AEM Journals.ASM.org

Farm Management, Environment, and Weather Factors Jointly Affect the Probability of Spinach Contamination by Generic *Escherichia coli* at the Preharvest Stage

Sangshin Park,[a] Sarah Navratil,[b,c] Ashley Gregory,[d] Arin Bauer,[d] Indumathi Srinath,[a,e] Barbara Szonyi,[a] Kendra Nightingale,[b,c] Juan Anciso,[d] Mikyoung Jun,[f] Daikwon Han,[g] Sara Lawhon,[h] Renata Ivanek[a]

Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, USA[a]; Department of Animal Sciences, Colorado State University, Fort Collins, Colorado, USA[b]; Department of Animal and Food Sciences, Texas Tech University, Lubbock, Texas, USA[c]; Department of Horticultural Sciences, Texas A&M AgriLife Extension Service, Weslaco, Texas, USA[d]; Tarleton State University, Stephenville, Texas, USA[e]; Department of Statistics, Texas A&M University, College Station, Texas, USA[f]; Department of Epidemiology and Biostatistics, Texas A&M University, College Station, Texas, USA[g]; Department of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, USA[h]

The National Resources Information (NRI) databases provide underutilized information on the local farm conditions that may predict microbial contamination of leafy greens at preharvest. Our objective was to identify NRI weather and landscape factors affecting spinach contamination with generic *Escherichia coli* individually and jointly with farm management and environmental factors. For each of the 955 georeferenced spinach samples (including 63 positive samples) collected between 2010 and 2012 on 12 farms in Colorado and Texas, we extracted variables describing the local weather (ambient temperature, precipitation, and wind speed) and landscape (soil characteristics and proximity to roads and water bodies) from NRI databases. Variables describing farm management and environment were obtained from a survey of the enrolled farms. The variables were evaluated using a mixed-effect logistic regression model with random effects for farm and date. The model identified precipitation as a single NRI predictor of spinach contamination with generic *E. coli*, indicating that the contamination probability increases with an increasing mean amount of rain (mm) in the past 29 days (odds ratio [OR] = 3.5). The model also identified the farm's hygiene practices as a protective factor (OR = 0.06) and manure application (OR = 52.2) and state (OR = 108.1) as risk factors. In cross-validation, the model showed a solid predictive performance, with an area under the receiver operating characteristic (ROC) curve of 81%. Overall, the findings highlighted the utility of NRI precipitation data in predicting contamination and demonstrated that farm management, environment, and weather factors should be considered jointly in development of good agricultural practices and measures to reduce produce contamination.

Produce-related food safety concerns have been on the rise due to reported large-scale outbreaks related to contaminated produce, including leafy greens (1, 2). The exact number of food-borne illnesses and outbreaks attributable to produce is unknown due to underreporting and difficulties in attributing food-borne illnesses to a particular food commodity (3, 4). However, based on food-borne disease outbreak data reported to the U.S. Centers for Disease Control and Prevention between 1998 and 2008, among almost 68,000 illnesses in outbreaks assigned to one of the 17 considered food commodities, the commodities associated with the most outbreak-related illnesses were poultry (17%), leafy vegetables (13%), beef (12%), and fruits/nuts (11%) (1). Not only were leafy greens responsible for a considerable proportion of food-borne illnesses, but the mean percentage of outbreaks attributed to leafy greens has been on the rise; it increased from 6% (1998 to 1999) to 11% (2006 to 2008) (1). Therefore, a reduction in the number of human food-borne cases attributable to leafy greens is of timely importance.

Enteric food-borne pathogens are shed into the environment through the feces of colonized or infected hosts, and *Listeria monocytogenes* is naturally found in soil. During their persistence in the environment, the pathogens may contaminate distant locations through diverse vehicles and activities, such as spreading of animal manure and contaminated irrigation water. Therefore, contamination of produce, including leafy greens, with these food-borne pathogens is affected by contamination events and the ability of the pathogen to survive in the environment. Contamination events may occur through routes such as application of raw or inadequately composted manure (5, 6), exposure to contaminated irrigation water (7, 8) or flooding (9), and unintentional deposition of feces by infected or carrier livestock or wildlife (8, 9). A pathogen's survivability is an inherent pathogen characteristic (10) that also varies depending on the environmental and weather conditions. For example, it has been reported that inactivation of enteric bacterial, viral, and protozoan pathogens in the environment may be affected by predation, competition, water stress/osmotic potential, temperature, UV radiation, pH, inorganic ammonia, and organic nutrients (11). In order to effectively control food-borne pathogens in leafy greens at the preharvest level, both the contamination routes and weather and environmental factors affecting pathogens' survivability should be considered.

While produce contamination with enteric food-borne pathogens is of high public health and economic concern, the contam-

Received 5 November 2013 Accepted 5 February 2014

Published ahead of print 7 February 2014

Editor: D. W. Schaffner

Address correspondence to Sangshin Park, dvm.spark@gmail.com.

doi:10.1128/AEM.03643-13

TABLE 1 Spinach sample collection scheme

State	Farm	Sampling dates	No. of samples: Collected	Positive
Colorado	1	July 26, Aug. 24, Sept. 7, 2010; May 31, July 11, Aug. 11, 2011	115	0
	2	June 7, July 6, Aug. 31, 2010; May 24, May 29, June 13, June 14, July 11, 2011	120	24
	3	June 21, July 12, Aug. 16, 2010; May 29, June 20, July 18, 2011	120	2
	4	June 13, Aug. 9, Sept. 25, 2010; June 4, July 4, Aug. 1, 2011	120	1
Texas	1	Nov. 19, 2010; Jan. 7, Feb. 7, Nov. 11, Dec. 2, 2011; Jan. 6, 2012	120	11
	2	Nov. 19, 2010; Jan. 7, Feb. 7, Nov. 11, Dec. 2, 2011; Jan. 6, 2012	120	4
	3	Dec. 3, 2010; Jan. 21, 2011	10	1
	4	Dec. 3, 2010; Jan. 21, Feb. 18, Mar. 4, Dec. 16, 2011; Jan. 20, Feb. 10, 2012	95	7
	5	Dec. 3, 2010; Jan. 21, Feb. 18, 2011	25	7
	6	Jan. 21, Feb. 18, Mar. 4, Dec. 16, 2011; Jan. 20, Feb. 10, 2012	60	1
	7	Dec. 3, 2010; Jan. 21, 2011	10	1
	8	Dec. 16, 2011; Jan. 20, 2012	40	4

ination events are relatively rare (6, 12, 13), thus requiring intensive but also expensive sampling and testing efforts. A common practice has been to use generic *Escherichia coli* as an indicator of fecal contamination of produce (5, 6, 13–15). Although generic *E. coli* can form stable populations in temperate soil and water environments (16–18), its survival is indicative of conditions favorable for the survival and persistence of pathogenic *E. coli* and *Salmonella* spp. (15, 19, 20). Data from our previous study confirmed the potential usefulness of generic *E. coli* as an index organism for the presence of *Salmonella* spp. (19). Other studies also suggested the usefulness of generic *E. coli* as an index organism for the presence of *E. coli* O157: H7 (20) and *Salmonella enterica* serovar Typhimurium (15). The utility of generic *E. coli* in evaluating the efficacy of the process to reduce the population of *E. coli* O157: H7 and *Salmonella* spp. was demonstrated (21). To improve the control of food-borne illnesses related to fresh leafy greens, it is meaningful to identify the risk factors for their contamination with generic *E. coli*.

Geographic information systems (GIS) integrated with standard statistical and epidemiological methods provide tremendous opportunities for research and control of food-borne pathogens (and indicators of fecal contamination) in produce (22). Only limited research and application efforts (23, 24) have been conducted in the United States and elsewhere to facilitate broad implementation of geospatial databases, methods, and technologies to improve produce food safety. Given that the role of the local weather and environmental factors should be considered to effectively control pathogen contamination of produce, the freely available National Resources Information (NRI) databases developed using GIS, may provide an abundant data source on the local landscape (topography, hydrography, soil characteristics, and road network) and weather conditions. Weather information from NRI databases may not represent the microscale conditions on the produce field but may still be very useful. NRI databases have successfully been used to study the epidemiology of food-borne diseases, such as identifying determinants of food-borne pathogen occurrences in the environment (25, 26). Formative research is needed to analyze whether or not NRI information can determine the probability of leafy green contamination when considered in isolation or jointly with farm management and environmental factors.

The identification of novel risk factors for contamination of leafy greens with food-borne pathogens is a key to developing effective strategies for their control so that microbial safety of leafy greens can be improved. In this study, we used spinach as a representative of leafy greens and generic *E. coli* as an indicator of fecal contamination. Our objectives were to (i) identify NRI weather and landscape factors that are associated with the probability of spinach contamination with generic *E. coli* prior to harvest and (ii) determine how these and farm management and environmental factors on a particular farm jointly affect the probability of spinach contamination. We addressed these objectives by applying spatial and statistical modeling to newly obtained data from NRI databases integrated with our previously described data on farm management and environmental factors and *E. coli* contamination of spinach on 12 produce farms (19).

