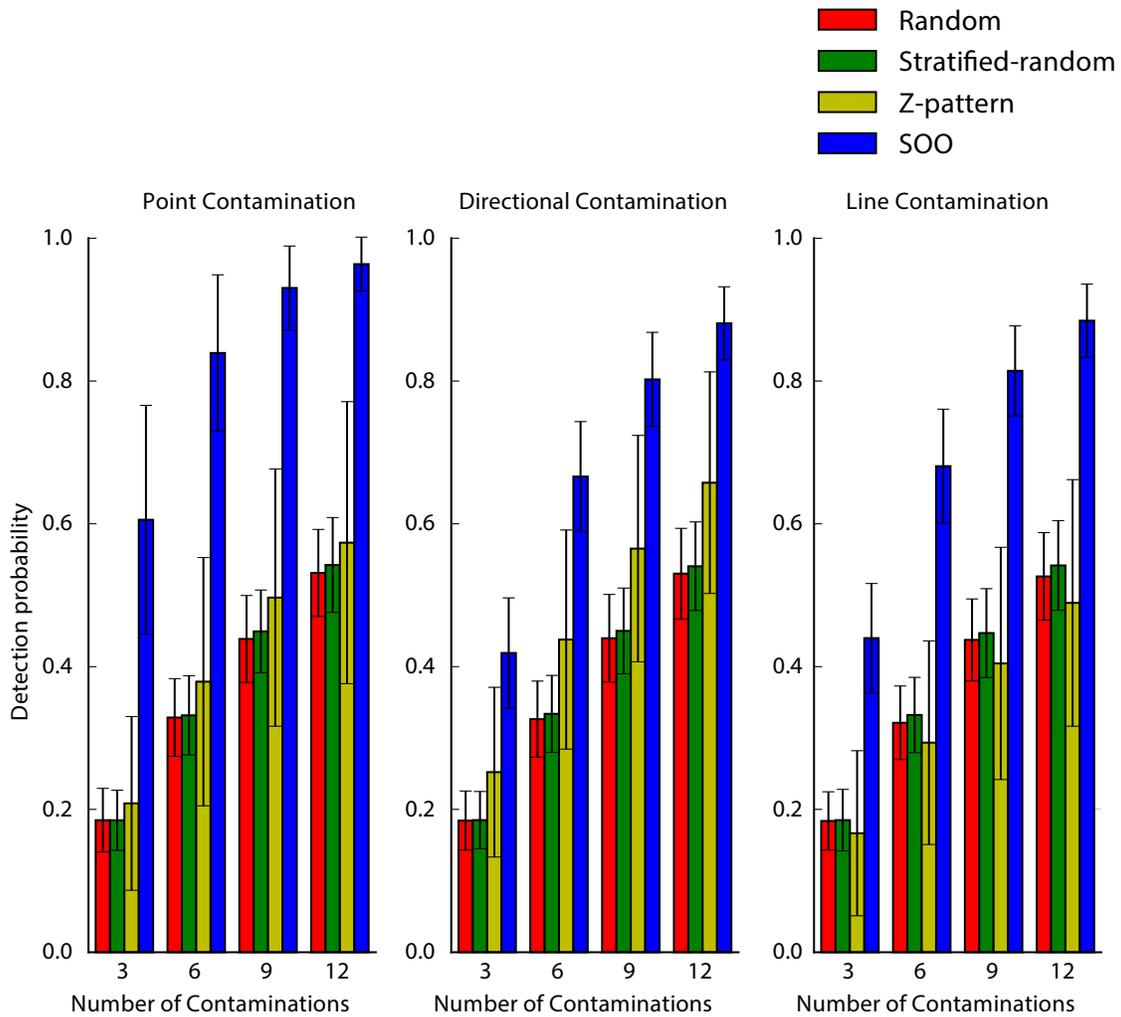


Figure 4.6 Detection probability as a function of number of contaminations in the field.

During the initial evaluation of SOO sampling plans, the focus was on scenarios where the SOO plan correctly identified the high-risk areas of the simulated field. In practice, it may not be possible to perfectly identify the risk

factors in the field. The sensitivity of SOO sampling plans to the correctness of the high-risk areas subsequently evaluated by simulation modeling. Also, in the simulations above, the SOO plan collected two thirds of the samples from the high risk area. Accordingly, the performance of SOO sampling plans were evaluated for the impact of the percentage of samples collected from the high risk area.

In this simulation, a fraction of SOO high risk plots were replaced with plots that fell in the low-risk area. The fraction of plots that remain in the true high-risk area was denoted by an overlap parameter: $\text{overlap}=1.0$ means the no high-risk plots are changed in the SOO plan, which corresponds to perfect identification of high-risk areas, whereas an $\text{overlap} = 0.0$ means all the high-risk plots were incorrectly identified, i.e., the SOO evaluation prior to sample site selection completely failed to detect a high risk area. The number of samples collected from the high risk area was also varied. The mean detection probability was plotted as a function of number of samples collected in the high risk area is presented in Fig. 4.7, with each colored line representing simulations where a fraction of high-risk plots were replaced.

When the number of samples collected from the high-risk area is high, the mean detection probability of the SOO plan decreases when more high-risk plots are replaced with low risk plots. However, the SOO plans still had better mean detection probability than random sampling plan (Figure 4.7 black dashed line) as long as more than 20% of the high-risk plots were correctly identified. In this case, the mean detection probability increased when a higher portion of the samples were collected from the high-risk area. When the SOO plan failed to identify the high-risk

area (overlap <20%), the detection probability decreased when more samples are collected from a presumed high-risk area (Figure 4.7 dark blue lines).

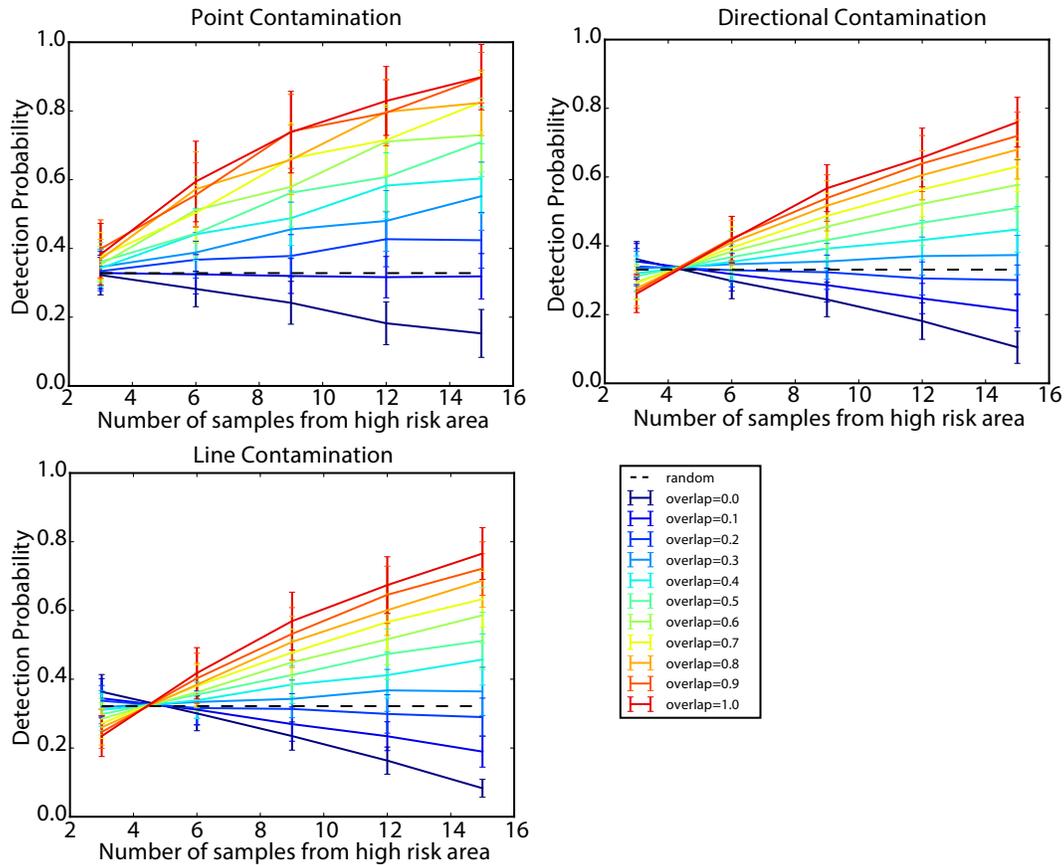


Figure 4.7 Detection probability of the SOO sampling plan while the high risk area cannot be accurately determined. Each colored line represent simulations when a fraction of the plots in the high risk area are randomly replaced with plots that lie outside of the high risk area. The overlap parameter denotes fraction of the plots that lies in the high risk area. The detection probability is plotted as a function of number of samples collected from the high risk area in the SOO plan (total number of samples=18).

4.4.2 Validation

The levels of five indicators microorganisms were tested: total aerobic bacteria (APC), total coliforms (TC 37°C)(TC37), thermotolerant coliforms (TC 44°C)(TC44), generic *E. coli* (*E. coli* 37°C) (EC37) and thermotolerant *E. coli* (*E.coli* 44°C)(EC44). The three sampling plans compare were samples of opportunity sampling, stratified random sampling and random sampling. The fraction of positive samples detected out of 18 samples and average levels of five indicators were calculated.

4.4.2.1 Flooded field

For the flooded field, in the first trial, the fraction of positive samples was significantly greater for the SOO sampling plan than the other two sampling plans based on the EC37 and EC44 levels. The fraction of positive samples based on the SOO sampling plan was significantly greater than stratified sampling for the TC44 testing ($p < 0.05$). The log number of average bacterial count, for SOO sampling was significantly higher than those for the other two sampling plans based for the two *E.coli* assays ($p < 0.01$), APC ($p < 0.01$) and TC44 ($p < 0.01$) (see Fig. 4.8).

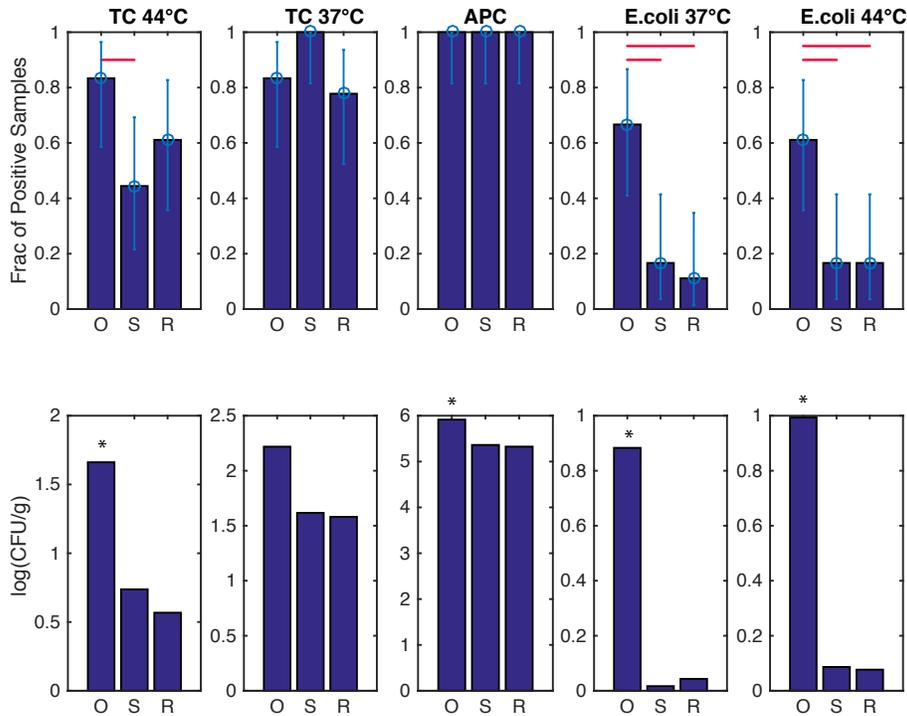


Figure 4.8 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in flooded field from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the first trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$).

For the flooded field, in the second trial, the fraction of positive samples was significantly greater for the SOO sampling plan than the other two sampling plans based on the EC37 and EC44 levels ($p < 0.05$). The log number of average bacterial count, for SOO sampling was significantly higher than those for the other two sampling plans based for all five assays. (see Fig. 4.9).

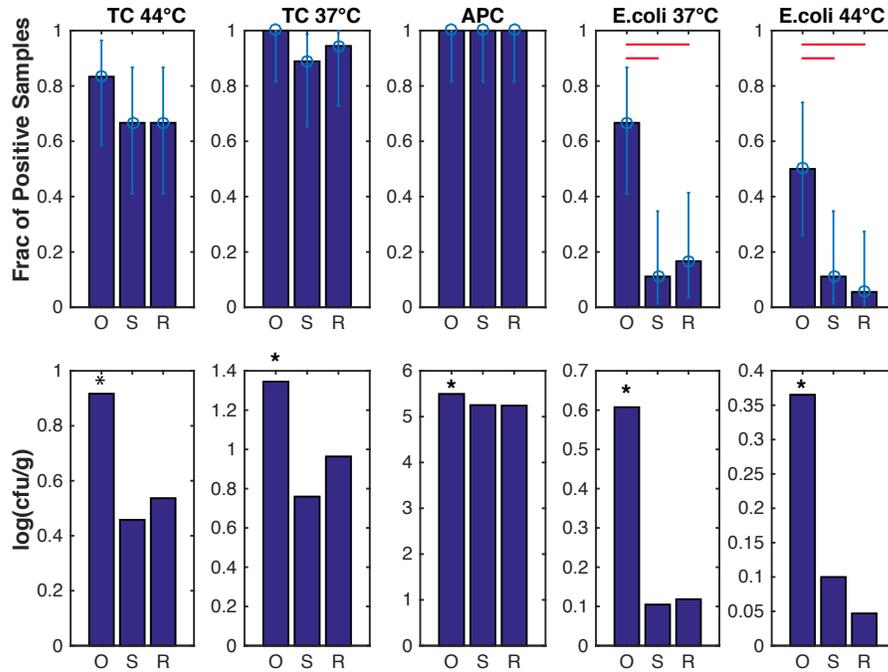


Figure 4.9 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in flooded field from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the second trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$).

For the flooded field, in the third trial, the fraction of positive samples was significantly greater for the SOO sampling plan than random sampling plan based on the EC37 ($p < 0.05$). The fraction of positive samples based on the SOO sampling plan was significantly greater than stratified sampling for the EC44 testing ($p < 0.05$). The log number of average bacterial count, for SOO sampling was significantly higher than those for the other two sampling plans based for the EC44 assay ($p < 0.01$) and APC ($p < 0.01$) (see Fig. 4.10).

