

# Spatially explicit uncertainty modeling of zoonotic pathogen distribution: a case of *Listeria monocytogenes* in New York State, USA

Rita Mwima<sup>1</sup> · Anthony Gidudu<sup>2</sup> · Allan Mazimwe<sup>3</sup> · Arika Ligmann-Zielinska<sup>4</sup> · Samuel Majalija<sup>5</sup> · Margaret Khaita<sup>6</sup> · John David Kabasa<sup>7</sup> · Peter Bergholz<sup>8</sup>

Received: 30 August 2015 / Accepted: 5 December 2016 / Published online: 24 December 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** *Listeria monocytogenes* is a bacterium that is responsible for causing Listeria, a disease that has a wide range of adverse effects such as meningitis, bacteremia, complications during pregnancy, and other fatal illnesses especially among those whose immune systems are compromised. The purpose of this study was to establish hotspot candidate sites in New York State where the *L. monocytogenes* pathogen could be found. Several suitability criteria which include proximity to water, pasture, forests, and urban development and slope among others in New York State were considered in this analysis. This study assessed which spatial habitat factors influence habitat suitability of the *L. monocytogenes* pathogen in the forested areas of New York State. Multicriteria evaluation was used to integrate the different habitat factors using their different weights expressed using probability distributions. Spatially explicit uncertainty and sensitivity analysis (UA and SA) was carried out to examine the robustness of

habitat suitability analysis. Suitability maps were generated and summarized using an average suitability map, a standard deviation uncertainty map, and sensitivity maps. Results showed that the shallowest depth to a wet soil layer (measured annually) and proximity to water are the habitat factors which contribute the most and individually to the distribution and survival of this pathogen. The slope gradient is singly insignificant but influential when associated with other factors like temperature, soil organic matter content, volume of water soil can store, proximity to forest, urban development, and pasture among others. It was established that water is the key habitat factor that favors the survival of this pathogen. Also, the ability to spatially model zoonotic pathogen hotspots is important in zoonoses control, informing and influencing policy. From these results, it is important to maintain the water quality of the water sources (lakes, rivers, ponds) and ensure that there is limited runoff from slopes.

✉ Rita Mwima  
tsubira2010@gmail.com

Anthony Gidudu  
anthony.gidudu@gmail.com

Allan Mazimwe  
allanmazimwe@yahoo.com

Arika Ligmann-Zielinska  
ligmannz@msu.edu

Samuel Majalija  
saraali67@gmail.com

Margaret Khaita  
mkhaita@cvm.msstate.edu

John David Kabasa  
kabasajd@yahoo.com

Peter Bergholz  
peter.bergholz@ndsu.edu

- <sup>1</sup> Makerere University, Kampala, Uganda
- <sup>2</sup> College of Engineering, Design, Art and Technology (CEDAT), Makerere University, Kampala, Uganda
- <sup>3</sup> Makerere University, CEDAT, Kampala, Uganda
- <sup>4</sup> Michigan State University (USA), Lansing, MI, USA
- <sup>5</sup> College of Veterinary Medicine, Animal resources and Biosecurity (CoVAB), Makerere University, Kampala, Uganda
- <sup>6</sup> Mississippi State University, Starkville, MS, USA
- <sup>7</sup> College of Veterinary Medicine, Animal Resources and Biosecurity, Makerere University, Kampala, Uganda, Uganda
- <sup>8</sup> North Dakota State University, Fargo, ND, USA

**Keywords** Spatial modeling · *Listeria monocytogenes* · Multi criteria evaluation · Uncertainty analysis (UA) · Sensitivity analysis (SA)

## Introduction

Zoonoses are commonly defined as diseases that are transmitted between animals and humans (WHO, 2015). A zoonotic agent may be a bacterium, fungus, or virus. At least 61% of all human pathogens are zoonotic and have represented 75% of all emerging pathogens during the last decade (WHO 2015; Lloyd-Smith Lab, n.d.). They are also responsible for the huge disease burden worldwide in terms of morbidity and mortality in humans, animals, loss of productivity, and economic losses. Zoonoses are often under-diagnosed, which reflects the limited capacity and coverage of health services (WHO 2015). Not only are they often under-diagnosed, but key aspects of zoonotic disease dynamics remain poorly understood (Lloyd-Smith Lab, n.d.).

There is a growing awareness of the increasing threats presented to humans by zoonoses especially those that originate from wildlife reservoirs (Woolhouse and Gowtage-Sequeria 2005). With changes in human living patterns, environmental and climate changes pose unprecedented challenges to the global health of people, animals, and ecosystems (Wood et al. 2012). According to Sachs et al. (2009), ecosystem health is related to human health, although their relationships remain poorly understood (Naeem 2009). Competent authorities agree that controlling a zoonotic disease through its animal reservoir is the most cost-effective intervention although some argue that increasing public awareness and education on preventive measures could also be effective in controlling the spread of zoonoses (WHO, 2015).

One of the major challenges that come with zoonoses control is the fact that it is very important to know the zoonotic pathogen reservoir and the disease transmission dynamics. Secondly, the transmission dynamics that involve wildlife are often complex and thus present a number of challenges (Alexander et al. 2012). For instance, practical characterization of wildlife host species and pathogen systems is often lacking, and insight into one system may have little application to another involving the same host species and pathogen (Alexander et al. 2012). Usually, the process by which a zoonotic pathogen moves from an animal host to a human host is referred to as spillover and occurs as a result of complex bidirectional interactions among people, animals, pathogen communities, and environments, and is a key step in the zoonotic disease spread process (Alexander et al. 2012). Each individual in the human population is not equally at risk for zoonotic pathogen and this can be proven using spatial disease models that show in which areas a particular pathogen thrives best depending on the factors that favor its survival.