MATERIALS AND METHODS

Spinach contamination data. The collection and microbiological testing of spinach samples have been described in detail in our previous study (19). Briefly, using a repeated cross-sectional study design, a total of 955 spinach samples was collected on 12 enrolled farms (4 in Colorado and 8 in Texas) during two spinach growing seasons between June 2010 and February 2012. Colorado and Texas were chosen as representative states of the Western and Southwestern United States, respectively. Spinach is best grown under relatively cool and dry conditions (18 to 24°C days and 4 to 7°C nights) (27), and so the spinach growing season in Texas is between November and March and that in Colorado is from April to September. Due to the different timing of the spinach growing seasons in these two states, the two states could also be viewed as representative of the spinach produced year round in the United States. In Texas, the enrolled farms were located in Cameron, Hidalgo, and Uvalde counties. In Colorado the farms were in Adams, Boulder, Larimer, and Saguache counties. Each farm was visited up to 4 times per growing season for a total of 2 to 8 sampling dates per farm over the study period (Table 1). During each farm visit, we collected 5 spinach samples from each of 1 to 6 spinach fields per farm. The five spinach samples were collected by selecting four spinach samples from each of the four corners of the field and the fifth sample from the field center. Each spinach sample consisted of at least 10 randomly selected individual plant leaves of different maturities, collected in an area within a 5-meter radius. The Global Positioning System (GPS) coordinates for the exact locations of spinach sample collections were recorded using a handheld GPS device (Garmin 12XL; Garmin Ltd., Olathe, KS). Samples were collected and placed into sterile Whirl-Pak bags (Nasco, Fort Atkinson, WI) using sterile gloves. They were shipped in coolers with ice packs to a laboratory and processed within 48 h. Twenty-five grams of spinach leaves was suspended in 75 ml phosphate-buffered saline (PBS) placed in stomacher bags. The contents of each bag were crushed using a blender (Smasher lab blender; AES-Chemunex, France), and then a 1-ml aliquot from the sample bag, followed by 1 ml of each of five 1:10 serial dilutions, was plated directly onto Petrifilm *E. coli*/coliform count plates (3 M Microbiology, St. Paul, MN). After incubation at 37°C for 48 h, the plates were visually assessed by counting blue to red-blue colonies with gas bubbles (according to the standard *E. coli* Petrifilm enumeration method [http://tmacog.org/Environment/SWW_07/PetrifilmInterpretation.pdf]). A spinach sample was considered contaminated with generic *E. coli* if at least one colony was observed. The approach had a detection limit of 4 CFU/ml of the plated dilution.

Spatial modeling of weather and landscape data. For each georeferenced location where spinach samples were collected, we obtained information from freely available NRI databases on potentially relevant weather and landscape factors following the general approach described in a study by Ivanek et al. (24). In total, we obtained information on 96 variables grouped under ambient temperature, precipitation, wind speed, soil properties, and distances to the nearest water body or road (Table 2). It was unclear if the average, minimum, or maximum daily ambient temperature would be a better predictor of the probability of spinach contamination, so we explored the potential role of all three of these temperature characteristics. Data on weather factors (temperature, precipitation, and wind speed) were obtained through the National Climatic Data Center (http://www.ncdc.noaa.gov/) based on information recorded at land-based weather stations. Specifically, for each sampled location and date of sampling, we used the nearest weather station that had recorded the particular weather information for the day or period of interest. Altogether, we used data from 22 weather stations (including 10 for temperature, 10 for wind speed, and 10 for precipitation information), which were located on average 11.9 km (range, 1.5 km to 34.7 km) away from the sampling locations. The effect of a considered weather factor on the probability of spinach contamination may occur instantly or may accumulate gradually over a period of time (24). If the effect occurs instantly, it is unknown whether we should be interested in the weather characteristics on the day of sample collection or on any particular day before that. Therefore, for each of the considered weather factors, we created 4 variables describing the particular weather factor on the day of sample collection and on days 1, 2, and 3 prior to sample collection. Likewise, if the effect of a weather factor accumulates over a period of time, it is unknown how long a period we should consider. Thus, we created additional 14 variables explaining the mean level of a particular weather factor for a period of time between the day of sample collection and days 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, and 29 prior to sample collection. We created and examined a total of 18 variables for each considered weather factor. All temperature measurements, recorded in degrees Fahrenheit (°F), were converted to degree Celsius (°C) using the conversion equation $°C = (°F - 32) \times 5/9$. The precipitation measurements were recorded in inches and converted to millimeters ($mm = in. \times 25.4$). The amount of rain recorded as "trace" was assigned a value of 0.0001 mm. Wind speed measurements were recorded in knots and converted to meters per second (m/s) using the equation $m/s = knots \times 0.514$. The weather variables and their notations are defined in Table 2. Additionally, we extracted information about wind gusts using the approach described for wind speed; however, the gust variables were not considered in statistical analyses because many values were missing. There were no missing values for wind speed, temperature, or precipitation variables.

NRI databases were also used to obtain information on landscape factors at the spinach sampling locations (Table 2). Landscape databases and GPS coordinates of spinach sample collection locations were imported into ArcGIS 10 (ESRI, Redland, CA) and reprojected into the Universal Transverse Mercator, North American Datum of 1983. Information about the local soil properties and distances to the nearest water bodies and roads for each sampled location were extracted from the overlay between the GPS coordinates and landscape layers. The soil properties (4 variables) were obtained from the Soil Survey Geographic (SSURGO) database (http://websoilsurvey.nrcs.usda.gov/). Distances (2 variables) to the nearest water body and road for locations in Texas were extracted from the National Hydrography and TxDOT Roadways data sets, respectively, obtained through the Texas Natural Resources Information System (http://www.tnris.org/). Information about the distances to the nearest water body and road for locations in Colorado was extracted from the Hydrography-1M and Transportation-1M data sets, respectively, obtained through the Colorado Department of Natural Resources (http://data.geocomm.com/catalog/US/61076/datalist.html). Some of the landscape variables had missing observations. Specifically, 35 observations were missing for the organic matter variable, and 25 observations were missing for each of the variables soil acidity, soil texture, and slope. Handling of missing observations is described "Statistical analyses" below.

Farm management and environmental factors. A survey of farm management and environmental factors has been previously described (19). Briefly, at the time of spinach sample collection, we used a questionnaire to survey farmers about the general farm-related management and environmental factors that were subsequently coded into a total of 76 explanatory variables (listed in Table 2 of reference 19). In the current study, a univariate statistical analysis was performed on 71 of these variables. We excluded the variables with suspected misinterpreted questions in the questionnaire survey ("portable toilet distances from the work area," "wildlife control," and "buffer zone with fence") (19). The variables for "terrain" and "proximity to the nearest road within 10-mile radius" from the survey (19) were replaced by variables describing the same landscape characteristics obtained from the NRI databases. Out of the 71 included variables, Table 2 lists and describes only those that were significant at the 20% level in univariate statistical analyses and were eligible for further statistical consideration. Three of the variables in Table 2 had missing values: "time since the last workers' visit," 265 missing values; "organic farming certified by the National Organic Program," 750 missing values; and "wildlife control by fences," 450 missing values. Handling of these missing observations is described in "Statistical analyses" below.

Statistical analyses. Statistical analyses were conducted using R (the R Project for Statistical Computing, http://www.r-project.org/). Except in univariate analyses, P values of <0.05 were considered significant. A liberal significance cutoff of 20% was used in univariate analyses to ensure that all potentially influential variables (including potential confounders) were evaluated in the multivariable analysis. For the same reason, a correction for multiple testing was not conducted. In univariate and multivariable modeling, associations between generic *E. coli* contamination of spinach (dependent variable) and individual explanatory (independent) variables were assessed using a mixed-effect logistic regression model with random effects for farm and date, implemented through the "lmer" function in the "lme4" package (28). The odds ratio (OR) was used as a measure of association. The phi coefficient was used to check for colinearity between two individual explanatory categorical variables (29). Spearman's rank correlation analyses were performed to assess correlations between individual explanatory variables when one or both of the explanatory variables were continuous. To assess similarity of the weather and landscape factors between the two states, the chi-square test (χ^2) for categorical data and the Wilcoxon rank sum test for continuous data were used. When any two independent variables considered for multivariable modeling had high colinearity or correlation ($>60\%$), these variables were considered one at a time in multivariable modeling.

