RESEARCH IN PROCESS

Simulation analysis of in-field produce sampling for risk-based sampling plan development

SUMMARY

Effective preharvest, field-level produce sampling is challenging because current practices typically yield few positive samples with fields rarely re-testing positive. Statistical theory suggests one reason is that detecting rare contamination events would require 100s to 1,000s of random samples, or targeted sampling of higher risk locations in fields. This project will develop and validate tools for the produce industry to evaluate exiting and improved produce field sampling plans. Results will be used to communicate to growers the number and location of samples needed to achieve a known power to detect contamination. We will validate these simulations against academic literature, industry partner data, and field-trials of controlled contamination of spinach. Our project will provide growers with tools to (i) develop improved sampling plans, (ii) customize those plans to their individual fields, and (iii) quantify the performance and costs of the plan – all to better identify and manage preharvest food safety risks.

OBJECTIVES

1. Simulate contamination of produce fields that are representative of commercial fields in four produce-growing regions of the United States.
2. Evaluate convenience, improved generic, and risk-based sampling plans.
3. Validate simulations against data from industry partners and academic literature.
4. Validate simulations against field-trials of controlled contamination events.

METHODS

We will simulate (scope in **Table 1**) a **produce field** and **product**, contaminated by a **food safety hazard**. A **sampling plan** will pass said product to a laboratory **assay** to determine + and − **outcomes**. These simulation results will then be used to evaluate sampling plans, and be validated against existing industry data and experimental data.

To the *use* the generic simulation model, we will simulate produce fields representative of (i) the Central Valley, CA; (ii) Yuma, AZ; (iii) the Delmarva Peninsula; and (iv) Upstate NY, through site visits and expert elicitation. We will simulate typical contamination for each field, including: (i) point sources, e.g. fecal deposits; (ii) systematic sources, e.g. contaminated irrigation water; and (iii) sporadic contamination, e.g. endemic soil bacteria. Finally, we will validate these results against data, including from experimentally contaminated spinach fields subject to high-resolution sampling.

RESULTS TO DATE

The Illinois team has programmed the generic produce field simulation (**Figure 1**). With our existing generic field simulation we are able to evaluate generic sampling plans such as n_{60} composite random sampling, stratified random sampling, and composite sampling (**Figure 2**). We are currently developing code for convenience sampling plans (such as Z-pattern sampling) and risk-based sampling. The Illinois team's graduate student has begun a literature review to extract parameters relevant to risk-based sampling, including the impact on foodborne pathogen prevalence of water practices (irrigation, surface water, rainfall), soil (amendments, presence of indicators), and landscape features (proximity to natural areas or other agricultural production). Our industry partner connected us with a data analyst who has provided parameters for four typical contamination scenarios in their fields: background contamination, small cluster, larger cluster, and major events.