Spatial modeling of zoonoses is very important in combating diseases, particularly with those who live in close proximity to both livestock and wildlife. Spatial models are used in making public health-related decisions; however, other factors such as individual and societal costs, perceived risk, strategic or policy-driven objectives, and resource allocation priorities are also essential and need to be considered when designing public health actions (Hongoh et al. 2011). In addition, the available spatial risk models for zoonoses describe the risk of exposure to a pathogen without including data on other spatially varying components such as the distribution of vulnerable human populations (Hongoh et al. 2011). Decision support tools are therefore essential in the prevention and control of zoonoses worldwide.

This study focused on developing a spatially explicit model for predicting areas more prone to the survival of *Listeria monocytogenes* bacteria in New York State, USA. *Listeria*, a zoonosis, is a serious highly fatal food borne illness caused by a gram-positive ubiquitous bacterium known as *L. monocytogenes* and one normally acquires it after eating food contaminated by this bacterium (Va'zquez-Boland et al. 2001). *L. monocytogenes* is commonly found in the environment and in food and mainly affects immune-compromised individuals such as; pregnant women, the elderly, and newborns even though those with normal immune systems can also be affected (Food Safety Authority of Ireland (FSAI) 2005).

Listeriosis is of great public importance especially in the USA because of its high mortality rate (CDC 2013; Pushkareva and Ermolaeva 2010) and its symptoms are often very severe (CDC 2013a; Va'zquez-Boland et al. 2001). The symptoms include the following: fever, muscle aches, headache, stiff neck, confusion, loss of balance, and convulsions among others, but apart from humans, *L. monocytogenes* also affects other vertebrate species like birds (CDC, 2013a; Va'zquez-Boland et al. 2001). This bacterium can be controlled and managed in the food chain by actively carrying out good hygiene practices (GHP), good manufacturing practices (GMP), and implementing a hazard analysis critical control point (HACCP) system (FSAI, 2005).

## Research objectives

There is a critical need for accurate spatial distribution information for zoonotic pathogens. The primary objective of this research was therefore to develop a spatial distribution model also known as a habitat suitability model (HSM) for the *L. monocytogenes* pathogen in New York State using quantitative tools in a geographical information system (GIS) platform. This model relied on spatial and temporal datasets to determine the distribution of the *L. monocytogenes* pathogen in the forested areas of New York State. This model will be a

predictive tool for showing areas prone to the survival of the *L. monocytogenes* bacteria and the uncertainties involved.

## Materials and methods

To study the spatial distribution of the *L. monocytogenes* pathogen in the forested areas of New York State, data used was collected as part of a larger study. Briefly, over a 2-year period (2009 and 2010), a total of 685 samples of soil, water, fecal, drag swab, and pond sediments were collected in the five areas of New York State that represent the forested areas: Finger Lakes National Forest (FLNF), Adirondack Park, Catskill Park, Connecticut Hill Wildlife Management Area (CHWMA), and Montezuma Park. These samples were obtained through two to three visits per study area during the spring, summer, and autumn. Global positioning system (GPS) data for each sample was collected and the date of sample collection was recorded. All samples were cultured for the presence of *Listeria* and Table 1 is a summary of results obtained.

The habitat factors that were used were chosen on the basis of an analysis of existing studies and expert knowledge; these aspects were used as habitat factors and were described as map layers so that each layer represented one criterion. Spatially dependent predictor data were acquired for five parks (Adirondacks, Catskills, Montezuma, Fingerlakes, and Connecticut Hill) located in New York State. The map layers for various indicators together with their sources are as shown in Table 2.

### Extraction of habitat factors from indicator map layers

Proximity data were extracted from the NLCD land cover base map by calculating the Euclidean nearest neighbor distance to the selected land cover type. Proximity to urban areas was calculated from the land cover map. Proximity to water was calculated from a map combining water body areas and flow

lines while the percent slope was derived from the DEM. A total of 15 different habitat factors were obtained for classification tree (CT) model development, such as soil type, slope, proximity to urban development, pastures, forests, and water among others. For each sample collection date, meteorological variables were obtained from the weather station nearest to that area, using the airport weather stations in the National Oceanic and Atmospheric Administration (NOAA) National Climate Data Center (NCDC) Local Climatology Database. This study aimed at capturing the association between remotely sensed meteorological data and *L. monocytogenes* prevalence. Meteorological data included temperature (maximum, minimum, and daily average) and precipitation amounts, with the average temperature, thaw cycles, and precipitation calculated for each time period, ranging from 1 to 10 days before sample collection (for a complete list of habitat factors considered together with their description, see Appendix A1). All the habitat factors were projected into the Universal Transverse Mercator (UTM) coordinate system, North American Datum of 1983.

### Assigning criteria weights to habitat factors

Classification and regression tree (CART) analysis, a data mining tool was used to reveal the relationship between *L. monocytogenes* bacteria and the habitat factors that were used. A CT method was used with CART using the Gini splitting criteria. Several classification trees (which result in random forests) were used to obtain the most important habitat factors for consideration. All the habitat factors were included in the classification tree analysis as possible predictors of the presence of *L. monocytogenes* pathogens and tree-based modeling were used to determine the rules that classified sites by pathogen presence or absence.