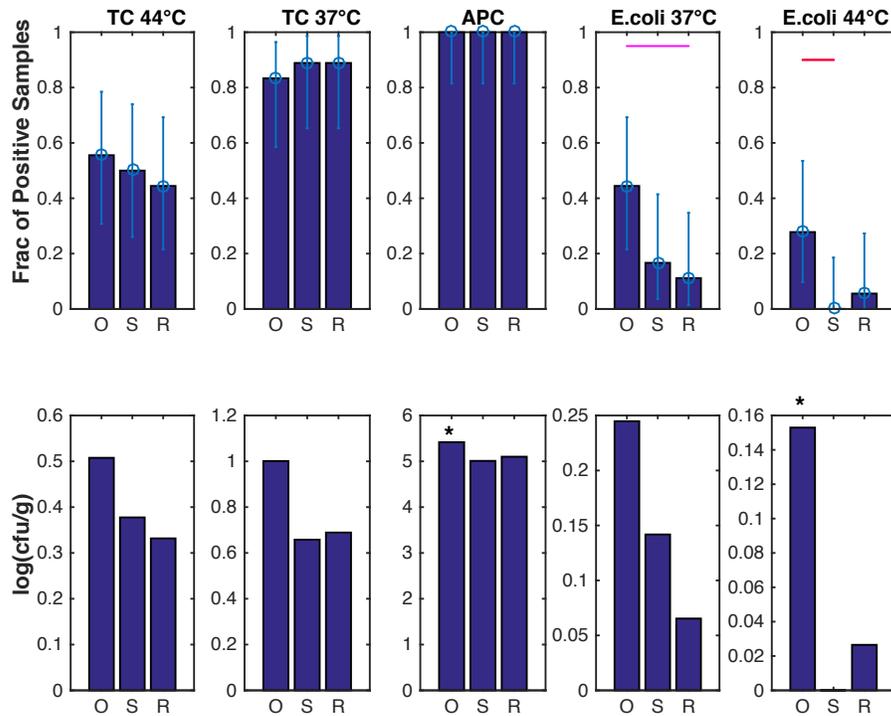


Figure 4.10 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in flooded field from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the third trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$).

4.4.2.2 Field with animal house nearby

For the field close to an animal facility, in the first trial, the fraction of positive samples from SOO sampling was significantly higher than stratified sampling based EC37 testing ($p < 0.05$). For the average bacterial count, SOO sampling had significantly higher counts than other two sampling plans for EC37 and EC44 assays ($p < 0.01$), and the TC37 and TC44 assays ($p < 0.01$) (Fig. 4.11).

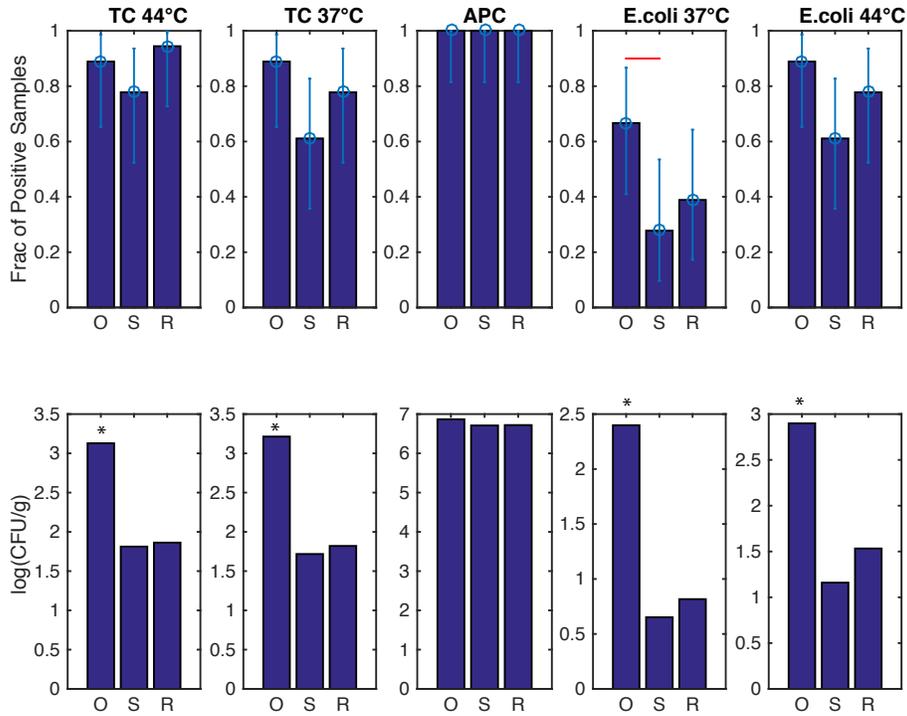


Figure 4.11 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with animal house nearby from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the first trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$).

For the field close to an animal facility, in the second trial, the fraction of positive samples from SOO sampling was significantly higher than stratified sampling based TC37 testing ($p < 0.05$). No *E.coli* were detected. For the average bacterial count, SOO sampling had significantly higher counts than other two sampling plans for TC37 and APC assays ($p < 0.01$) (Fig. 4.12).

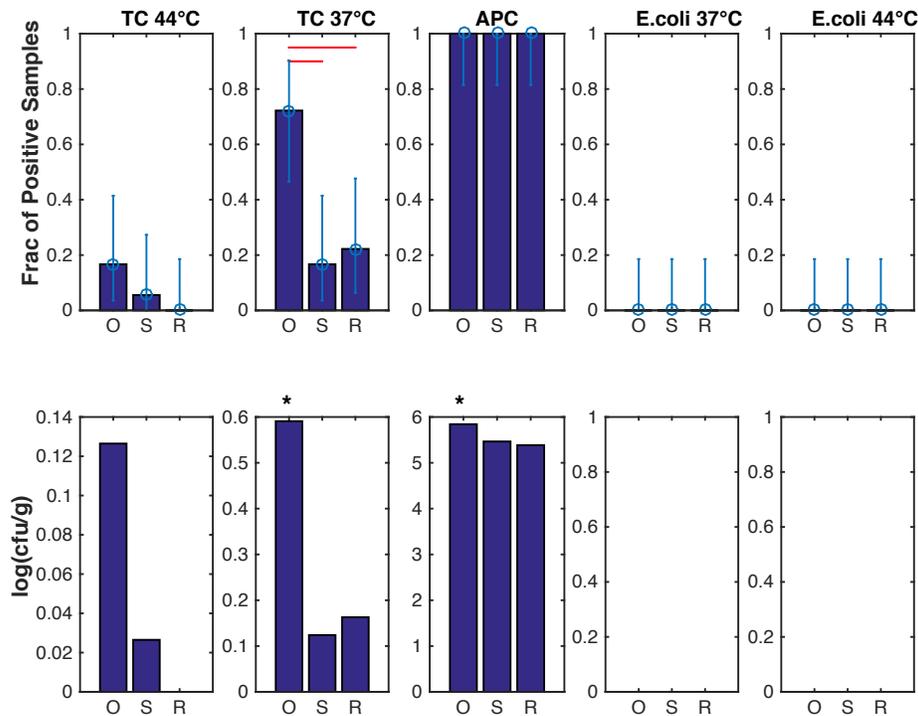


Figure 4.12 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with animal house nearby from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the second trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$).

For the field close to an animal facility, in the third trial, the fraction of positive samples from SOO sampling was significantly higher than stratified sampling based TC37 testing ($p < 0.05$). No *E.coli* and TC44 were detected. For the average bacterial count, SOO sampling had significantly higher counts than other two sampling plans for TC37 and APC assays ($p < 0.01$) (Fig. 4.13).

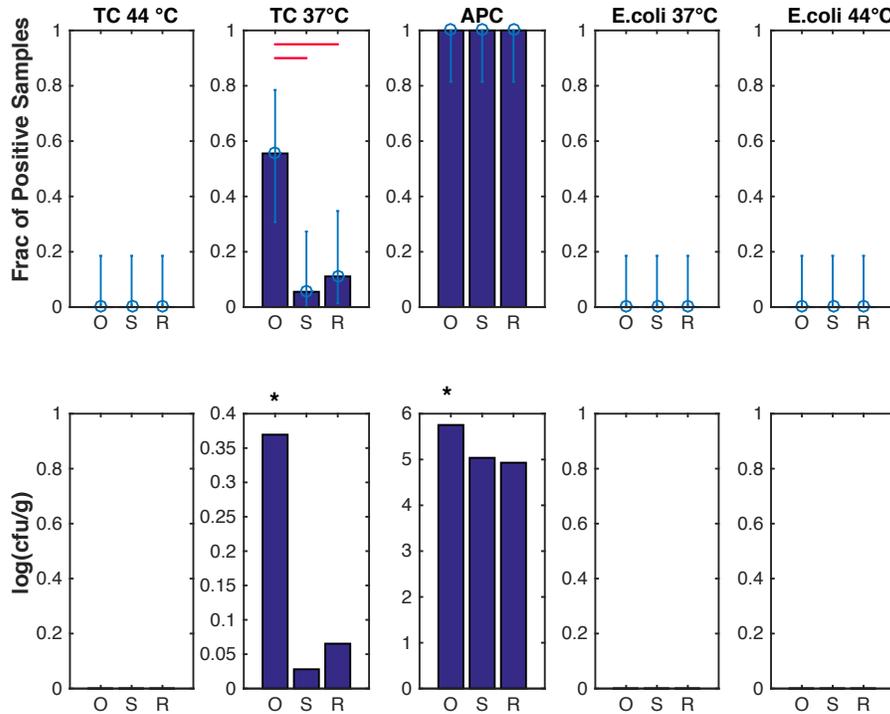


Figure 4.13 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with animal house nearby from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the third trial). Red bar and “*” mean significantly different based on Fisher’s exact test and ANOVA test ($p < 0.05$).

Below is an example of the positive field samples for the *E.coli*-37°C assay for the three sampling plans (Fig. 4.14). Nine samples from SOO collected close to the animals are all positive based on the *E.coli* indicator.

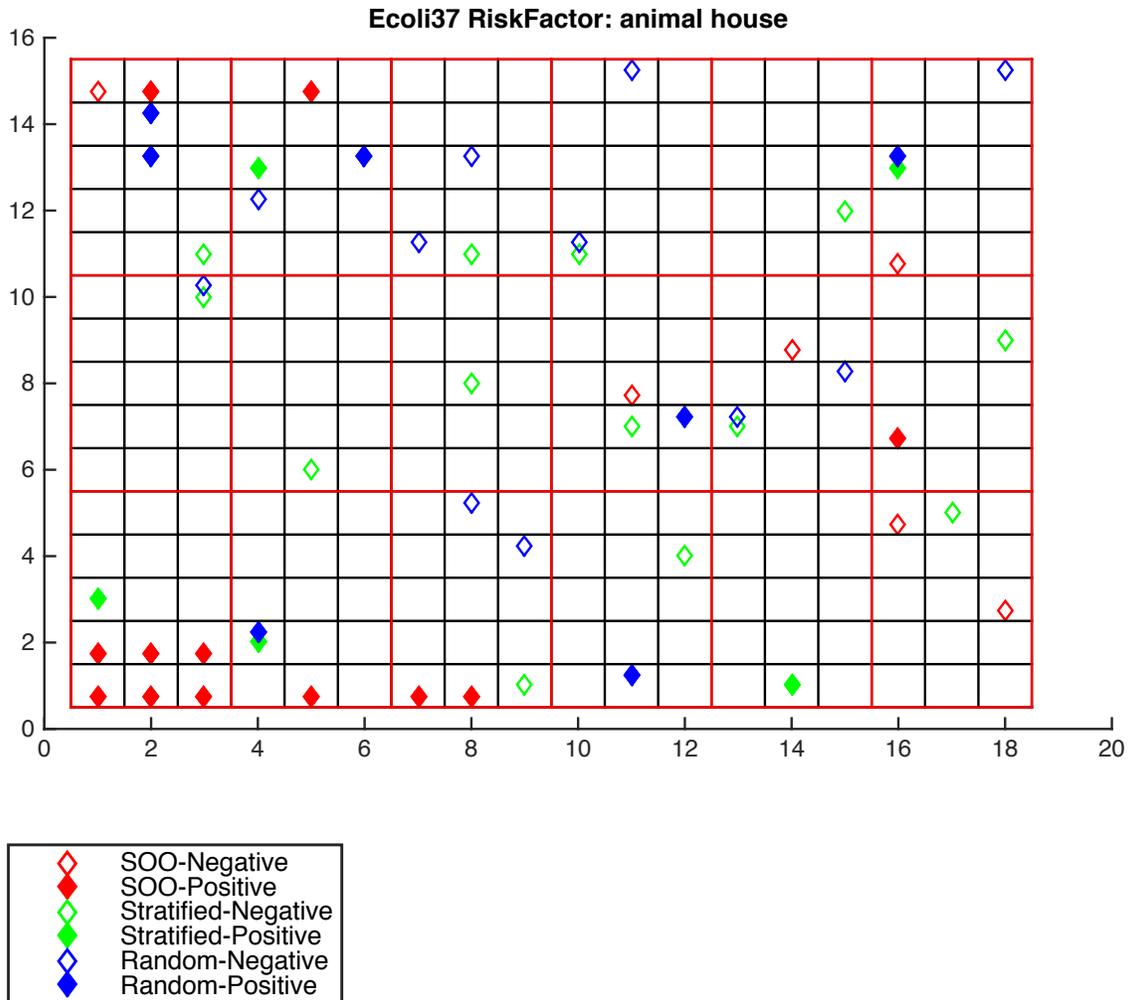


Figure 4.14 Example samples drawn from the field with animal house nearby according to three sampling plans (18 samples for each). Solid symbols represent positive samples of *E.coli* (37°C) indicator and empty symbols represent negative samples of *E.coli* (37°C) indicator based on 270 subplots in the field.

4.4.2.3 Field with power line above

For the field with an overhead power line, in the first trial, the fraction of positive samples from the presence/absence assays was significantly higher for SOO sampling plan than other two sampling plans for the TC37 assays ($p < 0.05$), but not for the EC37, EC44, or TC44 (Fig. 4.15). The mean bacterial counts for the

quantitative assays were significantly greater for the EC37, EC44, TC37 or TC44 assay with the SOO sampling, but not the APC assay (Fig. 4.15). Samples of “samples of opportunity” sampling have significantly higher counts than other two sampling plans based on EC and TC indicators ($p < 0.01$) (see Figure 4.15).

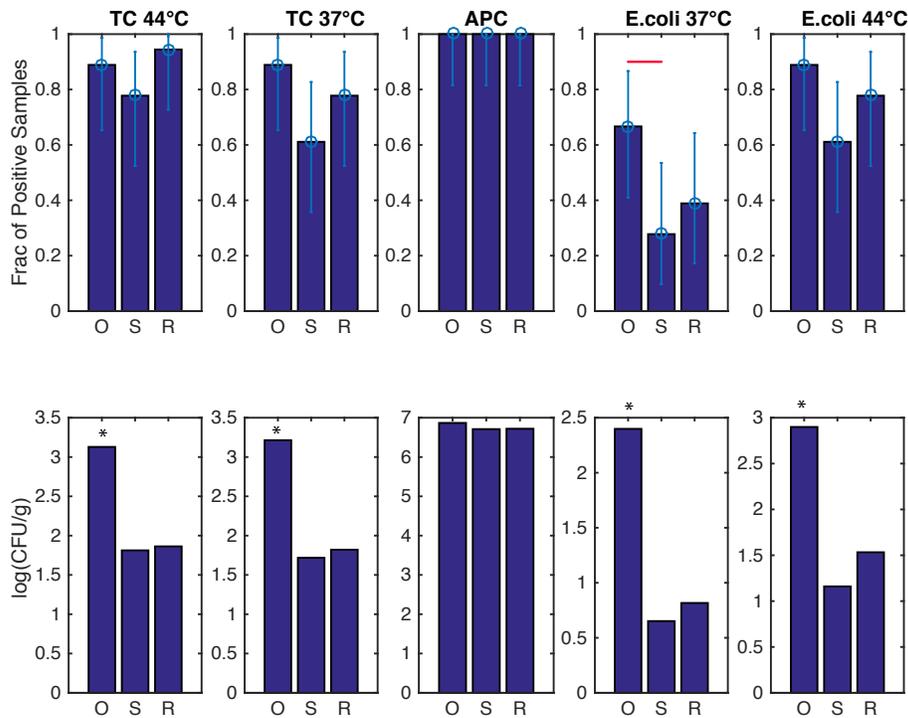


Figure 4.15 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with power line above from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the first trial). Red bar and “*” indicate that the means were significantly different ($p < 0.05$) based on Fisher’s exact test and ANOVA test, respectively.