The weather variables listed in Table 3 were standardized by subtracting the mean and dividing by the standard deviation for the variable. The resulting standardized variables were subjected to principal component analysis (PCA) to better understand the type and overlap of information contained in the weather variables. The number of meaningful components to retain was determined by considering the proportion of variance accounted for, the scree test, and the interpret-

TABLE 2 Description of the considered explanatory variables obtained from the NRI databases (weather and landscape factors) and from a survey of produce farmers (farm management and environmental factors)

Variable	Description and levels[f]	Unit
Weather factors[c]		
Temp		
tm$X$[a]	Avg daily temp on day of SC ($X = 0$) or day X prior to SC ($X = 1, 2, 3$)	°C
tmd$X$[b]	Mean of avg daily temp in the period between day of SC and day X prior to SC ($X = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 29$)	°C
ti$X$	Minimum daily temp on day of SC ($X = 0$) or day X prior to SC ($X = 1, 2, 3$)	°C
tid$X$	Mean of minimum daily temp between day of SC and day X prior to SC ($X = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 29$)	°C
tx$X$	Maximum daily temp on day of SC ($X = 0$) or day X prior to SC ($X = 1, 2, 3$)	°C
txd$X$	Mean of the maximum daily temp between the day of SC and day X prior to SC ($X = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 29$)	°C
Precipitation		
pX	Total amt of rain on day of SC ($X = 0$) or day X prior to SC ($X = 1, 2, 3$)	mm
pd$X$	Mean amt of rain between day of SC and day X prior to SC ($X = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 29$)	mm
Wind speed		
ws$X$	Wind speed on day of SC ($X = 0$) or day X prior to SC ($X = 1, 2, 3$)	m/s
wsd$X$	Mean wind speed between day of SC and day X prior to SC ($X = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 29$)	m/s
Landscape factors		
Soil properties		
Soil_acidity	Relative acidity or alkalinity of soil at SL (6.1–7.9/7.9–9.0)	pH
Soil_texture	Soil texture at SL (loam/clay loam/silty clay loam/other)	
Slope	Direction toward which surface of soil faces at SL (0.4–0.5/1–2/4–7)	°
Organic_matter	% of wt of decomposed plant and animal residue at SL (0.5–2.0/2.0–4.0)	%
Distance[c]		
D_road	Distance to nearest road from SL	m
D_water	Distance to nearest body of water from SL	m
Farm management factors[d]		
Workers_time	Time since the last workers' visit during CGS[c]	Days
Hygiene-field status[e]	Composite variable, with 1 indicating the use of portable toilets and washing stations in the field, training of staff/temporary workers to use portable toilets, and absence of grazing and hay production in the field before spinach planting and 0 otherwise (1/0)	
Organic	Organic farming practices currently applied on farm (yes/no)	
Organic_certified	Organic farming certified by National Organic Program (yes/no)	
Before_fallow	Field condition fallow before planting of spinach during CGS (yes/no)	
Manure_application	Manure spread on field for CGS (yes/no)	
Planting_time	Time since planting spinach[c]	Days
Environmental factors[d]		
Proximity_beef	Proximity of beef farm within 10-mile radius (yes/no)	
Proximity_poultry	Proximity of poultry farm within 10-mile radius (yes/no)	
Domestic_animal	Domestic animal intrusion of the field for CGS (yes/no)	
Wild_control_fences	Wildlife control methods of the farm (fences) (yes/no)	
State	Farm location (Texas/Colorado as representative states of Southwestern USA/Western USA)	

[a] For example, tm0 denotes the average daily temperature on the day of sample collection, while tm3 denotes the average daily temperature on day 3 prior to the day of sample collection.
[b] For example, tmd20 denotes the mean of the average daily temperatures recorded for the period between the day of sample collection and day 20 prior to the day of sample collection.
[c] Continuous variable.
[d] Farm management and environmental factors obtained through a survey in our previous study (19).
[e] Due to the simultaneous occurrence of the listed variables, a composite "hygiene-field status" variable was created and used to evaluate the effect of these factors.
[f] SC, sample collection; SL, sampling location; CGS, current growing season.

ability criteria. According to the interpretability criteria, (i) each retained component had to contain at least three variables with major loadings, (ii) the variables loading on a retained component had to share the same conceptual meaning, and (iii) the rotated pattern had to show a simple structure [meaning that (a) most of the variables had relatively high factor loadings on only one component and near-zero loadings on the other components and (b) most components had relatively high factor loadings for some variables and near-zero loadings for the remaining variables] (30). The results of PCA were used in two ways. First, a representative weather variable was chosen for each retained principal component for consideration in the multivariable modeling. The variable choice was based on the P value from the univariate analysis (usually the variable with the lowest P value) and the interpretability and robustness of a conclusion from the multivariable analysis. Second, we predicted the principal component scores for each retained principal component and used the scores as new explan-

TABLE 3 Significant associations between the individual weather variables and spinach contamination with generic *Escherichia coli* based on the univariate mixed-effect logistic regression models with farm and date as random effects

Variable[a]	OR (95% CI)	*P* value[b]
Temp		
tmd5	0.85 (0.67, 1.08)	0.190
tmd7	0.81 (0.61, 1.09)	0.164
tmd8	0.78 (0.58, 1.06)	0.118
tmd9	0.77 (0.56, 1.05)	0.095
tmd10	0.77 (0.56, 1.04)	0.091
tmd15	0.76 (0.57, 1.03)	0.077
tmd20	0.76 (0.56, 1.02)	0.065
tmd25	0.75 (0.56, 1.01)	0.061
tmd29	0.76 (0.57, 1.01)	0.055
tid10	0.82 (0.61, 1.10)	0.181
tid15	0.81 (0.60, 1.08)	0.148
tx2	0.87 (0.70, 1.08)	0.196
tx3	0.86 (0.73, 1.03)	0.104
txd3	0.84 (0.66, 1.08)	0.183
txd4	0.84 (0.67, 1.05)	0.127
txd5	0.84 (0.68, 1.04)	0.116
txd6	0.83 (0.66, 1.05)	0.126
txd7	0.81 (0.63, 1.05)	0.111
txd8	0.80 (0.61, 1.04)	0.089
txd9	0.80 (0.61, 1.03)	0.084
txd10	0.80 (0.62, 1.04)	0.096
txd15	0.80 (0.61, 1.04)	0.089
txd20	0.77 (0.59, 1.01)	0.055
txd25	0.76 (0.58, 0.99)	0.045
txd29	0.75 (0.58, 0.97)	0.028
Precipitation		
pd5	1.18 (0.96, 1.45)	0.113
pd6	1.19 (0.95, 1.48)	0.124
pd7	1.23 (0.95, 1.59)	0.113
pd8	1.28 (0.96, 1.71)	0.096
pd9	1.36 (0.99, 1.85)	0.057
pd10	1.32 (0.94, 1.84)	0.106
pd15	1.37 (0.86, 2.18)	0.179
pd25	1.78 (0.91, 3.47)	0.091
pd29	2.37 (1.16, 4.87)	0.018
PC score[c]		
PC1 (temp)	1.30 (0.99, 1.71)	0.062
PC2 (precipitation)	1.19 (0.95, 1.49)	0.122

[a] Variable definitions: tmd*X*, mean of the average daily temperatures (°C) in the period between the day of sampling collection (SC) and day *X* prior to SC; tid*X*, mean of the minimum daily temperatures (°C) between the day of SC and day *X* prior to SC; tx*X*, maximum daily temperature (°C) on day *X* prior to SC; txd*X*, mean of the maximum daily temperature (°C) between the day of SC and day *X* prior to SC; pd*X*, mean amount of rain (mm) between the day of SC and day *X* prior to SC.
[b] Only variables with *P* values of $<$0.20 are shown.
[c] Principal component (PC) scores were estimated for the variables identified through the principal component analysis in Table 4.

atory variables representing the whole group of weather variables of the corresponding principal component in the univariate and, if applicable, in the multivariable modeling.

The final multivariable mixed-effect logistic regression model was manually selected by conducting a backward-elimination process until only significant variables remained ($P < 0.05$ based on the Wald Z test), where each term deletion was followed by a likelihood ratio (LR) test and comparison based on the Akaike information criterion (AIC). To ensure comparability of nested models required for the LR test, the data set was reduced to observations with complete information (i.e., observations with missing values were excluded) for the part of the analysis that evaluated the variables with missing data. Because farm- and weather-related factors differed by state, suggesting a potential confounding effect by state, the effect of the state factor was examined in all considered multivariable models by comparing the fits of the models with and without the state factor. The presence of confounding was determined based on a >20 to 30% change in ln(OR) between the estimate obtained in the model without state (crude estimate) and the estimate obtained after controlling for the effect of state (adjusted estimate) (31).Two-way interactions of explanatory variables were also considered. The goodness of fit was evaluated based on how much the observed proportions agreed with the mean predicted probabilities using the "plot.logistic.fit.fnc" function in the "language" package (32). The latent variable approach was used to estimate the percentage of variation explained by random effects (farm and date) for the final model (31). The variance inflation factor (VIF) was used in the final model to diagnose colinearity. Locally weighted scatter plot smoothing was used in the final model to assess the linearity assumption between the logit of outcome {log ["probability of contamination"/(1 − "probability of contamination")]} and individual continuous explanatory variables (31).