### Criteria standardization (normalization) and valuation

The habitat factors that were used in this study were measured on different measurement scales and therefore in order to address the problem of incommensurate measurement scales, the criteria and their raw values were transformed into a common scale using a criterion standardization procedure. Standardization ensures that all the habitat criteria layers are presented in a [0.0, 1.0] scale. It was however important to maintain proportionality between the raw and standardized values to ensure that no unnecessary distortion was introduced (Ligmann-Zielinska and Jankowski 2014). Both linear and score range transformations utilize benefit and cost valuations, and for benefit valuation, higher values for a habitat factor are better

**Table 1** Sample types and number of samples collected in NYS

Sample type	Number of samples	Number of samples positive for <i>L. monocytogenes</i>
Fecal	418	42
Water	84	1
Soil	84	7
Drag swab	84	5
Pond sediment	15	1

**Table 2** Indicator data for habitat suitability modeling

	Indicator	Data source
1	National land cover (NLCD)	<a href="http://earthexplorer.usgs.gov/">http://earthexplorer.usgs.gov/</a>
2	Digital elevation model (DEM)	<a href="http://earthexplorer.usgs.gov/">http://earthexplorer.usgs.gov/</a>
3	Soil map layers	<a href="http://soils.usda.gov/survey/geography/ssurgo/">http://soils.usda.gov/survey/geography/ssurgo/</a>
4	Road maps	<a href="http://cugir.mannlib.cornell.edu/">http://cugir.mannlib.cornell.edu/</a>
5	Hydrologic line graphs	<a href="http://cugir.mannlib.cornell.edu/">http://cugir.mannlib.cornell.edu/</a>
6	Meteorological data (temperature and precipitation)	<a href="http://gis.ncdc.noaa.gov/map/lcd/">http://gis.ncdc.noaa.gov/map/lcd/</a>

while for cost valuation, lower values of a habitat factor are better (Malczewski, 1999).

Normalised habitat factor

$$= \left( \frac{(X - \min) * (\max_{\text{norm}} - \min_{\text{norm}})}{\max - \min} \right) + \min_{\text{norm}} \quad (1)$$

where  $X$  is the habitat factor to be normalized,  $\max$  and  $\min$  are the maximum and minimum values of old value range.  $\max_{\text{norm}}$  and  $\min_{\text{norm}}$  are maximum and minimum values of the new value range.

The normalized habitat factors and their weights were combined based on multicriteria evaluation (MCE) using a priority function given by  $P$ :

$$P = \sum_{i=1}^m a_i p_i (q_i) \quad (2)$$

(Store and Kangas 2001).

where  $P$  is the habitat suitability index,  $m$  the number of factors,  $a_i$  the relative importance of the habitat factor,  $p_i$  the sub-priority of function  $i$ , and  $q_i$  the amount of factor  $i$  (Store

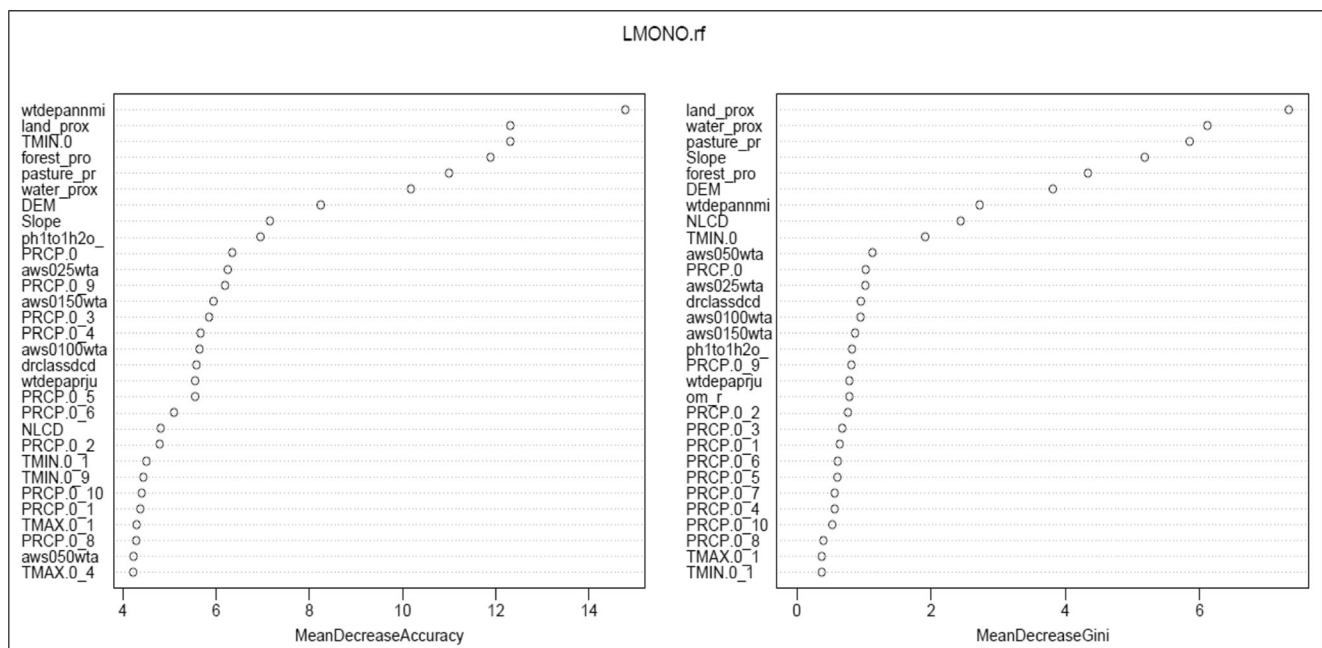
and Kangas, 2001). The relative importance of each habitat factor was acquired using classification trees and were used as coefficients  $a_i$  in this study.

### Weighting of habitat factors

The weights that were used in integrating habitat factors as a means to assess the contributions of each factor to the combined suitability index were derived by dividing the score of that particular factor by the total score of all factors. The reason for habitat factor weighting was to express the importance of each factor relative to other factors, with the more important factors having greater weights in the overall evaluation.