For the field with an overhead power line, in the second trial, both of the fraction of positive samples from the presence/absence assays and the mean bacterial counts for the quantitative assays were not significantly different among sampling plans (Fig. 4.16).

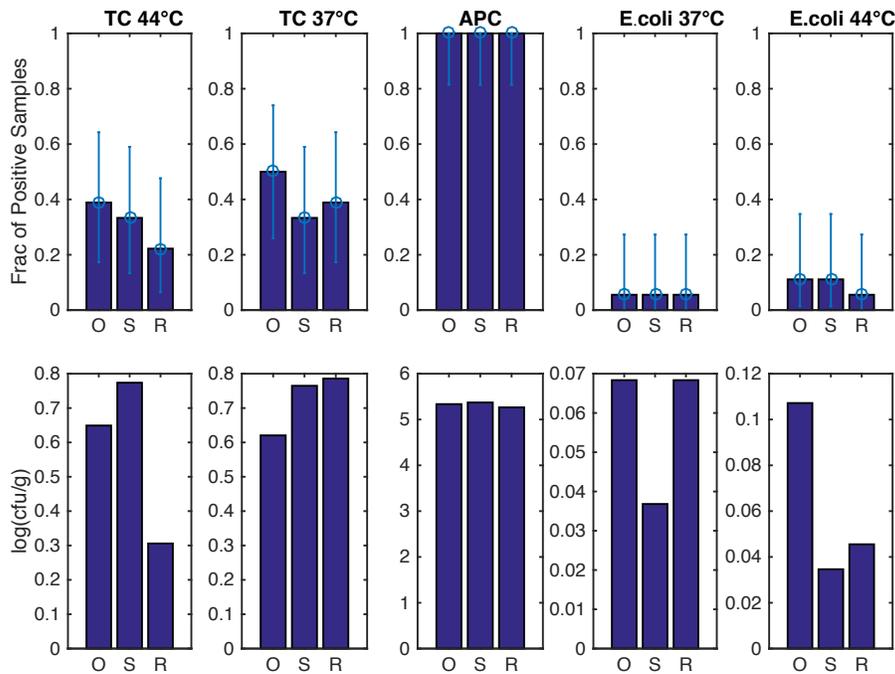


Figure 4.16 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with power line above from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the second trial). Red bar and “*” indicate that the means were significantly different ($p < 0.05$) based on Fisher’s exact test and ANOVA test, respectively.

For the field with an overhead power line, in the third trial, the fraction of positive samples from the presence/absence assays was significantly higher for S00 sampling plan than other two sampling plans for the TC37 assay ($p < 0.05$). No EC and TC44 were detected (Fig. 4.17). The mean bacterial counts for the quantitative assays were significantly greater for the EC37, TC37 and APC assays ($p < 0.01$) (Fig. 4.17).

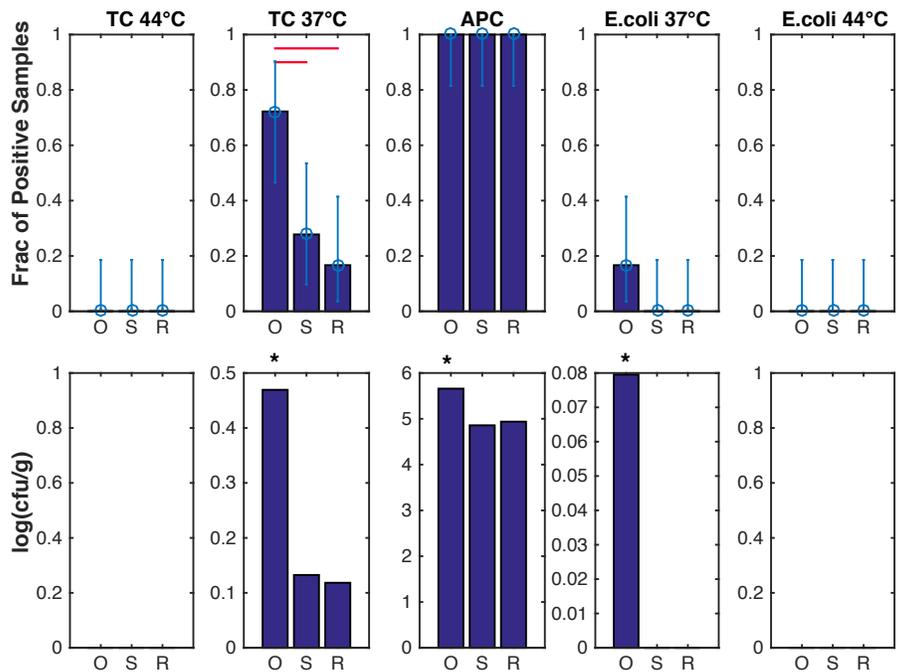


Figure 4.17 Fraction of positive samples out of 18 samples (first row) and levels of indicator microorganisms (second row) in samples in the field with power line above from three sampling plans: samples of opportunity sampling (O), stratified random sampling (R), and random sampling (R)(the third trial). Red bar and “*” indicate that the means were significantly different ($p < 0.05$) based on Fisher’s exact test and ANOVA test, respectively.

4.4.2.4 Comparison of sampling plans across three repeated trials

ANOVA repeated measure test was used to compare the three sampling plans based on three repeated trials. For the average count, samples of opportunity sampling had significantly higher TC-37 ($p = 0.008$) and APC ($p = 0.013$) counts in the flooded field trials (Table 4.2).

Table 4.2 The p-values of ANOVA repeated measures analysis on average bacterial indicators from three fields.

Indicator name	Animal House	Power line	Flooding
TC 44	0.37	0.44	0.11
TC 37	0.1	0.44	0.008*
APC	0.06	0.16	0.013*
EC37	0.44	0.05	0.067
EC44	0.44	0.44	0.129

“*” indicate that the mean detection probabilities of SOO sampling plan were significantly higher ($p < 0.05$) than other sampling plans based on ANOVA test.

For the fraction of positive samples, SOO sampling had significantly higher positive samples for TC37 ($p=0.02$) in the field with animal house nearby (Table 4.3). Samples of opportunity sampling had significantly higher positive samples for TC37 testing ($p=0.04$) in the field with the overhead power line. Samples of opportunity sampling had significantly higher positive samples for EC37 and EC44 assays ($p=0.004$) in the flooded field.

Table 4.3 The p-values of ANOVA repeated measures analysis on fraction of positive samples of five indicators out of 18 samples.

Indicator name	Animal house	Power line	Flooding
TC44	0.51	0.44	0.09
TC37	0.02*	0.04*	0.74
APC	NaN	NaN	NaN
EC37	0.44	0.13	0.004*
EC 44	0.44	0.44	0.004*

“*” indicates that the means were significantly different ($p < 0.05$) based on Fisher’s exact test. “NaN” means that the test cannot be done since all the samples are positive from all the sampling plans.

4.5 Discussion

Pre-harvest sampling and testing are important tool for verifying the microbiological safety of leafy greens. However, since the levels of foodborne pathogens are typically very low, it is hard to detect then with the number of samples typically used to evaluate pre-harvest fields. To improve the detection probability, it is important to develop efficient sampling plans for the pre-harvest produce. Several model studies have shown that stratified or systematic sampling plan performed better than random sampling for food products, particularly when the number of samples analyzed are limited [92– 94]. In the current study the effectiveness of different sampling plans were evaluated using both simulated fields and real fields to assess their ability to detect non-random contamination associated with three contamination source scenarios, point source contamination such as windblown contamination from a nearby animal facility, line contamination such as birds roosting on an overhead power line, and directional contamination as would be encountered in a partially flooded field. The basic hypothesis underlying the study was that sampling plans could be enhanced by taking advantage of knowledge

of the environment surrounding leafy green fields. Such SOO sampling plans are combination of knowledge informed biased sampling with traditional random or systematic sampling. Such hybrid sampling plans provided enhanced performance to detect target microorganisms in comparison to more traditional sampling plans for the three types of contamination sources using both simulation modeling and admittedly limited validation trials. Among the traditional approaches to sampling, stratified random sampling performed significantly better than random sampling based on the simulation modeling, which is consistent with the results of Jongenburger et al [94] who showed systematic sampling has higher mean detection probability than random sampling.

Several variables could affect the performance SOO sampling plans. The most important is whether expert knowledge led a sampler to correctly identify locations within a field with elevated, non-random contamination. The current study demonstrates that even if partially misjudging the likely location of non-random contamination, SOO sampling performed better than other sampling plans if the overlap between hypothesized and actual non-random contamination sites was $\geq 10\%$. Thus, SOO sampling would perform worse than traditional sampling only if the area of non-random contamination was completely missed. In such situation of increased uncertainty, the SOO approach could be adjusted by changing the percentage of sample that was devoted to biased sampling. Since many studies have been done on the risk factors leading to in-field contaminations, it would be very helpful to determine the most likely sources of contamination source for field prior to cultivation [89, 98].

In this study, only one contamination source was presented in the field. In real world scenarios, there might be more than one contamination sources. Therefore, all contamination sources should be considered before the sampler makes a decision about where to collect more samples. In future studies, multiple contamination sources could be simulated and subsequently validated. The current study provides a systemic way to estimate the detection probability of four sampling plans in the field. By considering three types of contamination with different spatial distributions, our model is readily able to detect non-random contaminations in the field for pre-harvest produce, such as *E. coli* or *S. enterica*. in lettuce or spinach. For the non-random contaminated field, a SOO sampling plan is suggested for detecting contaminations in pre-harvest leafy greens.

While the current study included limited validation studies, the number of replicate field trials coupled with the challenges of conducting blinded field studies mean that these trials must be considered preliminary in nature. This is partially overcome by simulation modeling where multiple iterations per simulation and multiple simulations allow a large number of *in silico* trials (e.g., 100 simulations each with 100 iterations) to be conducted. Ideally, partnerships with the leafy green growers could use the extensive pre-harvest testing data already being acquired to more extensively validate this hybrid approach to pre-harvest sampling.

4.6 Conclusion

This study has shown that the probability of detecting microbiological contamination by SOO plans is higher than random, stratified-random and z-pattern sampling plans based on the incorporation of knowledge concerning likely sources

of contamination. The enhanced detection probability of SOO sampling plan would be highly dependent on expert knowledge of the samplers and their ability to rapidly assess likely sources of contamination around the field environment and the percentage of samples that should be taken from high risk areas.

Chapter 5: *In silico* Evaluation of a Novel Iterative Bayesian Sampling Strategy for Efficient Detection of Pathogenic Bacteria in Pre-harvest Produce and Environments

5.1 Abstract

Sampling of pre-harvest leafy greens and their cultivation environments is increasingly used as a tool to enhance microbial food safety. In most sampling plans, the sample locations are determined beforehand and all samples are collected at once. This, in part, is because traditional methods of microbial detection take one or more days to yield results. However, recent developments in rapid microbial methods (RMMs) are significantly increasing the speed of analysis, which, in turn, makes it possible to begin considering iterative sampling strategies. The goal of this section of the current study was to compare the effectiveness of traditional sampling plans and a novel iterative sampling strategy based on Bayesian Global Optimization (BGO) using field simulations of realistic contamination sources. The effectiveness of iterative BGO sampling and two traditional sampling plans (random and stratified-random) were evaluated *in silico* using a simulation model. Pre-harvest fields similar to those described earlier with realistic contamination sources were generated *in silico*. Three types of contaminations were considered: stationary point contamination, stationary line contamination and directional line contamination. In these scenarios, it was assumed that (i) the likelihood of pathogen

presence was correlated with the levels of an indicator microorganism, and (ii) there is no *a priori* knowledge concerning the distribution of contamination. In the former it is assumed that the indicator microorganisms is associated with fecal contamination (e.g., *E. coli*, thermotolerant coliforms) and the specific pathogens of concern are transmitted via an oral-fecal route. The BGO plan uses prior results to inform the subsequent sampling locations to maximize overall detection probability. The same number of samples was collected in each sampling plan (n=18). In simulated fields with 5X6 plots and 9 subplots/plot (270 potential sampling locations and contamination sites), BGO sampling plan dramatically increased detection probability compared to traditional sampling plans (random: 0.30 ± 0.11 ; stratified random: 0.32 ± 0.11 ; BGO: 0.63 ± 0.23) with same number of samples. The difference was highly significant ($p < 0.0001$). This study provides a novel iterative sampling strategy for microbial quality testing. This alternative sampling approach would be particularly beneficial when implemented as part of testing program that monitors pre-harvest fields over the course of the cultivation cycle.

5.2 Introduction

Understanding microbiological distributions in pre-harvest settings can potentially improve sampling effectiveness by increasing detection probability, thereby improving food safety management decision-making based on the data [103]. Intuitively, the detection probability of a sampling plan can be improved if we collect more samples from areas with higher microbial contamination likelihoods. As discussed in previous chapters, one approach is to determine high risk areas based on knowledge of the field. For example, closeness to animal facilities,

presence of power lines above the field, and evidence of flooding can be used to define high risk areas[21, 76, 100]. However, when such knowledge is not available to the sampler, it can be difficult, if impossible to make informed decisions regarding high risk areas before samples are collected.

In this study, a novel iterative sampling approach is proposed. It is based on using knowledge gained as a result of prior samples to select additional sample locations in a manner that maximizes overall detection probability. It is based on treating sampling as a global optimization problem where the goal is the detection of the area or areas within a leafy green field with the greatest microbial contamination likelihood. Problems similar to this have been studied extensively in the field of geostatistics (e.g., estimating distribution of gold deposits based on samples from a few boreholes)[105] and in computer-based experimentation that requires optimizing algorithms for time-consuming simulations (e.g., automotive crash simulations) [106]. In the current study, a Bayesian Global Optimization (BGO) approach was selected to determine if detection efficiency could be enhanced by the iterative field testing over a relatively short time period.

The BGO method has several attractive features. First, it is able to maximize the expected improvement of a cost function which is the opposite of detection probabilities for each collected sample based on previously collected samples. Second, it provides a fast approximation to the overall distribution of microbial detection probability in the field, together with an estimate of the uncertainty at each location. Finally, it is able to provide a credible stopping rule based on the expected improvement of contamination probability from further searching. In this

chapter, the BGO sampling plan and traditional sampling plans are compared with the same number of samples on the detection probability of non-random contaminations.

5.3 Materials and Methods

5.3.1 Iterative sampling with Bayesian Global Optimization

The probability of detecting microbiological contamination in a field setting using a fixed number of samples is most effective when one can target areas with the highest likelihood of contamination. In the SOO sampling approach (see Chapter 4), this was achieved by knowledge of the potential sources of contamination in the cultivation location. In the BGO approach, this is alternatively achieved by using an iterative sampling approach where the results of the prior sampling inform the location of the next set of samples. We assumed that the contamination probability of each previously collected sample can be assessed by measuring indicator bacteria level, and that the contamination probability of nearby locations in a field is correlated. Given these assumptions, after obtaining the results of an initial sampling, a set of additional sampling iterations are developed as a global optimization problem, i.e., given the results of a set of previously collected samples, a set of new sample points are selected such that the expected contamination probability among all samples are maximized. This type of problem has been addressed in the field of Bayesian Global Optimization (BGO) [103, 104]. The mathematical formulation and the algorithmic solution to this problem are formulated in the context of microbiological sampling below.