We evaluated the utility of weather and landscape information from the NRI databases in predicting spinach contamination when these factors were considered alone or jointly with a survey of farm management and environmental factors. We compared the final statistical model developed based on the consideration of weather and landscape data only ("NRI model") with the final statistical model in which NRI data were considered alongside with the survey data ("NRI-survey model"). The predictive performances of the final NRI and NRI-survey models were compared by examining the receiver operating characteristic (ROC) curve and quantifying the area under the curve (AUC). Statistical testing of the difference between the AUCs was conducted using the "roc.test" function in the "pROC" package in R (33). In this assessment of a model's predictive performance, the data used for model development were also used for testing of the model's predictive performance, and it thus served as an internal validation of the developed statistical model. An independent data set for external validation of the developed models was not available. However, to asses the robustness of a model's predictive ability, we conducted a 5-fold cross-validation where the data set was randomly divided into five subsets and then four subsets were used for estimation of the model's coefficients while the fifth subset was used to test the model's predictive ability; this process was repeated five times, every time with a different test subset. The mean (and range) AUC from cross-validation was recorded and used for comparison of models.

RESULTS

A total of 955 spinach samples were collected on 37 days during the period between 7 June 2010 and 10 February 2012 (Table 1). The median temperature on the sampling days in Texas was 13.5°C (range, 8.6°C to 23.4°C), which was significantly lower than the temperature of 18.4°C (range, 9.6°C to 26.2°C) in Colorado (Wilcoxon rank-sum test, $P < 0.001$). The occurrences of rain (including trace rain) on the day of sample collection were similar in Texas and Colorado; on 2 out of 13 sample collection days it rained in Texas, while on 8 out of 24 sample collection days it rained in Colorado (Wilcoxon rank-sum test, $P = 0.432$). The median amounts of rain on the rainy sampling days in Texas (4.3 mm) and Colorado (2.8 mm) were slightly different, but this difference was not statistically significant (Wilcoxon rank-sum test, $P = 0.069$). The soil texture was significantly different between the farms enrolled in Texas and Colorado (χ^2, $P < 0.001$); most of sampled spinach was grown on silty clay loam (63%) in Texas, while it was grown on loam soil (70%) in Colorado. In Texas, a total of 93% sampled spinach was grown on a flat terrain (0.4 to

TABLE 4 Principal component analysis of weather variables in Table 3

Variable[a]	PC1[b]	PC2[b]
tmd5	**−0.20**	0.04
tmd7	**−0.22**	0.06
tmd8	**−0.22**	0.04
tmd9	**−0.22**	0.03
tmd10	**−0.22**	0.02
tmd15	**−0.22**	0.06
tmd20	**−0.21**	0.07
tmd25	**−0.20**	0.09
tmd29	**−0.20**	0.08
tx3	**−0.20**	0.03
txd3	**−0.20**	−0.01
txd4	**−0.21**	−0.02
txd5	**−0.21**	−0.02
txd6	**−0.22**	−0.02
txd7	**−0.22**	−0.01
txd8	**−0.22**	−0.02
txd9	**−0.22**	−0.04
txd10	**−0.21**	−0.04
txd15	**−0.22**	0.01
txd20	**−0.22**	0.03
txd25	**−0.21**	0.05
txd29	**−0.21**	0.04
pd5	0.03	**0.34**
pd6	0.03	**0.34**
pd7	0.02	**0.34**
pd8	0.02	**0.35**
pd9	0.02	**0.35**
pd10	0.02	**0.35**
pd15	0.05	**0.33**
pd25	0.05	**0.27**
pd29	0.06	**0.25**
SD (% of variance; cumulative %)	4.45 (0.64; 0.64)	2.73 (0.24; 0.88)

[a] tmd*X*, mean of the average daily temperatures (°C) in the period between the day of sampling collection (SC) and day *X* prior to SC; tx*X*, maximum daily temperature (°C) on day *X* prior to SC; txd*X*, mean of the maximum daily temperature (°C) between the day of SC and day *X* prior to SC; pd*X*, mean amount of rain (mm) between the day of SC and day *X* prior to SC.
[b] Bold indicates the highest component loadings.

TABLE 5 Final NRI multivariable mixed-effect logistic regression model with farm and date as random effects[a]

Variable (comparison level or unit)	Reference	OR (95% CI)	*P* value
pd29 (mm)[b]		2.9 (1.3, 6.3)	0.008
State (Texas)	Colorado	16.9 (1.4, 206.2)	0.027

[a] Variance component values (standard deviations) were 1.145 (1.070) for farm and 4.298 (2.073) for date. For the intercept-only model, variance component values were 0.938 (0.968) for farm and 6.106 (2.471) for date.
[b] pd29, mean amount of rain between the day of sampling collection (SC) and day 29 prior to SC.

0.5°), while in Colorado, sampled spinach was grown on a nearly level (1 to 2°, 73%) or gently sloping (4 to 7°, 27%) terrain. The enrolled farms in Texas had a significantly larger median distance to the nearest water body than the farms in Colorado. The median distance to the nearest water body was 1,607 m (range, 341 m to 8,156 m) in Texas and 352 m (range, 7 m to 1,153 m) in Colorado (Wilcoxon rank-sum test, $P < 0.001$). The median distance to the nearest road was 267 m (range, 3 m to 864 m) in Texas and 393 m (range, 21 m to 3,823 m) in Colorado (Wilcoxon rank-sum test, $P < 0.001$).

Generic *E. coli* was detected on 6.6% (63/955) of spinach samples. In the univariate mixed-effect models (Table 3), the odds of spinach contamination with generic *E. coli* were reduced when spinach was exposed to a higher average, minimum, or maximum temperature for several days before sample collection. On the other hand, the odds of spinach contamination were increased when spinach was exposed to a larger amount of rain for several days before sample collection. These results indicate that there is a protective effect of elevated temperatures and reduced moisture on the odds of detecting the indicator organism on spinach.

In PCA, two components explained 88% of the total variability (Table 4). Twenty-two variables that describe the average, minimum, and maximum daily or period temperatures were loaded on the first component (labeled the temperature component). Nine variables describing the precipitation were loaded on the second component (labeled the precipitation component). The principal component scores for the temperature and precipitation components were both significant as predictors of spinach contamination in univariate analysis (Table 3).

The final multivariable NRI model had a single NRI predictor, the mean amount of rain between the day of sample collection and 29 days prior to the sample collection date (pd29). However, adding the state factor into the NRI model with pd29 only (AIC = 366) improved the model fit (LR = 5.6; degrees of freedom [df] = 5 − 4 = 1; χ^2, $P = 0.018$; AIC = 362), and the model with pd29 and state was considered the final NRI model (Table 5). According to the model, after controlling for the effect of state, the OR describing the effect of rain on the probability of spinach contamination increased by 21% (adjusted OR = 2.9; crude OR = 2.4). No interaction was detected between the rain and state factors. The NRI model had a low predictive ability in internal validation (AUC = 69%). The cross-validation results indicated weak repeatability, with mean AUC = 68% (range, 64% to 74%) (Fig. 1). When we attempted to use the principal component scores for temperature and precipitation instead of the actual temperature and precipitation variables, the final model could not be selected at the 5% significance level (even after controlling for the effect of state).

Univariate mixed-effect logistic regression analysis (with farm and date as random effects) of survey variables indicated associations between the probability of spinach contamination and several farm management and environmental factors (Table 6). The variables listed in Tables 3 and 6 were considered in a multivariable mixed-effect logistic regression model to assess if weather data could complement the farm management and environmental factors in explaining the probability of spinach contamination. The variables in Table 6 that were highly correlated (e.g., the proximity of a beef farm and poultry farm and domestic animal intrusion) were considered one at a time in the multivariable modeling. The final NRI-survey model is shown in Table 7. Based on this model, the odds of spinach contamination with generic *E. coli* were reduced to approximately 1 in 17 (OR = 0.06) in the presence of "hygiene-field status" factors on the sampled field. However, the odds of contamination were increased (OR = 3.5) for every mm increase in the mean amount of rain between the day of sample collection and day 29 prior. The odds of contamination