### MCE-weighted linear combination

This study adopted a multicriteria evaluation that combines the habitat factors responsible for the survival of the



**Fig. 1** Below, are the variable importance plots for the habitat variables from random forests (RF) classifications which were used for predicting the presence of the *L. monocytogenes* pathogen in New York State, USA. Higher values of mean decrease accuracy show the habitat factors that are important for the pathogen’s survival.

**Table 3** Ten-point weighting scale for variable importance

Intensity of importance	Description	Suitability class
4	Equal importance	Lowest suitability
5	Equal to moderate importance	Very low suitability
6	Moderate importance	Low suitability
7	Moderate to strong importance	Moderately low suitability
8	Strong importance	Moderate suitability
9	Strong to very strong importance	Moderately high suitability
10	Very strong importance	High suitability
11	Very strong importance	High suitability
12	Very to extremely strong importance	Very high suitability
15	Extremely important	Highest suitability

*L. monocytogenes* pathogen using weighted linear combination (as is seen in Eq. 3) to a single aggregated index. These different habitat factors contribute differently depending on their weights to the overall habitat suitability.

$$V(x_i) = \sum_i w_i v_i(x_i) \tag{3}$$

(Kienberger et al. 2009)

*V* is the layer being considered, *x<sub>i</sub>* the variable indicators for suitability i.e., *x<sub>1</sub>*, *x<sub>2</sub>*, *x<sub>3</sub>*, ... , *x<sub>n</sub>*, *w<sub>i</sub>* the spatial weight of feature *i*, and *v<sub>i</sub>* the suitability index for each unit calculated from the formula given in Eq. (9). To generate habitat suitability partitions, the geons approach (Kienberger et al. 2009) which is based on regionalization concepts in OBIA was used to define habitat suitability units according to the given habitat factors (Lang et al. 2006). A region-based local mutual best fitting approach that combines image segments according to the gradient of degree of best fit (Kienberger et al. 2009) allows for controlling two corresponding criteria of similarity neighboring segments—likeness in “color or form” (Mazimwe, 2013). A multiresolution segmentation technique (in E-cognition) which organizes raster artifacts into image objects in a spatial hierarchy was used to fragment the habitat factor datasets (Mazimwe, 2013). The distance between pixels *P<sub>1</sub>* and *P<sub>2</sub>* in two image objects adjacent to each other in a feature space were expressed in terms of

$$h = \sqrt{\sum_{d=1}^n (P_1 - P_2)^2} \tag{4}$$

in a two-dimensional feature space (Mazimwe 2013). This distance is also expressed by the spectral distance SD of the two pixels *p<sub>1</sub>* and *p<sub>2</sub>* in a feature space expressed as:

$$SD = \sqrt{\sum_{d=1}^n (p_1 - p_2)^2} \tag{5}$$

(Kienberger et al. 2009)

or as a vector difference for a three-dimensional feature space expressed as:

$$SD = \left( \vec{v}_1 - \vec{v}_2 \right) \tag{6}$$

where  $\vec{v}_1 = (d_{11} \ d_{12} \ d_{13})$  and  $\vec{v}_2 = (d_{21} \ d_{22} \ d_{23})$

Specify the feature space location of pixel 1 and pixel 2 or the average value of object 1 and 2 (Kienberger et al. 2009). To optimize the degree of homogeneity between neighboring objects, the degree of fitting two adjacent image objects is defined by describing the change of heterogeneity *h<sub>diff</sub>* which is minimized at every merge (Baatz and Schape 2000).

$$h_{diff} = h_{min} - \frac{SD_1 + SD_2}{2} \tag{7}$$

To fulfill the requirement of producing objects of similar area, additional weights have to be added to the habitat factors with different object size (Kienberger et al. 2009). In order for form homogeneity to be realized, object boundary length (perimeter) is related to the perimeter of a circle, the deviation of which can be expressed by the shape index (Kienberger et al. 2009; De Smith et al. 2013)

$$Shape = \frac{P}{2\sqrt{\pi \times S}} \tag{8}$$

where *P* is the perimeter and *s* is the object size. A compactness ratio of 0.5 was therefore used for a scale level of 20. This approach (regionalization approach) provides algorithms that allow weighting of the different layers and the suitability index for each unit was calculated from the formula

$$|\vec{v}| = \sqrt{v_1^2 + v_2^2 + v_3^2 + \dots v_n^2} \tag{9}$$

for layers *v<sub>1</sub>*, *v<sub>2</sub>*, *v<sub>3</sub>*... to *v<sub>n</sub>* in a *j*th-dimensional space through the vector product (Mazimwe 2013). The habitat factors were standardized within the range of 0 and 1, whereas 1 reflects a high suitability and 0, low suitability.