In Bayesian global optimization, one considers optimization of a function f with respect to some parameters \mathbf{x} . The goal is to find an approximate solution to

$$\mathbf{x} = \underset{\mathbf{x} \in \mathbf{A}}{\operatorname{argmin}} f(\mathbf{x}) \quad (5.1)$$

In the case of microbiological sampling, \mathbf{x} represents the location of a sample in the field; the domain \mathbf{A} is set of all possible sample locations in the field. To maximize the microbiological contamination probability of the collected samples, $f = -\operatorname{risk}(\mathbf{x})$ is used, where $\operatorname{risk}(\mathbf{x})$ is the contamination likelihood at location \mathbf{x} . The evaluation of f is expensive and time-consuming, and that evaluations provide only the value of f at the evaluated point.

The function f is modeled as a Gaussian process, which is specified by its mean μ and a positive semi-definite covariance function $k(\mathbf{x}, \mathbf{x}')$. The concept of Gaussian process has a long history in the field of statistics. It can be thought as defining a distribution over functions or a collection of random variables, where any finite number of the random variables have a joint Gaussian distribution [109]. We denote the Gaussian process as,

$$f \sim GP(\mu, k), \text{ or } f(\mathbf{x}^{(i)}) = \mu + \epsilon(\mathbf{x}^{(i)}) \quad (5.2)$$

Where μ is the mean of the stochastic process, $\epsilon(\mathbf{x}^{(i)}) \sim \operatorname{Normal}(0, \sigma^2)$ and the correlation between errors is not zero. Instead, we assume the correlation between two locations in the field depends on the distance between the two locations via a square exponential function:

$$k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \text{Corr}[\epsilon(\mathbf{x}^{(i)}), \epsilon(\mathbf{x}^{(j)})] = \alpha \exp\left[-\sum_{l=1}^2 \frac{|x_l^{(i)} - x_l^{(j)}|^2}{\lambda_l^2}\right] \quad (5.3)$$

The model contains the parameters $\mu, \sigma, \alpha, \lambda_1, \lambda_2$. λ_1 and λ_2 , which are the spatial scales of the correlation along the horizontal and vertical direction, α controls the strength of the correlation. These parameters are estimated by choosing them to maximize the likelihood of the observed samples.

Before the BGO process is performed, an initial set of n samples is first collected according to random sampling plan. By analyzing these samples for a quantitative attribute, we can assess the microbiological contamination likelihood at a few locations of the field. Let $\mathbf{y} = (y^{(1)}, \dots, y^{(n)})$ denote the observed function value of n sample locations, Σ denotes the $n \times n$ matrix whose (i, j) th entry is $k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, and \mathbf{I} denote an n -vector of ones. The likelihood function [107] is given by:

$$\frac{1}{(2\pi)^{n/2} (\sigma^2)^{n/2} |\Sigma|^{1/2}} \exp\left[-\frac{(\mathbf{y} - \mathbf{I}\mu)^T \Sigma^{-1} (\mathbf{y} - \mathbf{I}\mu)}{2\sigma^2}\right] \quad (5.4)$$

Note that the dependences on α and λ is via the correlation matrix Σ . Given the correlation parameters, the values of μ, σ can that maximize the likelihood function can be solved in closed form

$$\hat{\mu} = \frac{\mathbf{I}^T \Sigma^{-1} \mathbf{y}}{\mathbf{I}^T \Sigma^{-1} \mathbf{I}} \quad (5.5)$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{I}\hat{\mu})^T \Sigma^{-1} (\mathbf{y} - \mathbf{I}\hat{\mu})}{n} \quad (5.6)$$

Given the estimated $\hat{\mu}$ and $\hat{\sigma}^2$, α and λ can be estimated by optimizing the likelihood function [107]. Once the estimated parameters were calculated, the posterior distribution of $f(\mathbf{x})$ based on knowledge of the n samples can be formed.

Given this posterior distribution, the expected function value for an arbitrary location \mathbf{x}^* is given by [110],

$$\hat{y}(\mathbf{x}^*) = \hat{\mu} + \mathbf{r}^T \Sigma^{-1}(\mathbf{y} - \mathbf{I}\hat{\mu}) \quad (5.7)$$

where \mathbf{r} is a n -dimensional vector containing correlations between the error terms at \mathbf{x}^* and the error terms at previously sampled point. That is, the i th element of \mathbf{r} is $r_i(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^{(i)})$. Intuitively, if \mathbf{x}^* is far away from any previously collected samples, its expected function value will be close to $\hat{\mu}$. The predictor interpolates the data between previously sampled points.

The mean squared error of this predictor can also be calculated, which is given by,

$$s^2(\mathbf{x}^*) = \sigma^2 \left[1 - \mathbf{r}^T \Sigma^{-1} \mathbf{r} + \frac{(1 - \mathbf{r}^T \Sigma^{-1} \mathbf{r})^2}{\mathbf{I}^T \Sigma^{-1} \mathbf{I}} \right] \quad (5.8)$$

In the expression above, the term $-\mathbf{r}^T \Sigma^{-1} \mathbf{r}$ represents the reduction in prediction error due to the fact that \mathbf{x}^* is correlated with the sample points. This adjustment would be zero if the \mathbf{x}^* is not correlated with any previously sampled points. The last term reflects uncertainty due to not knowing μ exactly.

The next step is to choose the set of points to evaluate next based on the posterior distribution. This is achieved using a decision-theoretic approach by optimizing a metric called the "expected improvement". Let $f_n^* = \min_{k \leq n} f(\mathbf{x}^{(k)})$ indicate the value of the best sample point evaluated. Suppose q more samples from

the field were collected, the value of the best point evaluated after all q samples are collected will be $\min(f_n^*, \min_{i=1, \dots, q} f(\mathbf{x}^{(i)}))$. The difference between these two values is called the improvement, which is equal to $[f_n^* - \min_{i=1, \dots, q} f(\mathbf{x}^{(i)})]^+$, where $[x]^+ = \max(x, 0)$. Note that although the exact value of $f(\mathbf{x}^{(i)})$, ($i = 1, \dots, q$) cannot be calculated, the estimated posterior distribution based on knowledge from previously collected samples can be calculated. An expected improvement can be calculated as,

$$q - EI(\mathbf{X}) = \mathbb{E}_n[[f_n^* - \min_{i=1, \dots, q} f(\mathbf{x}^{(i)})]_0] \quad (5.9)$$

where $\mathbb{E}_n[\cdot] = \mathbb{E}_n[\cdot | \mathbf{x}^{(1:n)}, \mathbf{y}^{(1:n)}]$ is the expectation taken with respect to the posterior distribution. Then, the set of points that maximizes the expected improvement is evaluated,

$$\operatorname{argmax}_{\mathbf{x}} q - EI(\mathbf{X}) \quad (5.10)$$

This can be achieved using stochastic gradient ascent algorithm and Monte Carlo simulation [108]. In this study, an open source library called Metric Optimization Engine (MOE) was used [111].

5.3.2 Model of non-random contamination factors

A field was divided into $N_x \times N_y$ plots of equal size. Each plot represents the smallest unit that can be sampled from. It is based on the assumption that the likelihood of contamination for a given plot depends on its spatial relationship with contamination sources. Three types of contamination are considered, including (1)

line contamination (e.g., below a power line), (2) point contamination (e.g., around an animal house) and (3) directional contamination (e.g., contamination due to flooding) (the same as chapter 4).

The contamination likelihood of due to line contamination is modeled as

$$I_{line}(x, y) = \exp\left\{-\frac{[\cos \theta (y - y_0) - \sin \theta (x - x_0)]^2}{2\sigma_{line}^2}\right\} \quad (5.11)$$

where θ is the orientation of the line, (x_0, y_0) is a point in the field that the line goes through, σ_{line} is the spatial spread of the contamination likelihood around the line(See Chapter4).

The contamination likelihood of the point contamination factor is modeled as a two dimensional Gaussian function

$$I_{point}(x, y) = \exp\left\{-\left[\frac{(x - x_0)^2}{2\sigma_{point}^2} + \frac{(y - y_0)^2}{2\sigma_{point}^2}\right]\right\} \quad (5.12)$$

where (x_0, y_0) is the location of the contamination source, σ_{point} controls the spatial spread of the point contamination(see Chapter 4).

The contamination likelihood of the directional contamination factor is modeled as a planar equation,

$$I_{directional}(x, y) = \exp\left\{k \left[\cos \theta \frac{(x - x_0)}{N_x} + \sin \theta \frac{(y - y_0)}{N_y} - 1 \right]\right\} \quad (5.13)$$

where θ is the direction of contamination, (x_0, y_0) is the corner of the field that is closest to the contamination source, k controls how fast the contamination decays along the direction of contamination (see Chapter 4).

The overall contamination likelihood is the sum of the three types of contamination likelihood.

$$I(x, y) = I_{line}(x, y) + I_{point}(x, y) + I_{directional}(x, y) \quad (5.14)$$

The contamination probability of a plot is proportional to the risk factor,

$$P(x, y) = \frac{I(x, y)}{\sum_{(x,y)} I(x, y)} \times N_c \quad (5.15)$$

where N_c is the number of contamination sites in the field. In the simulations we generate contamination sites according to the contamination probability equation without replacement to avoid contaminating the same plot multiple times (See Chapter 4).

5.3.3 Generation of simulated contaminated fields.

A large number of fields with non-random contamination were generated to systemically evaluate different sampling plans. Each simulated field has 18×15 plots and can be affected by only one of the three contamination types. Each plot represents a 1 m^2 square. A list of default parameters is given in table 5.1.

For each simulated field with line contamination the line orientation θ is randomly selected in the range of $[0, 180^\circ]$. A point in the center part (12×9) of the field is randomly selected as (x_0, y_0) to make sure that the line contamination goes

through the center of the field. The spatial spread of the line contamination σ_{line} is chose to be 1.

For each simulated field with point contaminations, a location on the boundary of the field is selected as the point location (x_0, y_0) . The spatial spread of the point contamination is set to be $\sigma_{point} = 3$.

For each simulated field with directional contamination, the direction of contamination is selected among $[0^\circ, 90^\circ, 180^\circ, 270^\circ]$ with equal probabilities, and (x_0, y_0) is selected to be $[(0, 0), (N_x, 0), (N_x, N_y), (0, N_y)]$ respectively. The parameter k is set to be 10 for directional contaminations.

5.3.4 Sampling plans

Three sampling plans are considered in this study, random, stratified random and BGO sampling. Each sampling plan collected the same number of distinct samples with different strategies. The number of samples per each sampling plan is denoted as N_{sample} . An example of samples drawn according to different sampling plans are shown in Figure 5.1.

The random sample plan collects N_{sample} samples randomly without replacement from all plots. The stratified random sample plan first divides the field into 5×6 plots, with each main plot containing 3×3 subplots. There is at most one sample collected from each plot.

The BGO sampling plan first collects $N_{sample}^{initial}$ samples randomly from the field (without replacement), and then collects the rest of the samples across multiple N_{iter} iterations, where N_{sample}^{iter} samples are collected during each iteration. The total

number of samples ($N_{sample}^{initial} + N_{iter}N_{sample}^{iter}$) is the same as other sampling plans. We chose $N_{sample}^{initial} = 6$ for all simulations if not specified otherwise. The number of iterations and samples per iteration are varied in the simulations. We included a noise term to mimic uncertainty of contamination likelihood evaluation of collected samples. The measured contamination likelihood is the true contamination likelihood times a multiplicative noise.

$$f_{measured}(\mathbf{x}) = f(\mathbf{x})z \quad (5.16)$$

where $z \sim \text{Gamma}(\frac{1}{\theta}, \theta)$ is a random variable that follows a gamma distribution with mean of one and variance of θ . The parameter θ is the scale parameter of the Gamma distribution and controls the strength of noise. The use of the Gamma noise ensures that the measured noise is always greater or equal to zero. θ was varied between 0.1 and 5 in the simulation.

Table 5.1 Summary of model parameters

Parameter	Description	Default value
N_x	Number of plots along the x-axis	18
N_y	Number of plots along the y-axis	15
σ_{line}	Spatial spread of line contamination	1
σ_{point}	Spatial spread of point contamination	3
k	Decay rate of directional contamination	10
N_{sample}	Number of samples	18
N_c	Number of contaminated plots in the field	6

$N_{sample}^{initial}$	Number of samples collected during the first iteration of the BGO sampling	6
N_{iter}	Number of iterations of the BGO sampling	12
N_{sample}^{iter}	Number of samples collected in each iteration	1

5.3.5 Evaluation of sampling plans

Sampling plans were compared using computer simulations. At the beginning of each simulation, the likelihood of contamination on a simulated field was generated with one of the three contamination sources (Eq. 5.14). Then N_c plots were selected as contaminated plots based on the contamination likelihood (Eq. 5.15). This simulated field was then sampled using the three types of sampling plans (random, stratified-random and BGO), each with 20 iterations. With each iteration, the sample locations were generated according to different sampling plans independently. For random and stratified-random sampling plans, the sample locations were generated in one pass without replacement. For BGO sampling plans, the sample locations were generated in an iterative fashion (eq. 5.10). The BGO algorithm generated new sampling locations based on the contamination likelihood value of the samples collected previously. The same number of unique samples was collected for all three sampling plans. A detection probability was estimated based on the number of iterations that a sampling plan successfully detected at least one of the contaminated sites. The simulation was then repeated 20 times, each with a different contaminated field, to estimate the distribution of detection probabilities for each sampling plan.

5.4 Results

The effectiveness of random, stratified-random and BGO sampling plans were compared on simulated contaminated fields with non-random contamination patterns. The BGO sampling plan collects samples iteratively. At each iteration, BGO uses information of previously collected samples on the contamination likelihood to decide where to collect samples next such that the overall sampling effectiveness can be maximized. The set of samples drawn based on the BGO plan is shown together with samples drawn from random and stratified random sampling plans (Figure 5.1A). To highlight the iterative sampling of BGO, we labeled the order of each collected samples. The BGO sampling typically have three phases. The first six samples are collected at one time randomly to have an initial estimation of the contamination patterns in the field. The BGO collected a few more samples iteratively (one sample each time) from locations with high uncertainty (sample 7-11). The final set of samples (sample 12-18) are collected iteratively (one sample each time) from areas with the highest contamination likelihood. Overall BGO tends to collect more samples from areas with high contamination likelihood than the random and stratified random sampling plan.

After a set of samples is collected, the BGO sampling plan provides the estimated distribution of microbial contamination likelihood (Fig. 5.1B), together with the uncertainty (standard deviation) of this estimated contamination likelihood at each location in the field (Fig. 5.1C). In this case the estimated distribution closely resembles the true contamination likelihood distribution (Fig. 5.1A) even though only a small number of samples are collected. The uncertainty is

low near the sampled locations. These estimators are used to guide selection of the next sample points such that the expected improvement with the new samples will be maximized (see section 5.3.1).