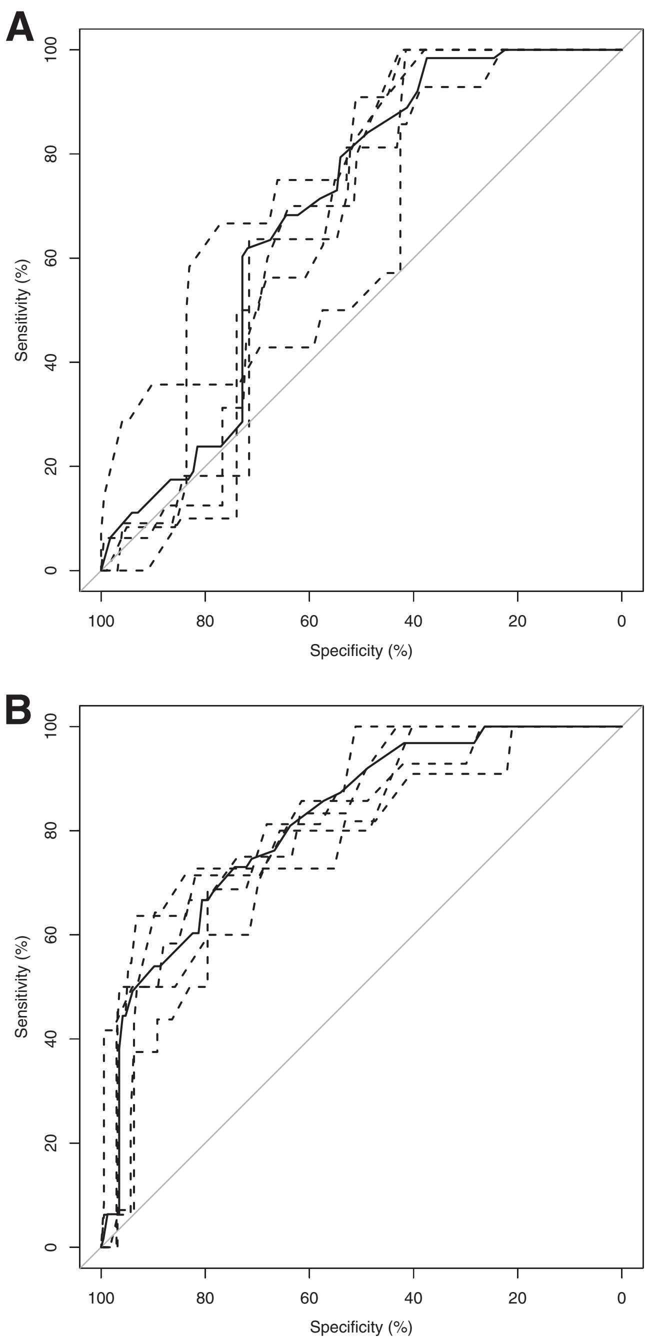

FIG 1 Receiver operating characteristic (ROC) curves from each of the five runs of the 5-fold cross-validation (dashed lines) and from the internal validation (solid line). (A) NRI model, including state and pd29; (B) NRI-survey model, including state, pd29, "hygiene-field status," and the use of manure fertilizer.

were elevated to approximately 52 in 1 (OR = 52.2) if manure fertilizer was applied onto the field before the current growing season. Likewise, the probability of spinach contamination was higher in Texas than in Colorado (OR = 108.1). It is interesting to note that the final NRI-survey model in Table 7 with state included (AIC = 346) had a statistically better fit than the model without state (LR = 14.8; df = 7 − 6 = 1; χ^2, $P < 0.001$; AIC = 359). None of the possible 2-way interactions in the NRI-survey model were significant. However, there was an indication of a possible confounding effect of state on (i) the association between rain and spinach contamination (crude OR = 2.4; adjusted OR = 2.9) and (ii) the association between the use of manure fertilizer and spinach contamination (crude OR = 10.4; adjusted OR = 68.9). The association between "hygiene-field status" factors and spinach contamination also seemed to have been confounded by the effect of rain (crude OR = 0.14; adjusted OR = 0.05). The causal diagram in Fig. 2 depicts the identified determinants of spinach contamination in the final NRI-survey model and their relationships. Internal validation confirmed that the predictive ability of the final NRI-survey model (AUC = 82%) was significantly better than that of the final NRI model ($P < 0.001$). Cross-validation analyses showed that the mean AUC was 81% (range, 80% to 84%) (Fig. 1). In the multilevel intercept-only model, the proportions of variation accounted for in the NRI-survey model were 1.4% and 51.1% at the farm and date levels, respectively, and in the multilevel final model, they were 9.1% and 59.1%, respectively.

In multivariable modeling of temperature and precipitation component scores jointly with the farm management and environmental factors, the final model included the same variables as the final NRI-survey model counterpart described above. The two models also had equal predictive performance (results not shown). However, the model coefficients were different, indicating slightly different magnitudes of association (e.g., for the precipitation component scores, estimated OR = 1.26 and 95% confidence interval [CI] = 1.01 to 1.56). This particular model is not being shown here because the models had equal structure and predictive performance and because the principal component scores were difficult to interpret in the model with precipitation component scores.

DISCUSSION

This study evaluated the effectiveness of using easily accessible weather and landscape data from the NRI databases to explain the probability of spinach contamination with generic *E. coli* at the preharvest level when considered alone or in combination with a produce farmer survey about farm management and environmental factors. The study results demonstrated that NRI databases could be used relatively easily to obtain information on precipitation as a determinant of spinach contamination with generic *E. coli*. The predictive ability of the developed statistical model was significantly improved when precipitation was considered together with the farm management and environmental factors provided by surveyed farmers. This finding supports the idea that farm management, environment, and weather factors should be considered together to predict spinach contamination and to design novel control strategies, including good agricultural practices (GAPs) to harvest produce when the risk for contamination is predicted to be at its lowest.

The study found that the odds of spinach contamination with generic *E. coli* increased with the mean amount of rain between the day of sample collection and day 29 prior to sample collection. Previous studies have also shown that storms and rain can increase produce contamination (34–36). These studies identified that rain splashed *Salmonella* Typhimurium onto tomato plants (34), splashed *Colletotrichum acutatum* onto strawberry fruits (36), and led to the transport of *Salmonella* Typhimurium to tomato fruits by aerosols (35). A modeling study showed that the probability of

TABLE 6 Farm management and environmental factors identified through analysis of the variables from a survey of spinach farmers that were significantly associated with spinach contamination with generic *Escherichia coli* based on the univariate mixed-effect logistic regression models with farm and date as random effects

Variable[a]	Level	Frequency[b]	OR (95% CI)	*P* value[c]
Farm management factors				
Workers_time	>3 days	13/330	0.31 (0.09, 1.04)	0.057
	≤3 days	32/360	Reference	
Hygiene-field status[d]	1	56/930	0.14 (0.02, 0.86)	0.034
	0	7/25	Reference	
Organic	Yes	31/260	3.7 (0.8, 16.4)	0.089
	No	32/695	Reference	
Organic_certified	Yes	4/175	0.01 (0.00, 0.98)	0.049
	No	7/25	Reference	
Before_fallow	Yes	15/165	5.8 (1.2, 27.6)	0.027
	No	48/790	Reference	
Manure_application	Yes	25/160	10.4 (1.4, 78.4)	0.024
	No	38/795	Reference	
Planting_time	>66 days	46/465	2.2 (0.8, 6.2)	0.144
	≤66 days	17/490	Reference	
Farm environmental factors				
Proximity_beef	Yes	24/120	9.0 (0.6, 145.1)	0.120
	No	39/835	Reference	
Proximity_poultry	Yes	24/110	11.4 (0.8, 168.6)	0.077
	No	39/845	Reference	
Domestic_animal	Yes	7/25	7.09 (1.20, 42.02)	0.031
	No	56/910	Reference	
Wild_control_fences	Yes	19/325	0.16 (0.01, 2.16)	0.168
	No	28/180	Reference	
State	Texas	36/480	12.6 (0.9, 180.1)	0.061
	Colorado	27/475	Reference	

[a] Variables are defined in Table 2.
[b] Number of observations with generic *E. coli*-contaminated spinach/total number of recorded observations for the variable.
[c] Only variables with *P* values of <0.20 are shown.
[d] The estimated OR (95% CI) value applies to all factors within the composite variable "hygiene-field status"; level 1 indicates the presence of toilet training and use of toilets and washing stations but absence of field grazing and hay production before planting of the spinach during the current growing season.

lettuce contamination with *E. coli* O157:H7 from manure-amended soil was significantly correlated with the number of times it rained (37). If microorganisms persist on produce prior to rain, high humidity may contribute to the survival or growth of the microorganisms on produce (38–40). Strawn et al. (23) identified precipitation as an important factor influencing the isolation of *Salmonella* on produce farms. The results presented here demonstrate that for every mm of rainfall during the 29-day period before sample collection, the odds of spinach contamination increase by 3.5 (95% CI, 1.7 to 7.3). This estimate was obtained after controlling for the effect of manure application, state, and "hygiene-field status" factors. This may prompt farmers to use NRI weather databases to monitor the amount of rain on their

TABLE 7 Final NRI-survey multivariable mixed-effect logistic regression model with farm and date as random effects[a]

Variable (comparison level or unit)	Reference	OR (95% CI)	*P* value
Hygiene-field status (1)[b]	0	0.06 (0.01, 0.30)	0.001
pd29 (mm)[c]		3.5 (1.7, 7.3)	0.001
Manure_application (yes)	No	52.2 (2.8, 968.0)	0.008
State (Texas)	Colorado	108.1 (4.8, 2447.3)	0.003

[a] Variance component values (standard deviations) were 0.100 (0.316) for farm and 3.534 (1.880) for date. For the intercept-only model, variance component values were 0.938 (0.968) for farm and 6.106 (2.471) for date.
[b] The estimated OR (95% CI) value applies to all factors within the composite variable "hygiene-field status" group; level 1 indicates the presence of toilet training and use of toilets and washing stations but absence of field grazing and hay production before planting of the spinach during the current growing season.
[c] pd29, mean amount of rain between the day of sampling collection (SC) and day 29 prior to SC.