### Uncertainty and sensitivity analysis

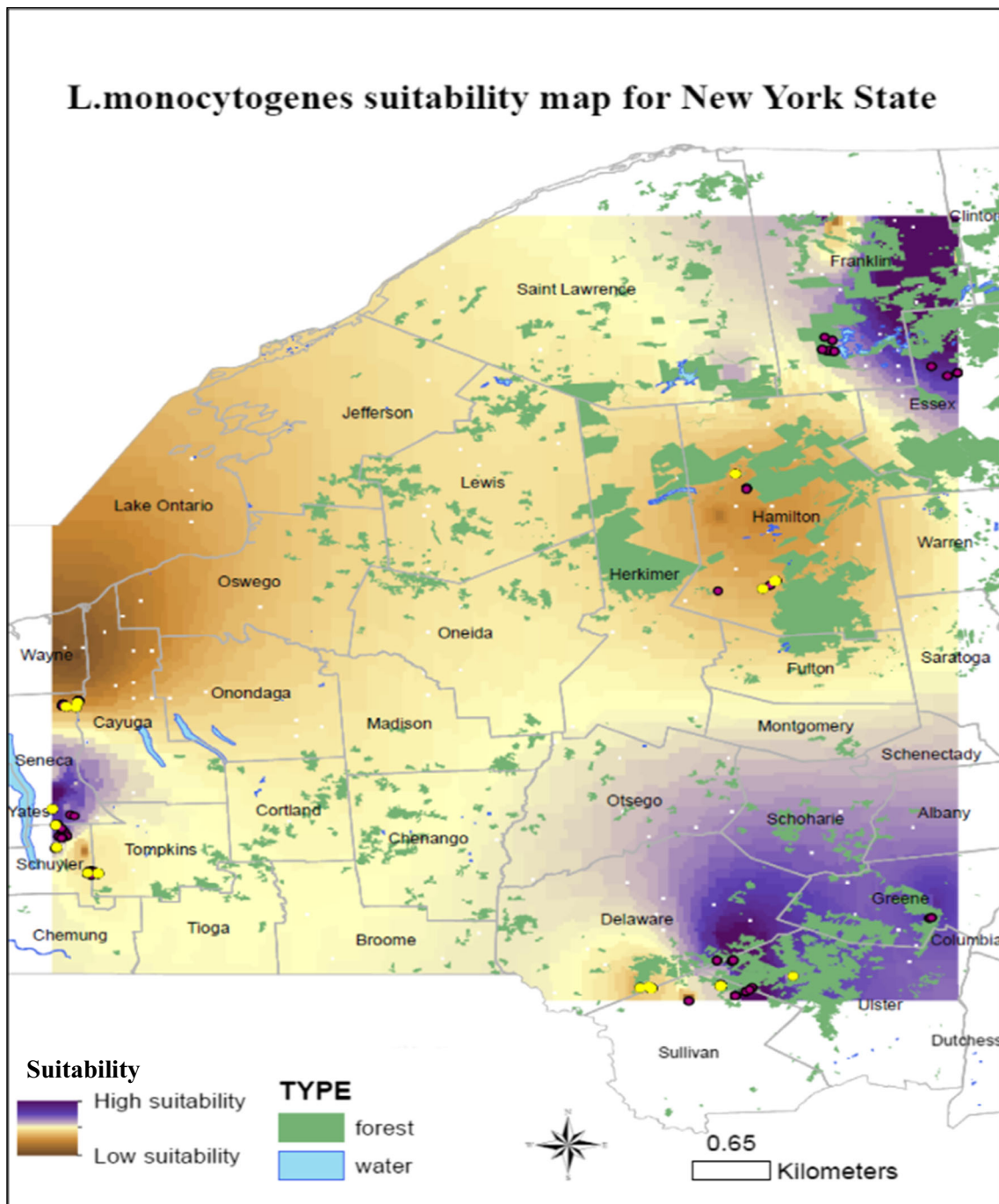
Spatially explicit uncertainty and sensitivity analyses were then applied using the integrated uncertainty and sensitivity analysis (iUSA) approach to examine the validity of habitat suitability evaluation. Monte Carlo simulation was used to traverse through criteria weight space, where weights were expressed using probability distributions.

Integrated uncertainty and sensitivity analysis quantified the variability of outcomes of MCE (uncertainty analysis) and identified which habitat factors were responsible for this variability (sensitivity analysis) (Ligmann-Zielinska and Jankowski 2014). This method resulted in the identification of habitat factors that require more attention and got rid of the factors that have little influence on outcome uncertainty (Saltelli and Annoni 2010; Lilburne and Tarantola 2009; Saltelli et al. 2008).

## Results and discussion

### Classification and regression tree analysis

From this analysis, a variable importance plot was attained. The variable importance plot which shows the habitat factors in their order of importance is shown in Fig. 1. The variable importance plot was generated to assess the importance of the selected habitat factors. The importance score for each habitat factor was