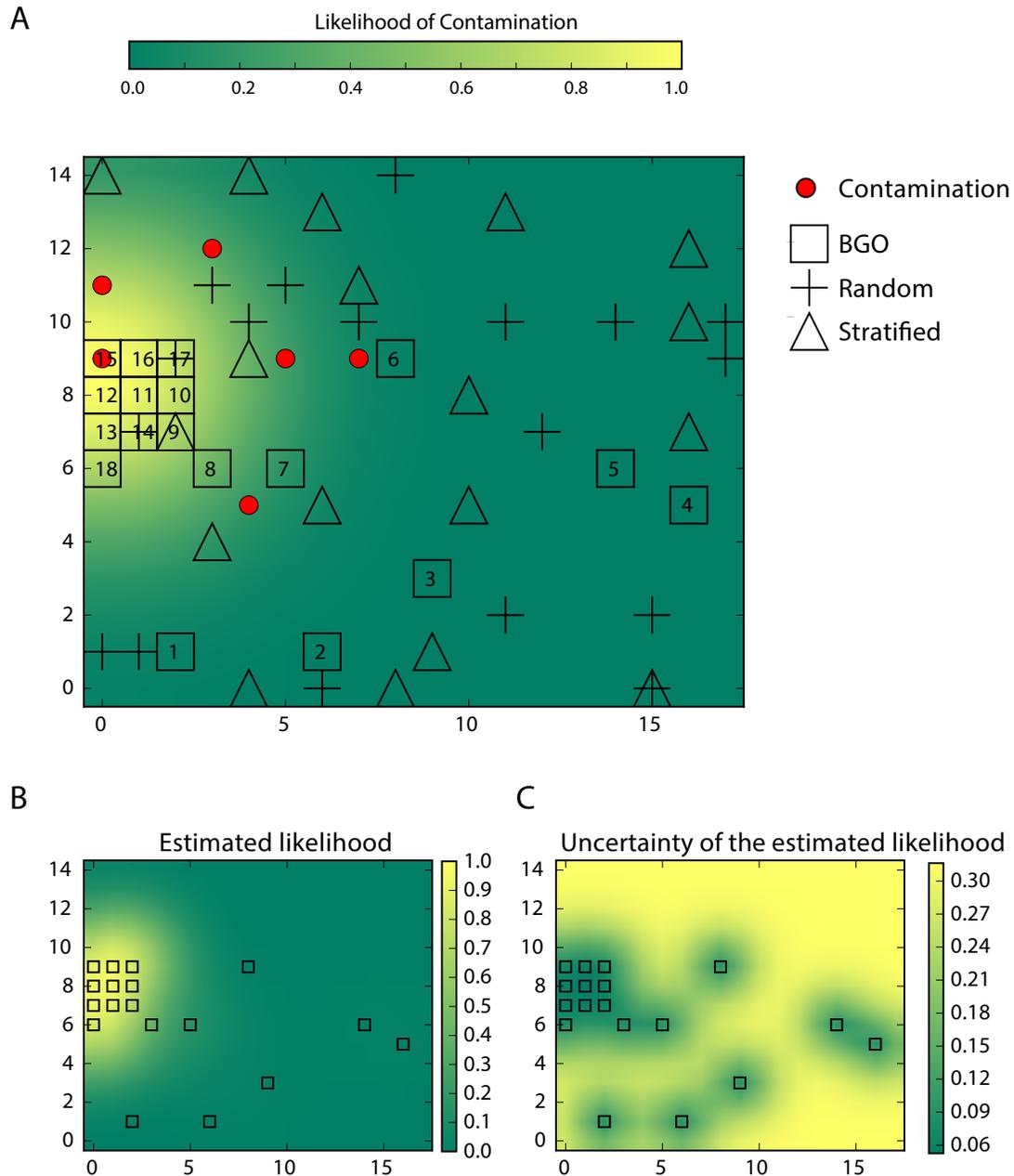


Figure 5.1 Example samples drawn according to different sampling plans. A. The samples drawn according to the random (cross), stratified random (triangle) and BGO (square) are shown together with contaminations and likelihood of contaminations. The order of the BGO samples are labeled with numbers. B.

Estimated distribution of contamination likelihood in the field from the BGO sampling plan. C. Standard deviation of the estimated likelihood based on the BGO sampling plan.

The performance of the three sampling plans on a large number of simulated fields with different types of contamination was quantified, including point contamination, line contamination and directional contaminations. The average detection probability is shown in Fig. 5.2A. BGO has significantly better detection probability than the other two sampling plans for all three types of contamination patterns ($p < 0.05$). The improvement is most dramatic for fields that had high contamination near a hot spot or along a line. The stratified sampling plan has small, but significant better performance than the random sampling plan on point contamination using a pairwise test ($p < 0.05$). The difference is not significant for other types of contaminations.

Detected contamination likelihood of samples collected from different sampling plans was analyzed. The samples collected by the BGO plan have higher likelihood on average than samples collected by the random or stratified sampling plan (Fig. 5.2B). The difference is significant for all three types of contaminations. The maximum of the detected likelihood among the 18 collected samples also differs across sampling plans (Fig. 5.2C). The BGO plan typically can find locations with high contamination likelihood towards the end of sampling.

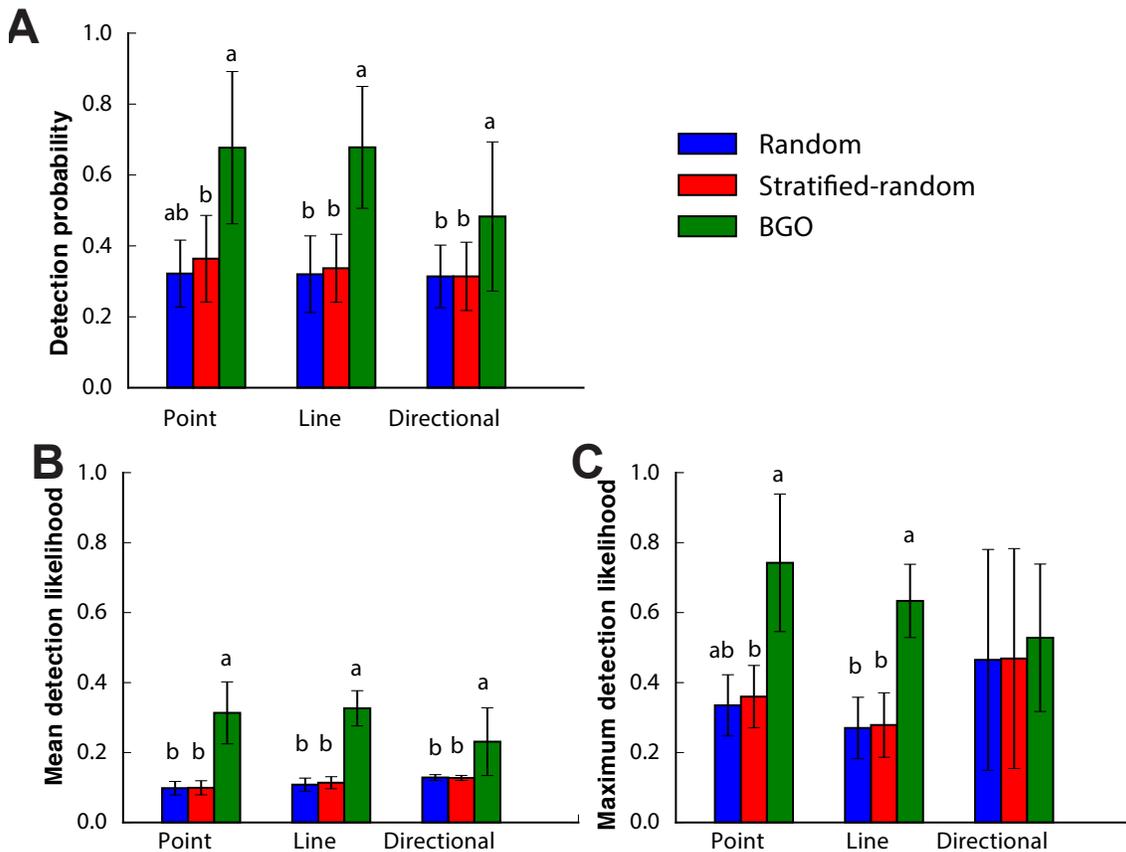


Figure 5.2 Performance of three sampling plans with different types of contamination patterns. (A) Detection probability of the contaminated site for different sampling plans. (B) Average contamination likelihood of samples collected from different sampling plans. (C) The maximum contamination likelihood among the 18 samples collected from different sampling plans. The error bars represents standard deviation across different fields with the same types of contaminations. Values from bars with different letters are significantly different based on an ANOVA and Tukey's *post hoc* test ($P < 0.05$).

Detection probability as a function of number of samples for the three sampling plans was analyzed (Fig. 5.3). The detection probability increases when more samples are allowed to be collected. Because BGO iterative sampling only collected samples iteratively after more than six samples, there is no difference between BGO and random sampling plans for $N_{sample} \leq 6$. BGO starts to have higher

detection probability than random and stratified-random plan when $N_{sample} \geq 10$.

The difference increases as a function of number of samples.

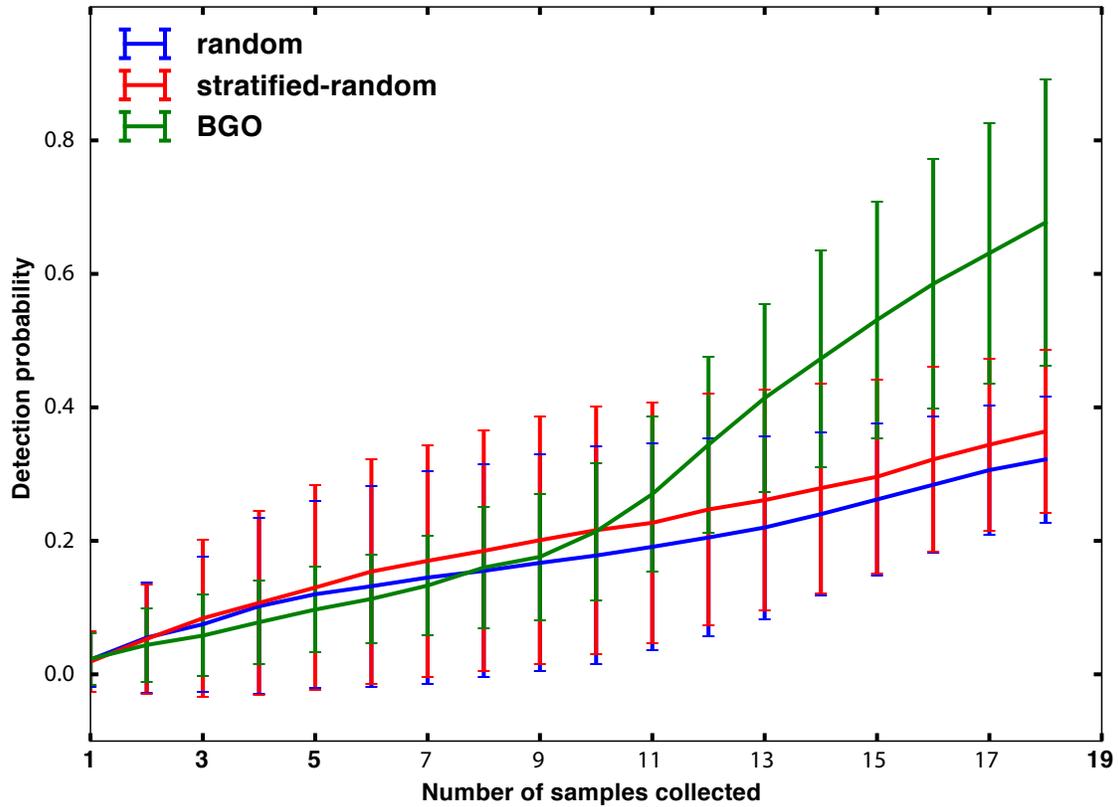


Figure 5.3 Performance of three sampling plans with different number of samples collected. The detection probability is plotted as a function of number of samples. Note that the first 6 samples of the BGO plan are drawn randomly. The difference between BGO and other sampling plans is significantly when number of samples is higher than 10.

The simulations in Figure 5.2 and Figure 5.3 collect one sample per iteration in the BGO plan. How the detection probability changes when the number of samples per iteration changes while the total number of sample is fixed is evaluated. It found that the detection probability decreases if fewer numbers of iterations is allowed (more number of samples per iteration) (Fig. 5.4). The BGO plan is still

significantly better than random and stratified random sampling plans with 6 samples per iterations (3 iterations total). However, the difference between BGO and other sampling plans is small and insignificant if with 9 samples per iterations (2 iterations).

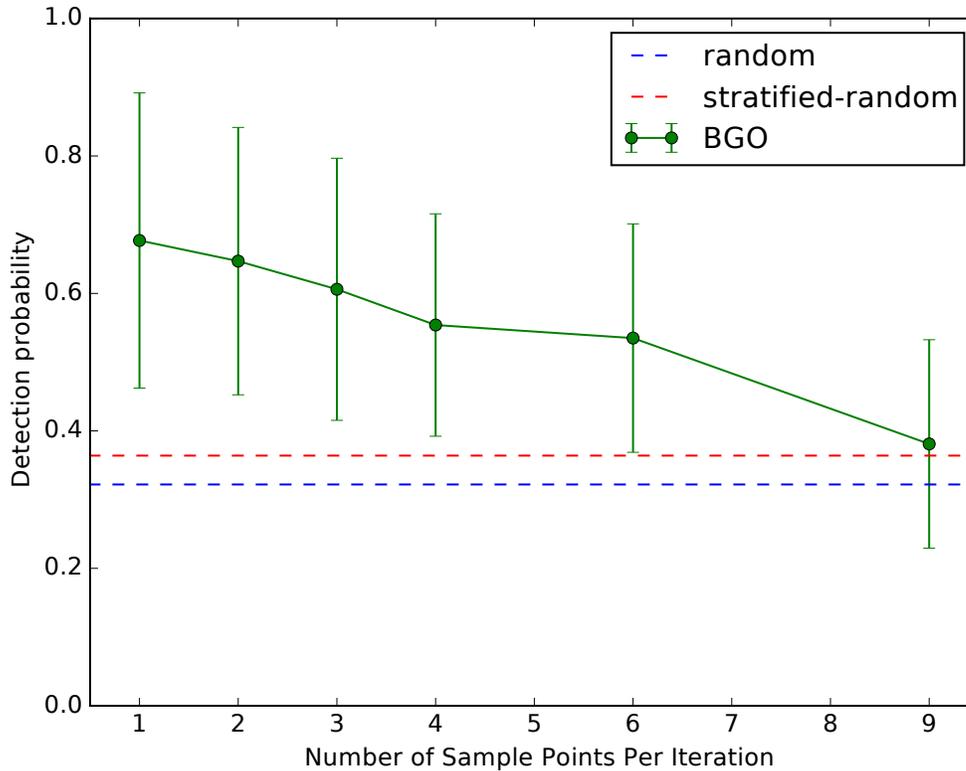


Figure 5.4 Detection probability of the BGO plan as a function of number of samples per iteration. The detection probability of BGO is calculated with increasing number of samples per iteration (decreasing number of iterations) on fields with point contaminations. The number of samples collected in the initial random sampling period is 6 for $N_{sample}^{iter} \leq 6$ and 9 for $N_{sample}^{iter} = 9$. The total number of iterations is 12, 6, 4, 3, 2, 1 for $N_{sample}^{iter} = 1, 2, 3, 4, 6, 9$ respectively. The detection probability for random and stratified-random plans are shown as blue and red dashed lines respectively.