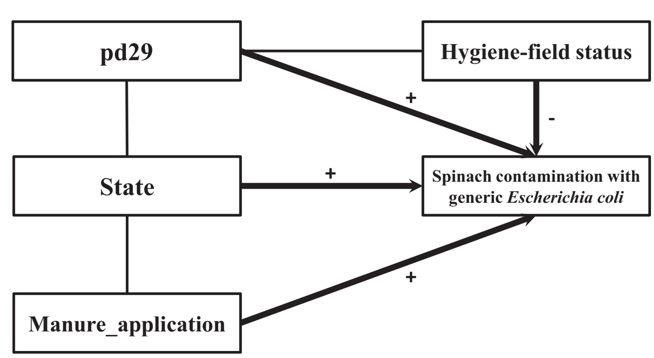

FIG 2 Proposed causal diagram of how farm management, environmental, and weather factors jointly influence spinach contamination with generic *Escherichia coli*. Thick arrows indicate exposure effects; the thin line indicates an association between explanatory variables and possible confounding effects; plus and minus signs indicate positive and negative associations, respectively, between the exposure variables and spinach contamination.

produce fields during a growing season, because they can use those data in conjunction with information on their particular farm management practices to predict the microbial safety of harvested spinach. This study also identified a possible distorting confounding effect of state (31), which made the association between the amount of rain during the period before sample collection and spinach contamination seem weaker than it actually was. It is unknown whether the state effect is a true population or sample confounder (31). The fact that controlling for the effect of state significantly improved the model fit and dramatically enhanced the effect of identified risk factors suggests that the effect of state should be considered in multistate (and multicountry) field studies of produce contamination. The likely explanation for the confounding effect of state may be in the different distributions of weather, landscape, and management practices across the states. For that same reason, generalizing our results to produce farms in other states should be done with caution.

Univariate analysis demonstrated that the odds of spinach contamination with generic *E. coli* decreased significantly (at a 5% significance level) with an increasing mean of the maximum daily temperatures between the day of sample collection and day 25 or 29 prior to sample collection. These results were not expected, as the growth rate of *E. coli* is optimal at warm temperatures between 27 and 39°C (41, 42). After controlling for the effect of state in multivariable modeling, these temperature variables were no longer significantly associated with spinach contamination, supporting that state acted as a distorter variable (31). While temperature may be a true predictor of spinach contamination with generic *E. coli*, based on the NRI temperature data we were unable to confirm its effect. Interestingly, the most significant precipitation variable in the univariate analysis (Table 3) was the mean precipitation during the 29-day period before sample collection. That result might be attributed to the long-term, or even seasonal, effect of weather on the survival and growth of generic *E. coli* on spinach at the preharvest level. This also indicates that the weather that occurs more than 29 days prior to sample collection may be an even more informative predictor of spinach contamination. Future study designs should account for long-term and seasonal effects of weather on the probability of produce contamination.

The odds of spinach contamination with generic *E. coli* were significantly elevated if the farm applied manure fertilizer (OR = 52.2). A previous study (5) also demonstrated that the use of manure fertilizer increased the produce contamination with *E. coli* in organic (OR = 13.2) and semiorganic (OR = 12.9) farms. Strawn et al. showed that manure application within a year prior to sample collection increased the likelihood of *Salmonella* being detected in a produce field (8). The results of our previous study that used the same survey data (but not NRI data) did not include the use of manure fertilizer (19). A possible reason for that may be a confounding effect of state on the association between the manure use and spinach contamination identified in the current study and the forward selection of the final model used in our previous work (19). Indeed, after controlling for the effect of state, the association between the use of manure fertilizer and spinach contamination became much stronger.

The "hygiene-field status" group included the use of portable toilets and washing stations in the field, training staff/temporary workers to use portable toilets, and absence of grazing and hay production in the field before spinach planting. This group had a protective effect (OR = 0.06) on spinach contamination with generic *E. coli* when considered along with weather and landscape factors as well as with the other farm management and environmental factors. This is in agreement with the results from our previous study, where "hygiene-field status" was the strongest protective factor (OR = 0.15) among the 76 surveyed farm management and environmental factors (19). The repeatability of the finding confirms the potential importance of this group of factors in produce food safety, highlighting the need to further elucidate their role in produce contamination. The association between the spinach contamination and "hygiene-field status" group of factors seemed to have been confounded by the amount of rainfall during the period prior to sample collection by making the association seem weaker than in actually was. A closer examination of the data suggests that confounding may be explained by an uneven distribution of the amount of rainfall with respect to the presence of "hygiene-field status" group factors on the sampled fields. It is likely that the identified confounding was of statistical origin rather than there being a biologically meaningful effect of rain on the association between the "hygiene-field status" group of factors and spinach contamination.

In addition to the "hygiene-field status" group of factors, our previous study (19) identified an association between spinach contamination with generic *E. coli* and the following risk factors: an irrigation lapse time of >5 days, a >66-day period since the planting of spinach, the farm location in Texas versus Colorado, the use of pond water for irrigation, and the proximity (within 10 miles) of a poultry farm. In the current study, these factors were not identified as significant in the final NRI-survey model. There are two possible explanations for these discrepancies. First, the previous study considered "farm" and "farm visit" random variables, whereas in the current study "farm" and "date" were considered random factors, because "date" explained the highest proportion of variation for weather factors. For example, in the univariate analyses for the average temperature of the sample collection day, "farm" and "date" explained 8.4% and 60.6% of variation, respectively, whereas "farm" and "farm visit" explained 32.9% and 12.1%, respectively. The difference in the considered random effects explains the differences between the results shown Table 6 and those in Table 4 in our previous paper (19). The second reason for the discrepancies between the two studies may be the presence of additional explanatory variables (e.g., weather factors) in the current study. Nevertheless, the fact that two models commonly retained the "hygiene-field status" group of factors is important, because farmers can improve the produce safety by supporting workers' hygiene and by managing the field condition before planting spinach.

Previous studies showed the difference of persistence and growth of microorganisms according to soil acidity (43) and the degree of organic matter in soil (44). However, the quality of data obtained from NRI databases for those variables was not satisfactory. For example, the soil acidity data had several overlapping levels (e.g., pH 6.1 to 7.9, 6.6 to 8.4, and 7.9 to 9.0). We encountered a similar problem with the degree of organic matter. Thus, the statistical analysis of these factors was not meaningful, supporting that future evaluation of these soil property data should consider sources other than the NRI databases considered here. Previous studies conducted under controlled conditions showed that the survivability of microorganisms in soil or on produce was significantly affected by slope (45) and soil texture (46). However, our findings did not support the results of those controlled trials.

The inconsistencies between the controlled trials and our observational study may be attributed to farm management or weather factors that may have obscured the relationship between these soil factors and spinach contamination. We additionally tested soil salinity and soil type (e.g., classified as entisols, inceptisols, and mollisols) to determine if they influenced the probability of spinach contamination with *E. coli*, but the results were not significant.

Spinach samples collected in close proximity (i.e., on the same farm or even in the same county) were likely to be more similar to each other than samples collected in distant locations. Failing to account for such an autocorrelation could introduce a certain level of bias into the results of the statistical analysis by deflating standard errors (47). In the current study, the spatial autocorrelation was partially accounted for by considering farm as a random effect (48).

Based on the models' estimates of AUC, the ability of the final NRI-survey model to correctly identify spinach contamination with generic *E. coli* was better than that of the final NRI model. However, the final NRI-survey model was not significantly better than the model based on survey data only from our previous study (19) (results not shown). This suggests that the predictor variables from a survey are equally good indicators of spinach contamination whether used alone or together with weather variables obtained from NRI databases. This makes sense because farm management practices will tend to adapt to the local weather events. While the final NRI-survey model showed a relatively high and repeatable predictive performance in cross-validation, caution is needed in generalizing model results to other locations before the model is evaluated on an independent data set.