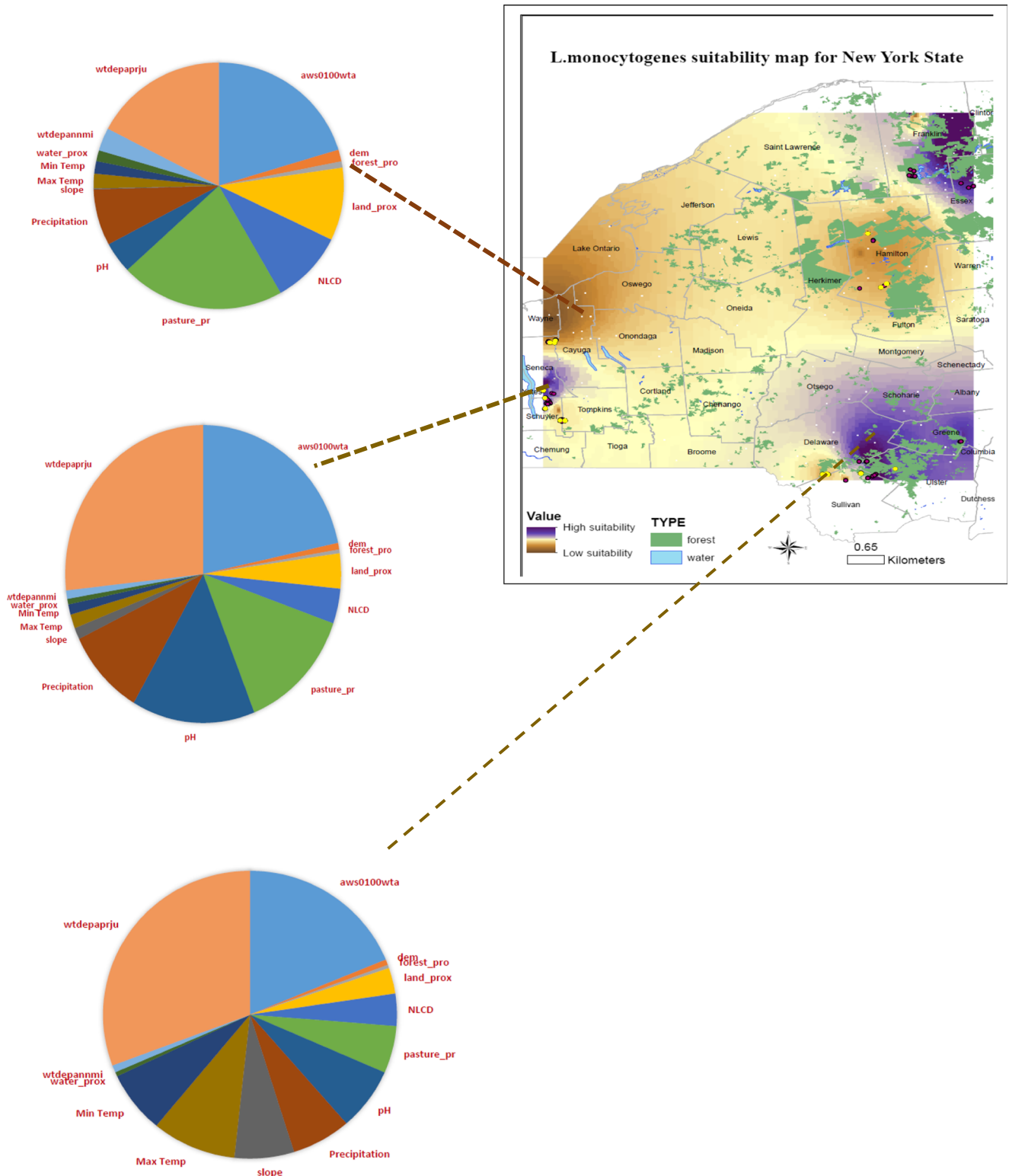


**Fig. 2** *L. monocytogenes* habitat suitability map for the forested areas of New York State

calculated by dividing the intensity of importance by 15 which is the highest value after which the result was multiplied by 100. Table 3 is a ten-point weighting scale for variable importance which explains the intensity of importance of each habitat factor.

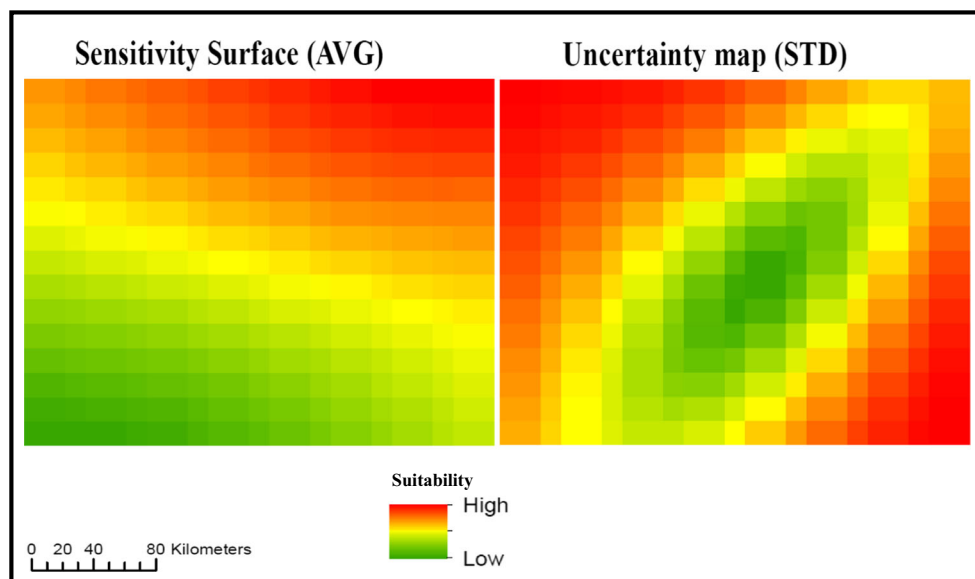
**MCE-weighted linear combination**

The result in Fig. 2 is a habitat suitability map of *L. monocytogenes* hotspots prone to listeria epidemics in the



**Fig. 3** *L. monocytogenes* habitat suitability map for the forested areas of New York State showing the values each factor contributed at a particular area

**Fig. 4** Results of uncertainty analysis (UA)



sense of habitat factors that contribute to the survival of the pathogen, obtained from integrating habitat suitability factors using MCE. Figure 2 shows the habitats for the *L. monocytogenes* pathogen with high suitability in the southern part of Montezuma National Forest (made up of Seneca, Cayuga, and Wayne counties), North Eastern part of Adirondacks Park (consists of Essex, Franklin, Hamilton, and St. Lawrence counties), North Western part of Finger Lakes National Forest (consists of Seneca, Cayuga, Schuyler and Tompkins counties), Northern part of Connecticut Hill Wildlife Management Area (consists of Tompkins and Schuyler counties) and the whole of Catskills Park (consists of Delaware, Greene, Sullivan and Ulster counties) except North Western part. The counties with the highest suitability are Franklin, Delaware, Greene, and Ulster while those with the least suitability are Hamilton, Wayne, and Tompkins. Regarding the characteristics of the suitable habitats as shown from the pie charts in Fig. 3, it is apparent that *wtdpannmi* (shallowest depth to a wet soil layer measured annually), *water\_prox* (proximity to water) and *forest\_pro* (proximity to forests) strongly contribute to an increase in suitability values in the Southern part of Montezuma National Forest and the Northern, Western, Southern, and Eastern parts of Catskills Park (lesser values of these habitat factors contribute the most to survival of the *L. monocytogenes*).