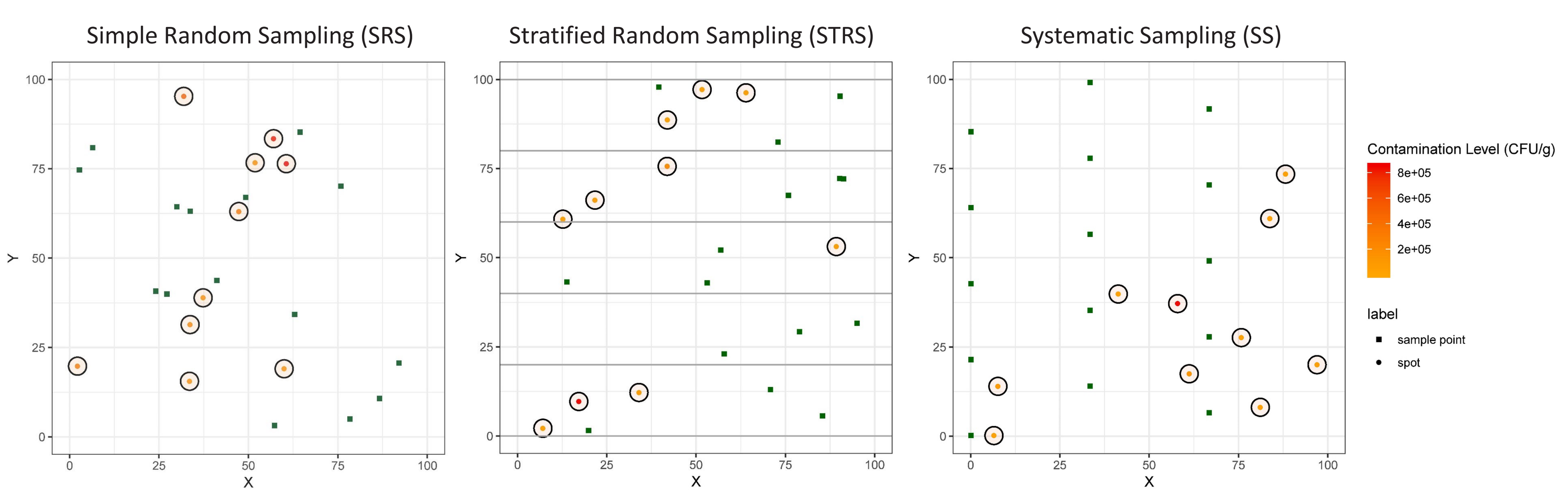

Figure 1. Example simulation of a generic 100-unit by 100-unit produce field with randomly located point sources of contamination, each with a 2-unit radius of spread, and 15 samples selected by different sampling schemes. The simulation predicts the contamination level at each sampling point based on the distance to contamination points (such as fecal droppings) and decay out to the indicated radius of spread. Then the individual sample points are composited for enrichment testing, and the simulation outputs if the hazard would be detected. Our next steps are to parameterize these simulations to represent the real geometry of fields, incorporate convenience (Z-pattern) and risk-based sampling, and represent likely hazards.

BENEFITS TO THE INDUSTRY

The key beneficiaries of this project are growers and those individuals who are responsible for pre-harvest produce sampling. Specifically, this project will provide beneficiaries with generic guidance for improving sampling to target specific types of contamination. We will also provide tools to help growers adapt generic guidance to field-specific, risk-based sampling plans. In the short-term, our findings will increase the understanding of the limitations of existing sampling plans – particularly since some current n_{60} composite plans are more a response to buyer requirements than to a science-based understanding of how to best detect contamination. It will also describe expected performance and resource requirements of improved sampling plans, which are critical to building the business case for investing additional resources in food safety risk management.