Our results demonstrated that produce contamination is affected jointly by a net of risk and protective factors related to farm management, environment, and weather (Fig. 2). These factors must be considered together both when interpreting preharvest produce contamination data and when designing and implementing strategies to control the microbial safety of fresh produce. Information on the identified risk factors should ideally be translated into new and improved control strategies to enhance the safety of fresh produce at the preharvest level. Specifically, avoiding application of animal manure would significantly reduce produce contamination. An example of this can be found in the federal laws overseeing fruit and vegetable food safety issued as part of the Food Safety Modernization Act (FSMA) Proposed Rule for Produce Safety (49), the Produce GAPs Harmonized Food Safety Standard (http://www.ams.usda.gov/AMSv1.0/getfile?dDocName=STELPRDC5105384), and a commodity-specific GAP "California Leafy Green Marketing Agreement" (LGMA) (50), which already include specific directions for farmers to avoid applying raw manure or soil amendments that contain improperly composted animal manure to produce fields. These documents also outline acceptable treatments, application intervals, testing, certification, and record keeping in case manure and compost are used. Our results indicate that portable toilets and hand-washing stations should be provided in the fields for workers, and the farmers should train their field workers on how to use them in order to reduce the produce contamination with microorganisms. Existing standards and regulations already require the adequate personnel training of field workers on sanitation and hygiene practices and the presence of toilet and hand-washing facilities around fields for produce safety (49, 50). Interventions to address risk factors related to farm management, such as manure application and the use of portable toilets and hand-washing stations, not only are part of the existing produce safety standards and regulations but also could be self-regulated by growers. The real challenge is how to translate information about the role of rainfall in produce safety into an intervention that growers could implement on a daily basis. For example, growers could aim to adjust farm management practices (e.g., harvest time) according to the rainfall conditions with the goal of producing microbiologically safer fruits and vegetables. However, this may be difficult to achieve in practice because rainfall is largely unpredictable and because its consequences are difficult to prevent/alleviate (e.g., there is not much flexibility in adjusting harvest time). Therefore, developing strategies to control the effects of rainfall on the microbial safety of fresh produce at preharvest requires further combined efforts of the grower industry, government, and academia and perhaps even establishment of an insurance program for weather-related microbial safety of produce. An example of such a real-world application would be to take produce harvested during high-risk weather conditions, or at times that are high risk for other reasons, and divert it for processing (cooking or freezing) as opposed to fresh produce. This would improve the safety of the produce but inadvertently have an effect on the profit. Also, changing from fresh market product to processing may be difficult, because processing products have contracts already spelled out before the crop is planted and different produce varieties are often grown for fresh and processing purposes. As an additional strategy, routine sampling of product at the field, which is currently used by some growers, could be further developed and standardized to test for indicator organisms in order to manage the risk of potential pathogen contamination.

To our knowledge, this is the first study on the association between produce contamination and the combination of farm management, environmental, and weather factors. Previous field studies have evaluated the effects of only a subset of these factors on produce contamination (5, 6, 14, 19). Although a recent study (23) identified risk factors (e.g., temperature, precipitation, and available water storage in soil) among landscape and meteorological factors for the isolation of *Salmonella* and *L. monocytogenes* on fruit and vegetable farms, they did not use crop samples but collected only soil, water, feces, and drag swab samples. The current study has some limitations worth noting. First, the causes of spinach contamination with generic *E. coli* could not be determined explicitly due to the cross-sectional nature of the study. Second, measurement errors in the weather data may be substantial considering that some of the nearest weather stations were up to 35 km away from the enrolled farms. Third, caution is needed in generalizing the results to all spinach farms in the United States because our study was based on only 12 produce farms growing spinach located in Colorado and Texas as representative of the Western and Southwestern United States, respectively. Fourth, the source of generic *E. coli* contamination on spinach was not identified in this study.

While the study focused on spinach, the findings are likely to be generalizable to other leafy greens due to the common properties of leafy greens, such as similar cultivation and harvest methods and less direct contact of the edible portion with soil. Considering that many of the enrolled farms produced other produce commodities in addition to spinach, the findings of the effect of farm management, environmental, and weather factors on spinach contamination may be applicable to produce in general. Finally, the developed geospatial and statistical modeling framework is adaptable to study determinants of produce contamination with other food-borne pathogens and on other produce commodities at the preharvest level.

In conclusion, our findings highlighted the utility of NRI precipitation data for predicting spinach contamination at the preharvest level. We also demonstrated how farm management, environmental, and weather factors jointly affect produce contamination with generic *E. coli*. These factors should be considered jointly in development of GAPs and measures to reduce produce contamination. Spinach contamination was significantly associated with the "hygiene-field status" group of factors, rainfall, and the use of manure fertilizer, with the effect of the latter two factors likely being confounded by state. Future studies of microbial contamination of leafy greens should focus on these factors.

ACKNOWLEDGMENTS

This work was supported by Agriculture and Food Research Initiative grant 2009-04261, program code 93233, to R.I. from the U.S. Department of Agriculture National Institute of Food and Agriculture (USDA-NIFA).

Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the USDA-NIFA.

We thank the two anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

REFERENCES

1. **Gould LH, Walsh KA, Vieira AR, Herman K, Williams IT, Hall AJ, Cole D.** 2013. Surveillance for foodborne disease outbreaks—United States, 1998-2008. MMWR Surveill. Summ. **62:**1–34.
2. **U.S. Food and Drug Administration.** 2008. Guide to minimize microbial food safety hazards for fresh fruits and vegetables; request for comments and for scientific data and information [docket no. FDA-2008-N-0455]. Fed. Regist. **73:**51306–51309. http://www.gpo.gov/fdsys/pkg/FR-2008-09-02/pdf/E8-20187.pdf. Accessed 29 January 2014.
3. **Lindqvist R, Andersson Y, de Jong B, Norberg P.** 2000. A summary of reported foodborne disease incidents in Sweden, 1992 to 1997. J. Food Prot. **63:**1315–1320.
4. **Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV.** 1999. Food-related illness and death in the United States. Emerg. Infect. Dis. **5:**607–625. http://dx.doi.org/10.3201/eid0505.990502.
5. **Mukherjee A, Speh D, Diez-Gonzalez F.** 2007. Association of farm management practices with risk of *Escherichia coli* contamination in preharvest produce grown in Minnesota and Wisconsin. Int. J. Food Microbiol. **120:**296–302. http://dx.doi.org/10.1016/j.ijfoodmicro.2007.09.007.
6. **Mukherjee A, Speh D, Dyck E, Diez-Gonzalez F.** 2004. Preharvest evaluation of *coliforms*, *Escherichia coli*, *Salmonella*, and *Escherichia coli* O157:H7 in organic and conventional produce grown by Minnesota farmers. J. Food Prot. **67:**894–900.
7. **Melloul AA, Hassani L, Rafouk L.** 2001. *Salmonella* contamination of vegetables irrigated with untreated wastewater. World J. Microbiol. Biotechnol. **17:**207–209. http://dx.doi.org/10.1023/A:1016686501953.
8. **Strawn LK, Grohn YT, Warchocki S, Worobo RW, Bihn EA, Wiedmann M.** 2013. Risk factors associated with *Salmonella* and *Listeria monocytogenes* contamination of produce fields. Appl. Environ. Microbiol. **79:**7618–7627. http://dx.doi.org/10.1128/AEM.02831-13.
9. **Orozco RL, Iturriaga MH, Tamplin ML, Fratamico PM, Call JE, Luchansky JB, Escartin EF.** 2008. Animal and environmental impact on the presence and distribution of *Salmonella* and *Escherichia coli* in hydroponic tomato greenhouses. J. Food Prot. **71:**676–683.
10. **Mitscherlich E, Marth EH.** 1984. Microbial survival in the environment. Bacteria and rickettsiae important in human and animal health. Springer-Verlag, New York, NY.
11. **Rogers S, Haines J.** 2005. Detecting and mitigating the environmental impact of fecal pathogens originating from confined animal feeding operations: review. http://storage.globalcitizen.net/data/topic/knowledge/uploads/2012012015737302.pdf. Accessed 29 January 2014.
12. **Hilborn ED, Mermin JH, Mshar PA, Hadler JL, Voetsch A, Wojtkunski C, Swartz M, Mshar R, Lambert-Fair MA, Farrar JA, Glynn MK, Slutsker L.** 1999. A multistate outbreak of *Escherichia coli* O157:H7 infections associated with consumption of mesclun lettuce. Arch. Intern. Med. **159:**1758–1764. http://dx.doi.org/10.1001/archinte.159.15.1758.
13. **Johannessen GS, Froseth RB, Solemdal L, Jarp J, Wasteson Y, L MR.** 2004. Influence of bovine manure as fertilizer on the bacteriological quality of organic Iceberg lettuce. J. Appl. Microbiol. **96:**787–794. http://dx.doi.org/10.1111/j.1365-2672.2004.02208.x.
14. **Mukherjee A, Speh D, Jones AT, Buesing KM, Diez-Gonzalez F.** 2006. Longitudinal microbiological survey of fresh produce grown by farmers in the upper midwest. J. Food Prot. **69:**1928–1936.
15. **Natvig EE, Ingham SC, Ingham BH, Cooperband LR, Roper TR.** 2002. *Salmonella enterica* serovar Typhimurium and *Escherichia coli* contamination of root and leaf vegetables grown in soils with incorporated bovine manure. Appl. Environ. Microbiol. **68:**2737–2744. http://dx.doi.org/10.1128/AEM.68.6.2737-2744.2002.
16. **Byappanahalli MN, Whitman RL, Shively DA, Sadowsky MJ, Ishii S.** 2006. Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. Environ. Microbiol. **8:**504–513. http://dx.doi.org/10.1111/j.1462-2920.2005.00916.x.
17. **Ishii S, Ksoll WB, Hicks RE, Sadowsky MJ.** 2006. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. Appl. Environ. Microbiol. **72:**612–621. http://dx.doi.org/10.1128/AEM.72.1.612-621.2006.
18. **Jimenez L, Muniz I, Toranzos GA, Hazen TC.** 1989. Survival and activity of Salmonella typhimurium and *Escherichia coli* in tropical freshwater. J. Appl. Bacteriol. **67:**61–69. http://dx.doi.org/10.1111/j.1365-2672.1989.tb04955.x.
19. **Park S, Navratil S, Gregory A, Bauer A, Srinath I, Jun M, Szonyi B, Nightingale K, Anciso J, Ivanek R.** 2013. Generic *Escherichia coli* contamination of spinach at the preharvest level: The role of farm management and environmental factors. Appl. Environ. Microbiol. **79:**4347–4358. http://dx.doi.org/10.1128/AEM.00474-13.
20. **Ogden LD, Fenlon DR, Vinten AJ, Lewis D.** 2001. The fate of *Escherichia coli* O157 in soil and its potential to contaminate drinking water. Int. J. Food Microbiol. **66:**111–117. http://dx.doi.org/10.1016/S0168-1605(00)00508-0.
21. **Eblen DR, Annous BA, Sapers GM.** 2005. Studies to select appropriate nonpathogenic surrogate *Escherichia coli* strains for potential use in place of *Escherichia coli* O157:H7 and salmonella in pilot plant studies. J. Food Prot. **68:**282–291.
22. **Clarke KC, McLafferty SL, Tempalski BJ.** 1996. On epidemiology and geographic information systems: a review and discussion of future directions. Emerg. Infect. Dis. **2:**85–92. http://dx.doi.org/10.3201/eid0202.960202.
23. **Strawn LK, Fortes ED, Bihn EA, Nightingale KK, Grohn YT, Worobo RW, Wiedmann M, Bergholz PW.** 2013. Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. Appl. Environ. Microbiol. **79:**588–600. http://dx.doi.org/10.1128/AEM.02491-12.
24. **Ivanek R, Grohn YT, Wells MT, Lembo AJ, Jr, Sauders BD, Wiedmann M.** 2009. Modeling of spatially referenced environmental and meteorological factors influencing the probability of *Listeria* species isolation from natural environments. Appl. Environ. Microbiol. **75:**5893–5909. http://dx.doi.org/10.1128/AEM.02757-08.
25. **Brown PE, Christensen OF, Clough HE, Diggle PJ, Hart CA, Hazel S, Kemp R, Leatherbarrow AJ, Moore A, Sutherst J, Turner J, Williams NJ, Wright EJ, French NP.** 2004. Frequency and spatial distribution of environmental *Campylobacter* spp. Appl. Environ. Microbiol. **70:**6501–6511. http://dx.doi.org/10.1128/AEM.70.11.6501-6511.2004.
26. **Martinez-Urtaza J, Lozano-Leon A, Varela-Pet J, Trinanes J, Pazos Y, Garcia-Martin O.** 2008. Environmental determinants of the occurrence and distribution of *Vibrio parahaemolyticus* in the rias of Galicia, Spain. Appl. Environ. Microbiol. **74:**265–274. http://dx.doi.org/10.1128/AEM.01307-07.
27. **Masabni J.** 2011. Spinach. http://aggie-horticulture.tamu.edu/vegetable/files/2011/10/spinach.pdf. Accessed 29 January 2014.
28. **Bates D, Maechler M, Bolker B.** 2012. lme4: linear mixed-effects models using S4 classes. http://cran.r-project.org/web/packages/lme4/lme4.pdf. Accessed 29 January 2014.
29. **Wearden S.** 2010. Phi coefficient, p 1234–1235. *In* Weiner IB, Craighead