In addition, other habitat factors that significantly influence the suitable habitats include slope, precipitation, pH, proximity to pasture, minimum temperature and proximity to water among others. Figure 3 also shows the value each factor contributed at a particular spot and from these values, the ten habitat factors that contributed the

most were chosen and considered in the uncertainty and sensitivity analyses (UA and SA). Also to note, the yellow dots shown in Figs. 2 and 3 represent samples that tested positive for *L. monocytogenes*. The purple dots on the other hand represent samples that tested negative for the same pathogen.

From the suitability map attained (as seen in Fig. 2), it is important to note that the suitability values were independent of the administrative boundaries, although these hotspots could influence the different listeria outbreaks in and out of New York State. The usual HSMs depend on the fact that the observed geographical distribution of a species reveals its ecological requirements (Hirzel and Lay 2008). Most of them assume that species are present in suitable habitats and absent from unsuitable ones. However, in practice, the relationship between niche and distribution is uncertain (Hirzel and Lay 2008). Hirzel and Le Lay (2008) explain that a species may be absent from a suitable site or present in an unsuitable site like what is shown on the suitability map using the samples that were collected from particular areas and were later tested for the presence of the *L. monocytogenes* pathogen. Unrecorded presences may be as a result of (i) incomplete sampling, (ii) selective sampling (iii) cryptic or rare species, or (iv) faulty determinations, and although time, effort, and skill may minimize this problem, it is often unavoidable (Hirzel and Lay 2008). These uncertainties in model predictions may lead to assumptions about data accuracy and outputs that are invalid and in turn impact the practices and decisions made thereafter (Regan et al. 2002). Integrated uncertainty and sensitivity analyses (iUSA) were therefore used to evaluate the degree of

uncertainty and also categorize sensitive parameters for habitat classifications and associated maps generated from expert opinion (Gillingham and Johnson 2004).

### Sensitivity and uncertainty analysis

Ten standardized habitat factors (input criteria maps) were considered for sensitivity and uncertainty analyses (Fig. 4). These factors are expressed as raster surfaces of 15 rows and 23 columns (hence 345 cells) with value range from 0.0 (worst) to 1.0 (best) and in ascii (text) format. The analysis is based on the premise that it would be easier to find the *L. monocytogenes* pathogen in areas with minimum temperature (temp), low pH (ph), close to wet soil layers (depth), high vertical elevation (elev), high slope values (slope), high amounts of rainfall (prec), and that are closer to pasture (pastureprox), water (waterprox), forests (forestprox), and urban development (landprox).

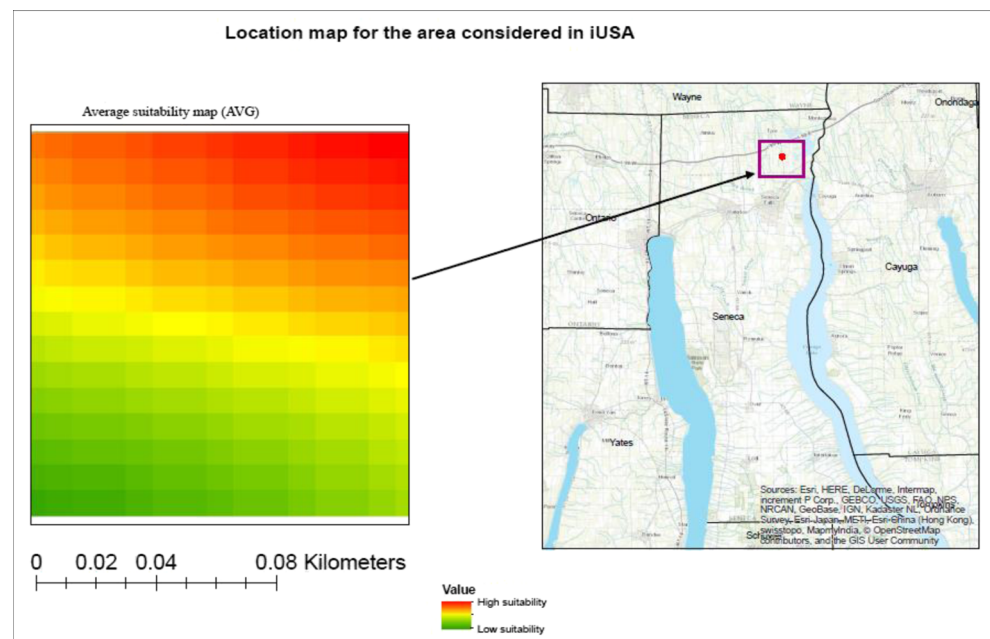
To account for possible criteria correlations, the ideal point (IP) aggregation function was used (Ligmann-Zielinska et al. 2012). For Monte Carlo simulations, Sobol's experimental design with 5376 samples and 64,512 model executions (runs) was used (Saltelli et al. 2010). The assumption is that the ten weights are all individually drawn from an identical probability distribution function (PDF) with a range of [0.0, 1.0] (Ligmann-Zielinska et al. 2012). Also, to note is that because of the high computational cost, a very small area was considered for this analysis as shown in Fig. 5. Considering a

larger area would require supercomputing to generate the sensitivity maps.