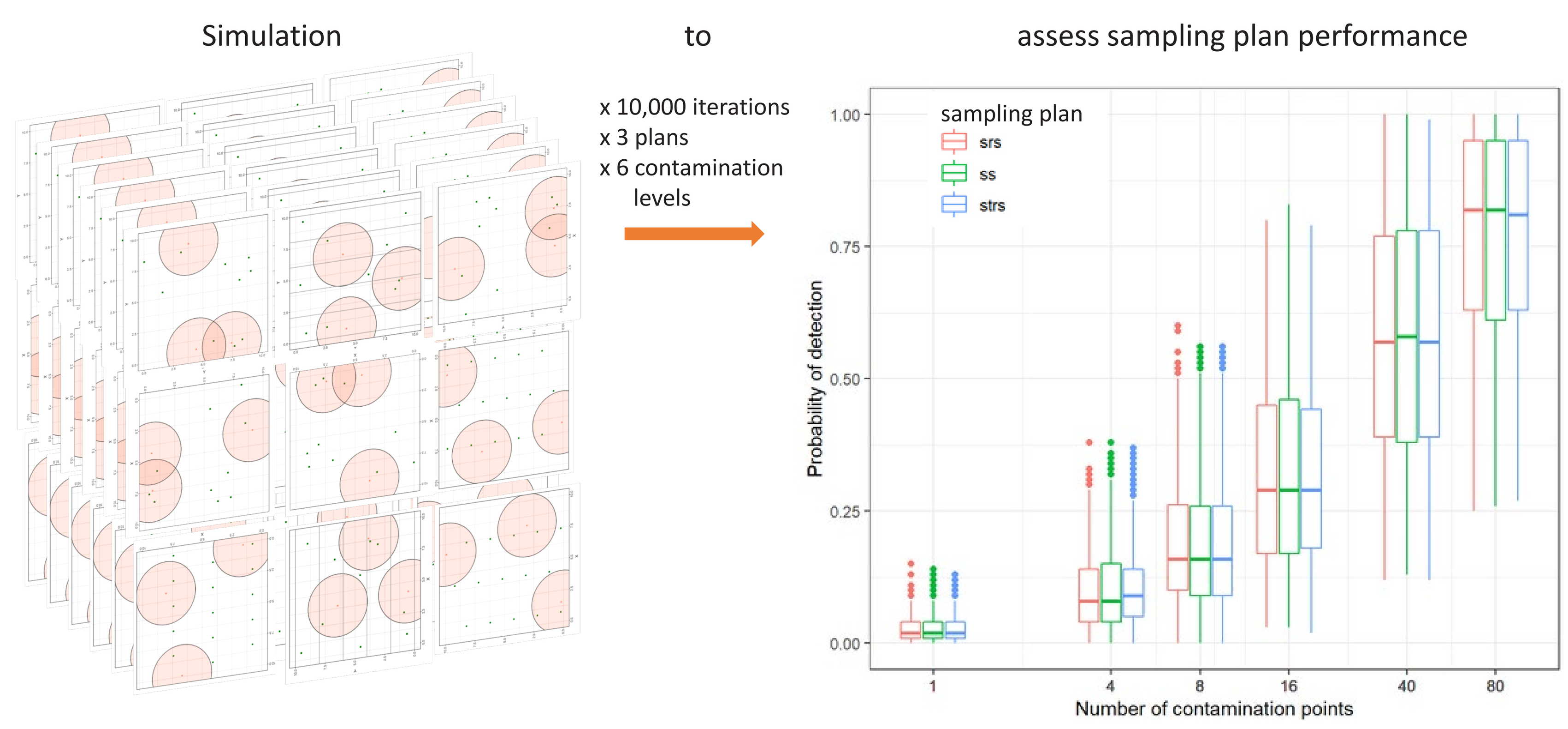

Figure 2. Here we (left) iterate a field model 10,000 times, each time simulating 3 different types of sampling plans and 6 different contamination levels, and then (right) compare the probability of detecting the hazard across these variables. In this example case, all three sampling plans perform equally well, and performance improves with the number of contamination points. Our next steps are to parameterize these simulations according to real-world scenarios to evaluate actual sampling plans of interest to the produce industry.

Simulation Domain	Parameters to Simulate	Variable Characteristics for Representative Fields
Produce Field	Geometry (field dimensions, field boundaries, slope) Risk Factors (landscape structure, non-crop buffers, irrigation system type and set-up, time since rain)	Field setup High- or low-risk features
Product	Location (seeding rate, row spacing) Plant characteristics (edible fraction, yield)	Major products (leafy greens, tomatoes); Planting patterns
Food Safety Hazard	Location and area contaminated Point source, e.g. feces Systematic source, e.g. irrigation Sporadic low-level, e.g. endemic soil bacteria Proportion of plants that are contaminated Pathogen concentration on contaminated plants	Pathogens of concern (e.g., *Salmonella* in tomatoes, STEC in leafy greens) Known contamination events Known research-testing results
Sampling Plan	Sampling Scope (number and size of composite samples) Sampling strategy Convenience: Z-shape Improved Generic: random, stratified, systematic Risk-Based: targeting field-specific high-risk areas	Typical sampling strategies Numbers Sizes Patterns
Assay *Performance*	Parameters (Limit of detection, sensitivity, specificity) Setup (Sample compositing, re-testing requirements) Inputs (Time per test, cost per test)	Typical assays Local assay costs, timing Retesting conditions, thresholds
Outcomes	Probability of detecting hazard above threshold Incremental cost for improved performance	Typical, ranges of sampling results Relative costs, efficacy

Table 1. Field simulation scope indicating a representative subset of the parameters that will be simulated.

CONTACT

Matthew Stasiewicz, PhD
University of Illinois at Urbana-Champaign
Department of Food Science and Human Nutrition
E: mstasie@illinois.edu

AUTHORS

Matthew Stasiewicz (PI)
Martin Wiedmann (Co-PI, Cornell University)

ACKNOWLEDGEMENTS:

We would like to thank the following students, staff, and collaborators: ILLINOIS – Jorge Quintanilla, Eric Cheng, Janice Trudell; CORNELL – Daniel Weller, Alexandra Belias, Shawna Sheehy; INDUSTRY – Jim Brennan, Eric Wilhelmsen

LENGTH OF FUNDING

January 1, 2019 – December 31, 2020

cps CENTER *for* PRODUCE SAFETY