WE (ed), The Corsini encyclopedia of psychology, 4th ed, vol 3. John Wiley & Sons, Inc., New York, NY.

30. **Hatcher L.** 1994. A step-by-step approach to using the SAS system for factor analysis and structural equation modeling. SAS Institute, Inc., Cary, NC.
31. **Dohoo IR, Martin W, Stryhn H.** 2010. Veterinary epidemiologic research, 2nd ed. VER Incorporated, Prince Edward Island, Canada.
32. **Baayen RH.** 2012. languageR: data sets and functions with "Analyzing linguistic data: a practical introduction to statistics." http://cran.r-project.org/web/packages/languageR/languageR.pdf. Accessed 29 January 2014.
33. **Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, Müller M.** 2013. Package "pROC." http://cran.r-project.org/web/packages/pROC/pROC.pdf. Accessed 29 January 2014.
34. **Cevallos-Cevallos JM, Danyluk MD, Gu G, Vallad GE, van Bruggen AH.** 2012. Dispersal of *Salmonella typhimurium* by rain splash onto tomato plants. J. Food Prot. **75**:472–479. http://dx.doi.org/10.4315/0362-028X.JFP-11-399.
35. **Cevallos-Cevallos JM, Gu G, Danyluk MD, Dufault NS, van Bruggen AH.** 2012. *Salmonella* can reach tomato fruits on plants exposed to aerosols formed by rain. Int. J. Food Microbiol. **158**:140–146. http://dx.doi.org/10.1016/j.ijfoodmicro.2012.07.009.
36. **Madden L, Yang X, Wilson L.** 1996. Effects of rain intensity on splash dispersal of *Colletotrichum acutatum*. Phytopathology **86**:864–874. http://dx.doi.org/10.1094/Phyto-86-864.
37. **Franz E, Semenov AV, van Bruggen AH.** 2008. Modelling the contamination of lettuce with *Escherichia coli* O157:H7 from manure-amended soil and the effect of intervention strategies. J. Appl. Microbiol. **105**:1569–1584. http://dx.doi.org/10.1111/j.1365-2672.2008.03915.x.
38. **Park S, Szonyi B, Gautam R, Nightingale K, Anciso J, Ivanek R.** 2012. Risk factors for microbial contamination in fruits and vegetables at the preharvest level: a systematic review. J. Food Prot. **75**:2055–2081. http://dx.doi.org/10.4315/0362-028X.JFP-12-160.
39. **Beuchat LR.** 2006. Vectors and conditions for preharvest contamination of fruits and vegetables with pathogens capable of causing enteric diseases. Br. Food J. **108**:38–53. http://dx.doi.org/10.1108/00070700610637625.
40. **Doyle MP, Erickson MC.** 2008. Summer meeting 2007—the problems with fresh produce: an overview. J. Appl. Microbiol. **105**:317–330. http://dx.doi.org/10.1111/j.1365-2672.2008.03746.x.
41. **Cooper VS, Bennett AF, Lenski RE.** 2001. Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. Evolution **55**:889–896. http://dx.doi.org/10.1554/0014-3820(2001)055[0889:EOTDOG]2.0.CO;2.
42. **Ingraham J, Marr A.** 1996. Effect of temperature, pressure, pH, and osmotic stress on growth, p 1570–1578. *In* Neidhardt FC, Curtiss R, III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (ed), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, DC.
43. **Jiang X, Morgan J, Doyle MP.** 2002. Fate of *Escherichia coli* O157:H7 in manure-amended soil. Appl. Environ. Microbiol. **68**:2605–2609. http://dx.doi.org/10.1128/AEM.68.5.2605-2609.2002.
44. **Jamieson R, Gordon R, Sharples K, Stratton G, Madani A.** 2002. Movement and persistence of fecal bacteria in agricultural soils and subsurface drainage water: a review. Can. Biosyst. Eng. **44**:1–9.
45. **Abu-Ashour J, Lee H.** 2000. Transport of bacteria on sloping soil surfaces by runoff. Environ. Toxicol. **15**:149–153. http://dx.doi.org/10.1002/(SICI)1522-7278(2000)15:2<149::AID-TOX11>3.0.CO;2-O.
46. **Ibekwe AM, Grieve CM, Papiernik SK, Yang CH.** 2009. Persistence of *Escherichia coli* O157:H7 on the rhizosphere and phyllosphere of lettuce. Lett. Appl. Microbiol. **49**:784–790. http://dx.doi.org/10.1111/j.1472-765X.2009.02745.x.
47. **Segurado P, Araújo MB, Kunin WE.** 2006. Consequences of spatial autocorrelation for niche-based models. J. Appl. Ecol. **43**:433–444. http://dx.doi.org/10.1111/j.1365-2664.2006.01162.x.
48. **Garber SM, Maguire DA.** 2003. Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. For. Ecol. Manage. **179**:507–522. http://dx.doi.org/10.1016/S0378-1127(02)00528-5.
49. **U.S. Food and Drug Administration.** 2013. Standards for the growing, harvesting, packing and holding of produce for human consumption, a proposed rule. http://www.gpo.gov/fdsys/pkg/FR-2013-01-16/pdf/2013-00123.pdf. Accessed 29 January 2014.
50. **Leafy Green Marketing Agreement.** 2013. Commodity specific food safety guidelines for the production and harvest of lettuce and leafy greens. http://www.caleafygreens.ca.gov/sites/default/files/California%20LGMA%20metrics%2008%2026%2013%20%20Final.pdf. Accessed 29 January 2014.