So far the microbiological contamination likelihood assumed can be perfectly assessed for previously collected samples. The robustness of the BGO sampling plan

with respect to noise during assessment of contamination likelihood is evaluated next. The uncertainty of contamination likelihood was modeled with a multiplicative Gamma noise. The measured likelihood is chosen to be the product of the true likelihood and a Gamma noise, where the variance of the Gamma noise is varied to control noise strength (see Materials and Methods Eq. 5.16). The detection probability decreases when small amount of noise is introduced. However, the detection probability of the BGO plan remains to be much higher than other sampling plans even with very large amount of noise (Fig. 5.5). The results are the same with other types of noise.

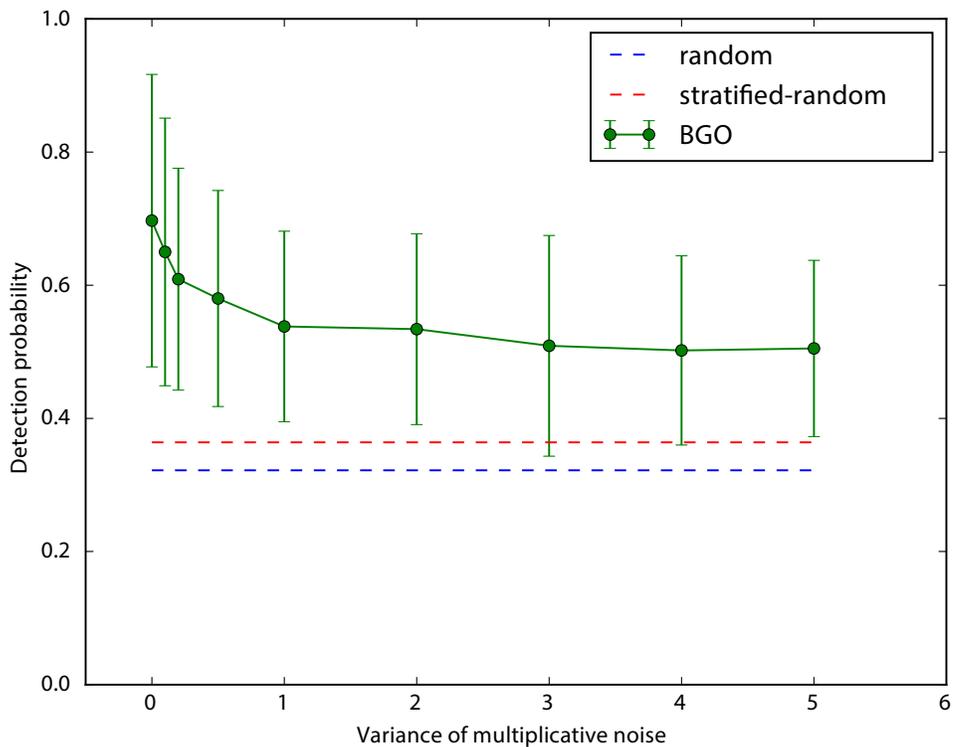


Figure 5.5 Robustness of the BGO plan to noise in contamination likelihood measurement. Mean detection probability is plotted as a function of the variance of the multiplicative Gamma noise. The mean detection probabilities for random and

stratified-random plans are shown as blue and red dashed lines respectively. This simulation is performed with point contamination.

5.5 Discussion

In this study, BGO gives much better detection probability than traditional sampling plans where there is not *a priori* knowledge of the distribution of contamination. The effectiveness of the BGO sampling plan relies on two assumptions. The first assumption is that the microbial contamination likelihood varies smoothly in the field. Locations that are close to each other have similar microbial contamination likelihood. Second, microbial contamination likelihood of previously collected samples could be evaluated in a quantitative fashion. This can be achieved, for example, by measuring the amount of indicator bacteria in each sample. There are many studies that report indicators bacteria are closely related to pathogen occurrence. For example, fecal indicators are often correlated with *Salmonella* spp. [112][113]. A study also shown that the correlation between indicators and pathogens are controversy due to insufficient data [114].

One potential drawback of the BGO sampling plan is the paradigm of iterative sampling requires longer sampling time. It is assumed that the contamination likelihood does not change during the course of BGO sampling. However, since microbiological contamination likelihood of previously collected samples are assessed before deciding where to sample next, the entire sampling process can take multiple days. It shows that the detection probability is significantly better than random sampling plan even with a small number of three iterations. In traditional ways, it would take at least three days to finish three iterations since it would take

at least 24h to get the detection results. Currently it would take at least a week if samples were collected 6 times iteratively. However, the contamination pattern could change overtime. For example, the pathogen or indicator microorganisms could die off over-time or re-emergence due to a rain event. New contamination source also could appear after first iterative sample is taken. So, traditional detection methods may not suitable for use with BGO method.

The development of rapid and ultra-rapid microbial-detection methods such as polymerase chain reaction (PCR)[115], quartz crystal microbalance (QCM)[116], surface plasmon resonance (SPR)[117], allows samplers to obtain microbiology result faster compared with traditional culture-plate methods [110 – 113]. In 2015, an ultra-rapid microbial-detection method using an immunoassay in combination with a 3D-printed helical microchannel device was invented[122]. Sometimes the result can be obtained in a matter of hours, as opposed to days or weeks in traditional cases [123]. These rapid methods for microbial detection can be sensitive and quick, which may allow multiple rounds of sampling in a relatively short period of time. The iterative sampling could be finished one or two days by using rapid detection methods. This is reasonable and possible with the rapid development of the rapid methods for microbial detection.

The BGO sampling plan requires evaluation of the microbial contamination likelihood of each sample. One important question is that the measurement of microbial contamination likelihood may not be accurate in practice. This can be due to inaccuracies of plate counts [124], low correlations between indicator bacteria and pathogens, and other experimental error. Inaccuracies of contamination

likelihood measurement are simulated by multiplying the true contamination likelihood with a noise variable that follows Gamma distribution with a mean of one. It found that although the performance of the BGO sampling plan decreases initially with small amount of noise, it remains much better than traditional sampling plans over a wide range of noise. This suggests that the improvement of BGO sampling is robust with respect to inaccuracy of risk assessment.

The BGO sampling plan and SOO sampling plans (in chapter 4) also can be combined. When the high-risk area was found by BGO sampling plan based on a microbial indicator, SOO sampling plan could then be used to assess the likelihood of a target pathogen. For example, 12 samples are collected in high-risk area identified by the BGO results, and 6 samples are collected randomly in the field.

Chapter 6: Summary and Future Studies

6.1 Summary

Pre-harvest microbiological testing of leafy green vegetables based on statistically valid sampling plans is increasingly becoming an important of food safety systems for fresh and fresh-cut produce. However, the efficacy of the current sampling plans was very low and there is little consensus on optimal sampling plans that are “fit for purpose”. The research project has attempted to systematically evaluate the effectiveness of current common pre-harvest sampling plans, and explored means of enhancing detection capabilities through understanding the underlying statistical concepts and parameters.

Chapter 3 focused on evaluating the performance three pre-harvest sampling plans, random, stratified random, and Z-pattern sampling by consideration of their mathematical derivations and computer simulations based on random contamination sites. It has been concluded that while the mean result obtained with all three sampling plans is similar, the performance of the random or stratified random sampling plans are less variable, particularly when the number of contamination sites or number of samples analyzed are small. Furthermore, validation field study was conducted to compare three sampling plans based on

random contamination. Fewer positive samples were detected in z-pattern sampling plan.

In the real world, contaminations always occur non-random due to many risk factors in the field, such as wild animals, poultry, flooding. In chapter 4, computer simulations were performed to compare the relative effectiveness of random, stratified-random, Z-pattern and SOO sampling which is the method based on the sampler's knowledge of risk factors. SOO sampling plan performed the best among all the sampling plans if the assumption of contamination source is incorrect. This study indicated that having the knowledge of the contamination source in the field would highly improve effectiveness of sampling. Field study was conducted to validate the model in Chapter 4 for the non-random contaminated fields. It concluded that SOO sampling plan performed better than other sampling plans based on the field with power line above, field with animal house nearby and flooded field.

However, if the assumption of contamination source were incorrect, the detection probability of SOO would be not significantly different than the other sampling plans. It indicate that it highly depend on the knowledge and decision of the samplers. It cannot be always right. Therefore, in chapter 5, a novel sampling method is invented, which is iterative sampling strategy based on Bayesian Global Optimization (BGO). In this chapter, the effectiveness of traditional sampling plans and a novel iterative sampling strategy were evaluated on simulated fields with realistic contamination sources. The study concluded that the novel sampling

strategy performed much better than traditional sampling plans in realistic scenarios and can save samples as well.

To our knowledge, this is the first systemic study to evaluate and compare pre-harvest sampling plans for food microbial safety. The current project compares all the current common pre-harvest sampling plans based on random and non-random contaminations using computer simulations and field validations. A novel sampling strategy is also developed to get better detection probability and improve sampling efficacy without field knowledge.

The outcomes of the current project provide scientific evidence and guidelines for researchers and samplers. The suggested sampling plans for different situations are discussed as follows.

- (1) If the field is randomly contaminated, such as contaminated by overhead irrigation water, random and stratified random sampling plan are suggested.
- (2) If the field is non-randomly contaminated or there is some potential risk factors near the field, such as power line above the field, animal house nearby, flooded field and wild animal activity in the field, then samples of opportunity sampling plan was recommended.
- (3) If you have no idea of the field and you also want to evaluate the contamination situation with least samples, then iterative sampling (BGO) is suggested.

6.2 Future studies

The current study represents our current best knowledge of pre-harvest sampling plans for produce microbial safety. Several gaps were identified and elaborated in each chapter. Some possible areas of research related to pre-harvest sampling plans are proposed as follows.

- (1) For the non-random contaminated field, this study covered the field with one contamination source, the field with multi-contamination sources can be researched.
- (2) For the model on samples of opportunity, more details need to be considered, such as, weather and wind strength, the height of the power line, the size and mass of the bird droppings, the animal species in the animal house. The contamination range can be changed due to the above factors.
- (3) Validation studies need to be done on the iterative sampling. The BGO sampling plan need to be applied in the field to further assess the feasibility and efficacy.

References

- [1] C. Naanwaab and O. Yeboah, "Demand for fresh vegetables in the United States: 1970–2010," *Econ. Res. Int.*, vol. 2012, p. 942748, 2012.
- [2] H. F. Wells, J. Bond, and S. Thornsby, "Vegetables and pulses outlook," *USDA Economic Research Service*, 2015. [Online]. Available: <http://usda.mannlib.cornell.edu/usda/ers/VGS//2010s/2015/VGS-05-01-2015.pdf>. [Accessed: 03-Nov-2015].
- [3] J. A. Painter, R. M. Hoekstra, T. Ayers, R. V. Tauxe, C. R. Braden, F. J. Angulo, and P. M. Griffin, "Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998–2008," *CDC*, vol. 19, no. 3, pp. 407–415, 2013.
- [4] A. Xu, D. M. Pahl, R. L. Buchanan, and S. A. Micallef, "Comparing the microbiological status of pre- and postharvest produce from small organic production.," *J. Food Prot.*, vol. 78, no. 6, pp. 1072–80, Jun. 2015.
- [5] FDA, "Food Code," 2013. [Online]. Available: <http://www.fda.gov/Food/GuidanceRegulation/RetailFoodProtection/FoodCode/ucm374275.htm>. [Accessed: 14-Feb-2016].
- [6] CDC, "Multistate outbreak of *E. coli* O157:H7 infections linked to fresh spinach (FINAL UPDATE)," 2006. [Online]. Available: <http://www.cdc.gov/ecoli/2006/spinach-10-2006.html>. [Accessed: 14-Feb-2016].

- [7] FDA, "Hazard analysis critical control point (HACCP)." Center for Food Safety and Applied Nutrition, 2015.
- [8] D. Zagory, "Microbial testing of fresh produce: Where is the value?," *Food Safety News*, 2014.
- [9] M. H. Zwietering, T. Ross, and L. G. M. Gorris, "Food safety assurance systems: Microbiological testing, sampling plans, and microbiological criteria - Wageningen UR," in *Encyclopedia of Food Safety, Volume 4: Food Safety Management*, MA : Academic Press, 2014, pp. 244–253.
- [10] E. Scallan, R. M. Hoekstra, F. J. Angulo, R. V Tauxe, M.-A. Widdowson, S. L. Roy, J. L. Jones, and P. M. Griffin, "Foodborne illness acquired in the United States--major pathogens," *Emerg. Infect. Dis.*, vol. 17, no. 1, pp. 7–15, Jan. 2011.
- [11] CDC, "FoodNet's progress reports," 2013. [Online]. Available: <http://www.cdc.gov/foodnet/data/trends/trends-2013-progress.html>. [Accessed: 01-Jan-2016].
- [12] L. A. Rodriguez-Romo and A. E. Yousef, "Inactivation of *Salmonella enterica* serovar Enteritidis on shell eggs by ozone and UV radiation.," *J. Food Prot.*, vol. 68, no. 4, pp. 711–7, Apr. 2005.
- [13] M. D. M. Mandal Shyamapada, "Plasmid encoded UV-resistance and UV induced ciprofloxacin resistance in *Salmonella enterica* serovar Typhi," *Int. J. Integr. Biol.*, vol. 2, no. 1, p. 43, 2008.
- [14] C. Pui, W. Wong, L. Chai, R. Tunung, P. Jeyaletchumi, M. Noor Hidayah, A.

- Ubong, M. Farinazleen, Y. Cheah, and R. Son, "Review article *Salmonella*: A foodborne pathogen," *Int. Food Res. J.*, vol. 18, pp. 465–473, 2011.
- [15] E. Franz and A. H. C. van Bruggen, "Ecology of *E. coli* O157:H7 and *Salmonella enterica* in the primary vegetable production chain.," *Crit. Rev. Microbiol.*, vol. 34, no. 3–4, pp. 143–61, Jan. 2008.
- [16] J. Konowalchuk, J. I. Speirs, and S. Stavric, "Vero response to a cytotoxin of *Escherichia coli*," *Infect. Immun.*, vol. 18, no. 3, pp. 775–9, Dec. 1977.
- [17] C. M. Thorpe, "Shiga toxin--producing *Escherichia coli* infection," *Clin. Infect. Dis.*, vol. 38, no. 9, pp. 1298–1303, May 2004.
- [18] R. K. S. Khalil, M. A. E. Gomaa, and M. I. M. Khalil, "Detection of shiga-toxin producing *E. coli* (STEC) in leafy greens sold at local retail markets in Alexandria, Egypt," *Int. J. Food Microbiol.*, vol. 197, pp. 58–64, Mar. 2015.
- [19] FDA, "FDA finalizes report on 2006 spinach outbreak," 2007. [Online]. Available:
<http://www.fda.gov/newsevents/newsroom/pressannouncements/2007/ucm108873.htm>. [Accessed: 20-Jul-2014].
- [20] S. J. Bach, T. A. McAllister, D. M. Veira, V. P. J. Gannon, and R. A. Holley, "Transmission and control of *Escherichia coli* O157:H7 — A review," *Can. J. Anim. Sci.*, vol. 82, no. 4, pp. 475–490, Dec. 2002.
- [21] E. D. Berry, J. E. Wells, J. L. Bono, B. L. Woodbury, N. Kalchayanand, K. N. Norman, T. V Suslow, G. López-Velasco, and P. D. Millner, "Effect of proximity

- to a cattle feedlot on *Escherichia coli* O157:H7 contamination of leafy greens and evaluation of the potential for airborne transmission.," *Appl. Environ. Microbiol.*, vol. 81, no. 3, pp. 1101–10, Feb. 2015.
- [22] I. T. Kudva, K. Blanch, and C. J. Hovde, "Analysis of *Escherichia coli* O157:H7 survival in ovine or bovine manure and manure slurry.," *Appl. Environ. Microbiol.*, vol. 64, no. 9, pp. 3166–74, Sep. 1998.
- [23] J. M. Sargeant, D. J. Hafer, J. R. Gillespie, R. D. Oberst, and S. J. Flood, "Prevalence of *Escherichia coli* O157:H7 in white-tailed deer sharing rangeland with cattle.," *J. Am. Vet. Med. Assoc.*, vol. 215, no. 6, pp. 792–4, Sep. 1999.
- [24] CDC, "Parasites - *Cryptosporidium* (also known as 'Crypto')," 2015. [Online]. Available: <http://www.cdc.gov/parasites/crypto/>. [Accessed: 02-Dec-2015].
- [25] A. Howe, S. Forster, S. Morton, R. Marshall, K. Osborn, P. Wright, and P. Hunter, "*Cryptosporidium* oocysts in a water supply associated with a cryptosporidiosis outbreak," *Emerg Infect Dis.*, vol. 8, no. 6, 2002.
- [26] S. Glaberman, J. Moore, C. Lowery, R. Chalmers, I. Sulaiman, K. Elwin, P. Rooney, B. Millar, J. Dooley, A. Lal, and L. Xiao, "Three drinking-water-associated *cryptosporidiosis* outbreaks, Northern Ireland," *Emerg Infect Dis.*, vol. 8, no. 6, pp. 631–3, 2002.
- [27] E. Naumova, A. Egorov, R. Morris, and J. Griffiths, "The elderly and waterborne *Cryptosporidium* infection: gastroenteritis hospitalization before and during the 1993 Milwaukee outbreak.," *Emerg Infect Dis.*, 2003.

- [28] B. Dixon, L. Parrington, A. Cook, F. Pollari, and J. Farber, "Detection of *Cyclospora*, *Cryptosporidium*, and *Giardia* in ready-to-eat packaged leafy greens in Ontario, Canada," *J. Food Prot.*, vol. 76, no. 2, pp. 307–13, Feb. 2013.
- [29] B. V. Maikai, E. B. T. Baba-Onoja, and I. A. Elisha, "Contamination of raw vegetables with *Cryptosporidium* oocysts in markets within Zaria metropolis, Kaduna State, Nigeria," *Food Control*, vol. 31, no. 1, pp. 45–48, May 2013.
- [30] A. K. Rai, R. Chakravorty, and J. Paul, "Detection of *Giardia*, *Entamoeba*, and *Cryptosporidium* in unprocessed food items from northern India," *World J. Microbiol. Biotechnol.*, vol. 24, no. 12, pp. 2879–2887, Aug. 2008.
- [31] V. Chandra, M. Torres, and Y. R. Ortega, "Efficacy of wash solutions in recovering *Cyclospora cayetanensis*, *Cryptosporidium parvum*, and *Toxoplasma gondii* from basil," *J. Food Prot.*, vol. 77, no. 8, pp. 1348–54, Aug. 2014.
- [32] E. F. Donaldson, L. C. Lindesmith, A. D. LoBue, and R. S. Baric, "Viral shape-shifting: norovirus evasion of the human immune system," *Nat. Rev. Microbiol.*, vol. 8, no. 3, pp. 231–241, Feb. 2010.
- [33] K. Debbink, L. C. Lindesmith, E. F. Donaldson, and R. S. Baric, "Norovirus immunity and the great escape," *PLoS Pathog.*, vol. 8, no. 10, p. e1002921, Jan. 2012.
- [34] M. M. Patel, M.-A. Widdowson, R. I. Glass, K. Akazawa, J. Vinjé, and U. D. Parashar, "Systematic literature review of role of noroviruses in sporadic gastroenteritis," *Emerg. Infect. Dis.*, vol. 14, no. 8, pp. 1224–31, Aug. 2008.

- [35] L. Baert, K. Mattison, F. Loisy-Hamon, J. Harlow, A. Martyres, B. Lebeau, A. Stals, E. Van Coillie, L. Herman, and M. Uyttendaele, "Review: Norovirus prevalence in Belgian, Canadian and French fresh produce: A threat to human health?," *Int. J. Food Microbiol.*, vol. 151, no. 3, pp. 261–9, Dec. 2011.
- [36] J.-Y. Bae, J.-S. Lee, M.-H. Shin, S.-H. Lee, and I.-G. Hwang, "Effect of wash treatments on reducing human norovirus on Iceberg lettuce and perilla leaf," *J. Food Prot.*, vol. 74, no. 11, pp. 1908–11, Nov. 2011.
- [37] L. Baert, I. Vandekinderen, F. Devlieghere, E. Van Coillie, J. Debevere, and M. Uyttendaele, "Efficacy of sodium hypochlorite and peroxyacetic acid to reduce murine norovirus 1, b40-8, *Listeria monocytogenes*, and *Escherichia coli* O157:H7 on shredded Iceberg lettuce and in residual wash water," *J. Food Prot.*, vol. 72, no. 5, pp. 1047–54, 2009.
- [38] E. Scallan, R. Hoekstra, F. Angulo, R. Tauxe, M. Widdowson, S. Roy, J. Jones, and P. Griffin, "Foodborne illness acquired in the United States--major pathogens," *Emerg Infect Dis*, vol. 17, no. 1, pp. 7–15, 2011.
- [39] B. Guzman-Herrador, L. Vold, H. Comelli, E. MacDonald, B. T. Heier, A. L. Wester, T. L. Stavnes, L. Jensvoll, A. Lindegard Aanstad, G. Severinsen, J. Aasgaard Grini, Ø. Werner Johansen, K. Cudjoe, and K. Nygard, "Outbreak of *Shigella sonnei* infection in Norway linked to consumption of fresh basil, October 2011.," *Euro Surveill.*, vol. 16, no. 44, Jan. 2011.
- [40] B. S. M. Mahmoud, "Effects of X-ray radiation on *Escherichia coli* O157:H7, *Listeria monocytogenes*, *Salmonella enterica* and *Shigella flexneri* inoculated on

- shredded Iceberg lettuce.," *Food Microbiol.*, vol. 27, no. 1, pp. 109–14, Feb. 2010.
- [41] B. S. M. Mahmoud, G. Bachman, and R. H. Linton, "Inactivation of *Escherichia coli* O157:H7, *Listeria monocytogenes*, *Salmonella enterica* and *Shigella flexneri* on spinach leaves by X-ray.," *Food Microbiol.*, vol. 27, no. 1, pp. 24–8, Feb. 2010.
- [42] R. Singla, A. Ganguli, and M. Ghosh, "An effective combined treatment using malic acid and ozone inhibits *Shigella* spp. on sprouts," *Food Control*, vol. 22, no. 7, pp. 1032–1039, Jul. 2011.
- [43] Y.-W. In, J.-J. Kim, H.-J. Kim, and S.-W. Oh, "Antimicrobial activities of acetic acid, citric acid and lactic acid against *Shigella* species," *J. Food Saf.*, vol. 33, no. 1, pp. 79–85, Feb. 2013.
- [44] E. B. Solomon, S. Yaron, and K. R. Matthews, "Transmission of *Escherichia coli* O157:H7 from contaminated manure and irrigation water to lettuce plant tissue and its subsequent internalization," *Appl. Environ. Microbiol.*, vol. 68, no. 1, pp. 397–400, Jan. 2002.
- [45] G. Wang and M. P. Doyle, "Survival of enterohemorrhagic *Escherichia coli* O157:H7 in water.," *J. Food Prot.*, vol. 61, no. 6, pp. 662–7, Jun. 1998.
- [46] D. Weller, M. Wiedmann, and L. K. Strawn, "Irrigation Is Significantly Associated with an Increased Prevalence of *Listeria monocytogenes* in Produce Production Environments in New York State.," *J. Food Prot.*, vol. 78, no. 6, pp. 1132–41, Jun. 2015.

- [47] M. L. Ndiaye, Y. Dieng, S. Niang, H. R. Pfeifer, M. Tonolla, and R. Peduzzi, "Effect of irrigation water on the incidence of *Salmonella* spp. on lettuces produced by urban agriculture and sold on the markets in Dakar, Senegal." *Academic Journals*, 2011.
- [48] K. Holvoet, I. Sampers, M. Seynnaeve, and M. Uyttendaele, "Relationships among hygiene indicators and enteric pathogens in irrigation water, soil and lettuce and the impact of climatic conditions on contamination in the lettuce primary production.," *Int. J. Food Microbiol.*, vol. 171, pp. 21–31, Feb. 2014.
- [49] B. A. Macler and J. C. Merkle, "Current knowledge on groundwater microbial pathogens and their control," *Hydrogeol. J.*, vol. 8, no. 1, pp. 29–40, Mar. 2000.
- [50] C. L. Meays, K. Broersma, R. Nordin, and A. Mazumder, "Source tracking fecal bacteria in water: a critical review of current methods.," *J. Environ. Manage.*, vol. 73, no. 1, pp. 71–9, Oct. 2004.
- [51] L. A. Jones, R. W. Worobo, and C. D. Smart, "Plant-pathogenic oomycetes, *Escherichia coli* strains, and *Salmonella* spp. Frequently found in surface water used for irrigation of fruit and vegetable crops in New York State.," *Appl. Environ. Microbiol.*, vol. 80, no. 16, pp. 4814–20, Aug. 2014.
- [52] C. Mukhopadhyay, S. Vishwanath, V. K. Eshwara, S. A. Shankaranarayana, and A. Sagir, "Microbial quality of well water from rural and urban households in Karnataka, India: a cross-sectional study.," *J. Infect. Public Health*, vol. 5, no. 3, pp. 257–62, Jun. 2012.
- [53] R. de Quadros Rodrigues, M. R. Loiko, C. Minéia Daniel de Paula, C. T. Hessel, L.

- Jacxsens, M. Uyttendaele, R. J. Bender, and E. C. Tondo, "Microbiological contamination linked to implementation of good agricultural practices in the production of organic lettuce in southern Brazil," *Food Control*, vol. 42, pp. 152–164, Aug. 2014.
- [54] E. B. Solomon, C. J. Potenski, and K. R. Matthews, "Effect of irrigation method on transmission to and persistence of *Escherichia coli* O157:H7 on lettuce.," *J. Food Prot.*, vol. 65, no. 4, pp. 673–6, Apr. 2002.
- [55] E. B. Solomon, H.-J. Pang, and K. R. Matthews, "Persistence of *Escherichia coli* O157:H7 on lettuce plants following spray irrigation with contaminated water," *J. Food Prot.*, vol. 66, no. 12, pp. 2198–202, Dec. 2003.
- [56] J. D. Wood, G. S. Bezanson, R. J. Gordon, and R. Jamieson, "Population dynamics of *Escherichia coli* inoculated by irrigation into the phyllosphere of spinach grown under commercial production conditions.," *Int. J. Food Microbiol.*, vol. 143, no. 3, pp. 198–204, Oct. 2010.
- [57] M. Islam, M. P. Doyle, S. C. Phatak, P. Millner, and X. Jiang, "Survival of *Escherichia coli* O157:H7 in soil and on carrots and onions grown in fields treated with contaminated manure composts or irrigation water," *Food Microbiol.*, vol. 22, no. 1, pp. 63–70, Jan. 2005.
- [58] M. Islam, J. Morgan, M. P. Doyle, S. C. Phatak, P. Millner, and X. Jiang, "Fate of *Salmonella enterica* Serovar Typhimurium on carrots and radishes hrown in fields treated with contaminated manure composts or irrigation water," *Appl. Environ. Microbiol.*, vol. 70, no. 4, pp. 2497–2502, Apr. 2004.

- [59] D. T. Ingram, J. Patel, and M. Sharma, "Effect of repeated irrigation with water containing varying levels of total organic carbon on the persistence of *Escherichia coli* O157:H7 on baby spinach.," *J. Food Prot.*, vol. 74, no. 5, pp. 709–17, May 2011.
- [60] M. C. Erickson, C. C. Webb, J. C. Diaz-Perez, S. C. Phatak, J. J. Silvoy, L. Davey, A. S. Payton, J. Liao, L. Ma, and M. P. Doyle, "Surface and internalized *Escherichia coli* O157:H7 on field-grown spinach and lettuce treated with spray-contaminated irrigation water.," *J. Food Prot.*, vol. 73, no. 6, pp. 1023–9, Jun. 2010.
- [61] J. Wei, Y. Jin, T. Sims, and K. E. Kniel, "Manure- and biosolids-resident murine norovirus 1 attachment to and internalization by Romaine lettuce.," *Appl. Environ. Microbiol.*, vol. 76, no. 2, pp. 578–83, Jan. 2010.
- [62] USDA, "Harmonized GAP," *USDA-Agricultural Marketing Service*, 2011.
[Online]. Available: <https://www.ams.usda.gov/services/auditing/gap-ghp/harmonized>. [Accessed: 13-Feb-2016].
- [63] "Guide to Minimize Microbial Food Safety Hazards for Fresh Fruits and Vegetables." [Online]. Available:
<http://www.fda.gov/downloads/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/ProduceandPlanProducts/UCM169112.pdf>.
[Accessed: 24-May-2014].
- [64] M. T. Jay-Russell, J. E. Madigan, Y. Bengson, S. Madigan, A. F. Hake, J. E. Foley, and B. A. Byrne, "*Salmonella* Oranienburg isolated from horses, wild turkeys

and an edible home garden fertilized with raw horse manure.," *Zoonoses Public Health*, vol. 61, no. 1, pp. 64–71, Feb. 2014.

- [65] M. C. Erickson, J. Liao, L. Ma, X. Jiang, and M. P. Doyle, "Thermal and nonthermal factors affecting survival of *Salmonella* and *Listeria monocytogenes* in animal manure-based compost mixtures.," *J. Food Prot.*, vol. 77, no. 9, pp. 1512–8, Sep. 2014.
- [66] J. Kim, F. Luo, and X. Jiang, "Factors impacting the regrowth of *Escherichia coli* O157:H7 in dairy manure compost.," *J. Food Prot.*, vol. 72, no. 7, pp. 1576–84, Jul. 2009.
- [67] M. C. Erickson, C. Smith, X. Jiang, I. D. Flitcroft, and M. P. Doyle, "Survival of *Salmonella* or *Escherichia coli* O157:H7 during holding of manure-based compost mixtures at sublethal temperatures as influenced by the carbon amendment.," *J. Food Prot.*, vol. 78, no. 2, pp. 248–55, Feb. 2015.
- [68] L. M. Avery, P. Booth, C. Campbell, D. Tompkins, and R. L. Hough, "Prevalence and survival of potential pathogens in source-segregated green waste compost.," *Sci. Total Environ.*, vol. 431, pp. 128–38, Aug. 2012.
- [69] G. Mootian, W.-H. Wu, and K. R. Matthews, "Transfer of *Escherichia coli* O157:H7 from soil, water, and manure contaminated with low numbers of the pathogen to lettuce plants.," *J. Food Prot.*, vol. 72, no. 11, pp. 2308–12, Nov. 2009.
- [70] D. Ongeng, G. A. Vasquez, C. Muyanja, J. Ryckeboer, A. H. Geeraerd, and D. Springael, "Transfer and internalisation of *Escherichia coli* O157:H7 and

- Salmonella enterica* serovar Typhimurium in cabbage cultivated on contaminated manure-amended soil under tropical field conditions in Sub-Saharan Africa,” *Int. J. Food Microbiol.*, vol. 145, no. 1, pp. 301–10, Jan. 2011.
- [71] M. Jay-Russell, “What is the risk from wild animals in food-borne pathogen contamination of plants?,” *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.*, vol. 8, no. 40, pp. 1–16, Dec. 2013.
- [72] M. Foti, A. Daidone, A. Aleo, A. Pizzimenti, C. Giacobello, and C. Mammina, “*Salmonella bongori* 48:z35:- in migratory birds, Italy,” *Emerg. Infect. Dis.*, vol. 15, no. 3, pp. 502–3, Mar. 2009.
- [73] M. Cooley, D. Carychao, L. Crawford-Miksza, M. T. Jay, C. Myers, C. Rose, C. Keys, J. Farrar, and R. E. Mandrell, “Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California,” *PLoS One*, vol. 2, no. 11, p. e1159, Jan. 2007.
- [74] J. L. Talley, A. C. Wayadande, L. P. Wasala, A. C. Gerry, J. Fletcher, U. DeSilva, and S. E. Gilliland, “Association of *Escherichia coli* O157:H7 with filth flies (muscidae and calliphoridae) captured in leafy greens fields and experimental transmission of *E. coli* O157:H7 to spinach leaves by house flies (diptera: muscidae),” *J. Food Prot.*, 2009.
- [75] M. C. Erickson, J. Liao, A. S. Payton, D. G. Riley, C. C. Webb, L. E. Davey, S. Kimbrel, L. Ma, G. Zhang, I. Flitcroft, M. P. Doyle, and L. R. Beuchat, “Preharvest internalization of *Escherichia coli* O157:H7 into lettuce leaves, as affected by insect and physical damage,” *J. Food Prot.*, vol. 73, no. 10, pp. 1809–16, Oct.

2010.

- [76] S. Park, S. Navratil, A. Gregory, A. Bauer, I. Srinath, M. Jun, B. Szonyi, K. Nightingale, J. Anciso, and R. Ivanek, "Generic *Escherichia coli* contamination of spinach at the preharvest stage: effects of farm management and environmental factors.," *Appl. Environ. Microbiol.*, vol. 79, no. 14, pp. 4347–58, Jul. 2013.
- [77] FDA, "Importing food products into the United States - The imported seafood safety program," 2015. [Online]. Available: <http://www.fda.gov/Food/GuidanceRegulation/ImportsExports/Importing/ucm248706.htm>. [Accessed: 09-Dec-2015].
- [78] A. Mukherjee, D. Speh, E. Dyck, and F. Diez-Gonzalez, "Preharvest evaluation of coliforms, *Escherichia coli*, *Salmonella*, and *Escherichia coli* O157:H7 in organic and conventional produce grown by Minnesota farmers.," *J. Food Prot.*, vol. 67, no. 5, pp. 894–900, May 2004.
- [79] I. Jongenburger, H. M. W. den Besten, and M. H. Zwietering, "Statistical aspects of food safety sampling.," *Annu. Rev. Food Sci. Technol.*, vol. 6, pp. 479–503, Jan. 2015.
- [80] S. Dahms, "Microbiological sampling plans –Statistical aspects," *Mitt. Leb. Hyg.*, vol. 95, pp. 32–44, 2004.
- [81] European Commission, "Commission regulation (EC) No 1441/2007 on microbiological criteria for foodstuffs," *Off. J. Eur. Union*, p. 1441/2007, L 322/312-L322/328, 2007.

- [82] I. Jongenburger, M. W. Reij, E. P. J. Boer, L. G. M. Gorris, and M. H. Zwietering, "Random or systematic sampling to detect a localised microbial contamination within a batch of food," *Food Control*, vol. 22, no. 8, pp. 1448–1455, Aug. 2011.
- [83] M. Rivas Casado, D. J. Parsons, R. M. Weightman, N. Magan, and S. Origgi, "Modelling a two-dimensional spatial distribution of mycotoxin concentration in bulk commodities to design effective and efficient sample selection strategies," *Food Addit. Contam. Part A*, vol. 26, no. 9, pp. 1298–1305, Sep. 2009.
- [84] C. Habraken, D. Mossel, and S. van der Reek, "Management of *Salmonella* risks in the production of powdered milk products," *Netherlands Milk Dairy J.*, vol. 40, pp. 99–116, 1986.
- [85] G. Zehnder, "Overview of Monitoring and Identification Techniques for Insect Pests - eXtension," *Extension*, 2014. [Online]. Available: <http://www.extension.org/pages/19198/overview-of-monitoring-and-identification-techniques-for-insect-pests#.ViqhaBNViko>. [Accessed: 23-Oct-2015].
- [86] A. Paetz and B.-M. Wilke, *Monitoring and Assessing Soil Bioremediation*, vol. 5. Berlin/Heidelberg: Springer-Verlag, 2005.
- [87] S. C. Marine, S. Pagadala, F. Wang, D. M. Pahl, M. V Melendez, W. L. Kline, R. A. Oni, C. S. Walsh, K. L. Everts, R. L. Buchanan, and S. A. Micallef, "The growing season, but not the farming system, is a food safety risk determinant for leafy

- greens in the mid-Atlantic region of the United States.," *Appl. Environ. Microbiol.*, vol. 81, no. 7, pp. 2395–407, Apr. 2015.
- [88] I. Jongenburger, M. W. Reij, E. P. J. Boer, L. G. M. Gorris, and M. H. Zwietering, "Random or systematic sampling to detect a localised microbial contamination within a batch of food," *Food Control*, vol. 22, no. 8, pp. 1448–1455, Aug. 2011.
- [89] I. Jongenburger, M. W. Reij, E. P. J. Boer, L. G. M. Gorris, and M. H. Zwietering, "Actual distribution of *Cronobacter* spp. in industrial batches of powdered infant formula and consequences for performance of sampling strategies.," *Int. J. Food Microbiol.*, vol. 151, no. 1, pp. 62–9, Nov. 2011.
- [90] I. Jongenburger, H. M. W. den Besten, and M. H. Zwietering, "Statistical aspects of food safety sampling.," *Annu. Rev. Food Sci. Technol.*, vol. 6, pp. 479–503, Jan. 2015.
- [91] D. Zagory, "Microbial Testing of Fresh Produce: Where is the Value? | Food Safety News," 2014. [Online]. Available: <http://www.foodsafetynews.com/2014/10/microbial-testing-of-fresh-produce-where-is-the-value/#.VsEM5zYrI0o>. [Accessed: 14-Feb-2016].
- [92] S. Park, S. Navratil, A. Gregory, A. Bauer, I. Srinath, M. Jun, B. Szonyi, K. Nightingale, J. Anciso, and R. Ivanek, "Generic *Escherichia coli* contamination of spinach at the preharvest stage: effects of farm management and environmental factors.," *Appl. Environ. Microbiol.*, vol. 79, no. 14, pp. 4347–58, Jul. 2013.

- [93] P. Battilani, C. Barbano, V. Rossi, T. Bertuzzi, and A. Pietri, "Spatial distribution of ochratoxin A in vineyard and sampling design to assess must contamination.," *J. Food Prot.*, vol. 69, no. 4, pp. 884–90, Apr. 2006.
- [94] C. S. LIN, C.S., POUHINSKY, G., MAUER, M., 1979. An Examination of Five Sampling Methods under Random and Clustered Disease Distributions Using Simulation. *Can. J. Plant Sci.* 59, 121–130. doi:10.4141/cjps79-017LIN, G. POUHINSKY, and M. MAUER, "An Examination of Five Sampling Methods under Random and Clustered Disease Distributions Using Simulation," *Can. J. Plant Sci.*, vol. 59, no. 1, pp. 121–130, Jan. 1979.
- [95] M. Rivas Casado, D. J. Parsons, R. M. Weightman, N. Magan, and S. Oraggi, "Modelling a two-dimensional spatial distribution of mycotoxin concentration in bulk commodities to design effective and efficient sample selection strategies," *Food Addit. Contam. Part A*, vol. 26, no. 9, pp. 1298–1305, Sep. 2009.
- [96] C. J. M. Habraken, D. A. A. Mossel, and D. Reek S. van, "Management of Salmonella risks in the production of powdered milk products.," *Netherlands Milk Dairy J.*, 1986.
- [97] CDC, "FoodNet 2015 Preliminary Data| FoodNet | CDC," 2015. [Online]. Available: <https://www.cdc.gov/foodnet/reports/prelim-data-intro.html>. [Accessed: 06-Mar-2017].
- [98] FDA, "2007 - FDA Finalizes Report on 2006 Spinach Outbreak." Office of the Commissioner, 2007.

- [99] S. C. Marine, S. Pagadala, F. Wang, D. M. Pahl, M. V Melendez, W. L. Kline, R. A. Oni, C. S. Walsh, K. L. Everts, R. L. Buchanan, and S. A. Micallef, "The growing season, but not the farming system, is a food safety risk determinant for leafy greens in the mid-Atlantic region of the United States.," *Appl. Environ. Microbiol.*, vol. 81, no. 7, pp. 2395–407, Apr. 2015.
- [100] E. van Nood, A. Vrieze, M. Nieuwdorp, S. Fuentes, E. G. Zoetendal, W. M. de Vos, C. E. Visser, E. J. Kuijper, J. F. W. M. Bartelsman, J. G. P. Tijssen, P. Speelman, M. G. W. Dijkgraaf, and J. J. Keller, "Duodenal Infusion of Donor Feces for Recurrent *Clostridium difficile*," *N. Engl. J. Med.*, vol. 368, no. 5, pp. 407–415, Jan. 2013.
- [101] A. Xu, R. L. Buchanan, and S. A. Micallef, "Impact of mulches and growing season on indicator bacteria survival during lettuce cultivation," *Int. J. Food Microbiol.*, vol. 224, pp. 28–39, May 2016.
- [102] E. D. Berry, J. E. Wells, J. L. Bono, B. L. Woodbury, N. Kalchayanand, K. N. Norman, T. V Suslow, G. López-Velasco, and P. D. Millner, "Effect of proximity to a cattle feedlot on *Escherichia coli* O157:H7 contamination of leafy greens and evaluation of the potential for airborne transmission.," *Appl. Environ. Microbiol.*, vol. 81, no. 3, pp. 1101–10, Feb. 2015.
- [103] *Impact of microbial distributions on food safety*. 2010.
- [104] M. Jay, M. Cooley, D. Carychao, G. Wiscomb, R. Sweitzer, L. Crawford-Miksza, J. Farrar, D. Lau, J. O'Connell, A. Millington, R. Asmundson, E. Atwill, and R. Mandrell, "*Escherichia coli* O157:H7 in Feral Swine near Spinach Fields and

Cattle, Central California Coast. ." *Emerg Infect Diseases*, vol.13, no.12,
Dec.2007

- [105] R. Olea, *Geostatistics for engineers and earth scientists*. Kluwer Academic Publishers, 1999.
- [106] J. Sacks, W. Welch, T. Mitchell, and H. Whnn, "Design and analysis of computer experiments (with discussion)," *Stat. Sci.*, vol. 4, pp. 409–435, 1989.
- [107] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient Global Optimization of Expensive Black-Box Functions," *J. Glob. Optim.*, vol. 13, no. 4, pp. 455–492, 1998.
- [108] J. Wang, S. C. Clark, E. Liu, and P. I. Frazier, "Parallel Bayesian Global Optimization of Expensive Functions," Feb. 2016.
- [109] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [110] J. Sacks, W. Welch, T. Mitchell, and H. Whnn, "Design and analysis of computer experiments (with discussion)," *Stat. Sci.*, vol. 4, pp. 409–435, 1989.
- [111] S. Clark, E. Liu, P. Frazier, J. Wang, D. Oktay, and N. Vesdapunt, "Metrics optimization engine," 2014. .
- [112] A. Touron, T. Berthe, G. Gargala, M. Fournier, M. Ratajczak, P. Servais, and F. Petit, "Assessment of faecal contamination and the relationship between pathogens and faecal bacterial indicators in an estuarine environment (Seine, France)," *Mar. Pollut. Bull.*, vol. 54, no. 9, pp. 1441–1450, Sep. 2007.

- [113] M. A. Moriñigo, R. Córna, M. A. Muñoz, P. Romero, and J. J. Borrego, "Relationships between *salmonella* spp and indicator microorganisms in polluted natural waters," *Water Res.*, vol. 24, no. 1, pp. 117–120, Jan. 1990.
- [114] J. Wu, S. C. Long, D. Das, and S. M. Dorner, "Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research."
- [115] A. K. Deisingh and M. Thompson, "Strategies for the detection of *Escherichia coli* O157:H7 in foods," *J. Appl. Microbiol.*, vol. 96, no. 3, pp. 419–29, 2004.
- [116] Z. Shen, J. Wang, Z. Qiu, M. Jin, X. Wang, Z. Chen, J. Li, and F. Cao, "QCM immunosensor detection of *Escherichia coli* O157:H7 based on beacon immunomagnetic nanoparticles and catalytic growth of colloidal gold," *Biosens. Bioelectron.*, vol. 26, no. 7, pp. 3376–3381, Mar. 2011.
- [117] Ö. Torun, İ. Hakkı Boyacı, E. Temür, and U. Tamer, "Comparison of sensing strategies in SPR biosensor for rapid and sensitive enumeration of bacteria," *Biosens. Bioelectron.*, vol. 37, no. 1, pp. 53–60, Aug. 2012.
- [118] A. Cundell, "Opportunities for Rapid Microbial Methods," *Eur. Pharm. Rev.*, vol. 1, pp. 64–70, 2006.
- [119] M. W. Griffiths, "Rapid microbiological methods with hazard analysis critical control point," *J. AOAC Int.*, vol. 80, no. 6, pp. 1143–50.
- [120] S. Parveen, S. Kaur, S. A. W. David, J. L. Kenney, W. M. McCormick, and R. K. Gupta, "Evaluation of growth based rapid microbiological methods for sterility testing of vaccines and other biological products," *Vaccine*, vol. 29, no.

45, pp. 8012–8023, Oct. 2011.

- [121] M. Miller, “The Implementation of Rapid Microbiological Methods,” *Eur. Pharm. Rev.*, pp. 24–26, 2010.
- [122] W. Lee, D. Kwon, W. Choi, G. Y. Jung, A. K. Au, A. Folch, and S. Jeon, “3D-Printed Microfluidic Device for the Detection of Pathogenic Bacteria Using Size-based Separation in Helical Channel with Trapezoid Cross-Section,” *Sci. Rep.*, vol. 5, no. 1, p. 7717, Jul. 2015.
- [123] E. Greb, “An Overview of Rapid Microbial-Detection Methods,” *Pharm. Technol.*, vol. 2010 Suppl, no. 1, 2010.
- [124] S. Sutton, “Accuracy of Plate Counts,” *J. Valid. Technol.*, vol. 17, no. 3, pp. 42–46, 2011.