### Uncertainty analysis

From the Monte Carlo (MC) simulations, multiple output suitability maps were generated and summarized to come up with an average suitability surface (AVG) and an uncertainty surface which represents a standard deviation of suitability maps (STD). Figure 4 shows a habitat suitability surface (left) of *L. monocytogenes* averaged over all the MC runs and its computed ideal point (IP) scores (from the ideal point aggregation function) which are within the range of 0.3 to 0.6. The results of MC simulations were summarized by calculating two summary suitability surfaces presented in Fig. 4. The map on the left depicts an average habitat suitability surface of *L. monocytogenes*, which was calculated as the mean of all MC runs. The red parts represent the areas which are very suitable for the survival of *L. monocytogenes*, while the rest represent the less suitable areas. On the other hand, the red parts on the uncertainty map represent high uncertainty, while the less red parts represent low uncertainty. The computed suitability scores fall within the 30 to 60% interval of the normalized suitability score range (0–100%). In regard to the areas which were of high suitability, the slope factor had high values and this can be attributed to the uniform PDFs allocated to each habitat factor. Therefore, the spatial distribution of high slope

**Fig. 5** Location map for the area considered in iUSA



values is complemented by high values of other factors, resulting in a similar spatial pattern of output high habitat suitability.

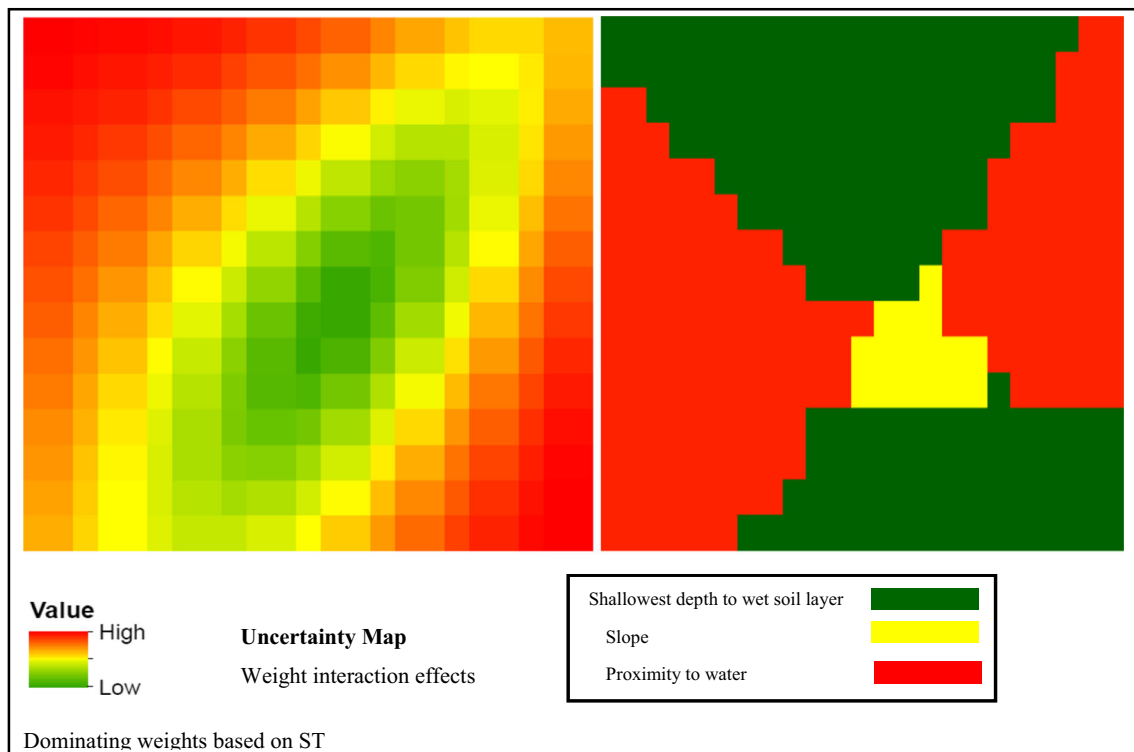
From the average suitability and uncertainty maps, it was assumed that the minimum AVG score for high suitability should be 0.4. The other assumption was that areas with STD values above 0.009 have a relatively high suitability uncertainty. From the AVG and STD maps, the robust areas are those with  $AVG \geq 0.4$  and  $STD < 0.009$  (also known as high-low areas) while the candidate areas are those with  $AVG \geq 0.4$  and  $STD \geq 0.009$  (also known as high-high areas). The high-low (HL) areas are robust for the survival of the *L. monocytogenes* pathogen while the high-high (HH) areas are the areas with high uncertainty and have less robust suitability scores.

When the suitability and uncertainty surfaces were visualized in tandem, the robust and candidate sites were confirmed. The robust sites were the suitable areas with less uncertainty while the candidate sites were those with high uncertainty. The AVG map (right of Fig. 4) provides a partial interpretation of habitat suitability, because the STD surface (uncertainty surface on the left of Fig. 4) has some high suitability sites which are also characterized by a relatively high uncertainty linked to spatial distributions of suitability criteria (Ligmann-Zielinska and Jankowski 2014). From UA, the upper left (UL) and upper right

(UR) areas of the sensitivity surface are robust for the presence of *L. monocytogenes* (Fig. 4). However, the upper left (UL) area on the uncertainty map has high STD values. These high-high (HH) areas would mean that they have less robust suitability scores, thus they may be suitable for the survival of the *L. monocytogenes* pathogen but need further analysis because of their uncertain suitability scores. To find out which habitat factors were responsible for this uncertainty, sensitivity analysis (SA) was carried out.

### Sensitivity analysis

Using SA, the influence of each habitat factor on shaping the uncertainty of suitability scores was determined. Here, the variability of suitability maps was broken down and assigned to every input weight and as a result, one first-order (S) and total-effect (ST) sensitivity index per weight was generated. The S-maps are shown in Appendix A2. These S-maps are quantitatively different in that if one weight has a high ST value, the other weights score lower. Ten ST maps were overlaid, after which the space was divided into regions of dominating weights as is seen on the right of Fig. 6. The uncertainty of high average suitability scores in the left half of the area is formed by the weights associated with *wtdepanmi* (shallowest depth to



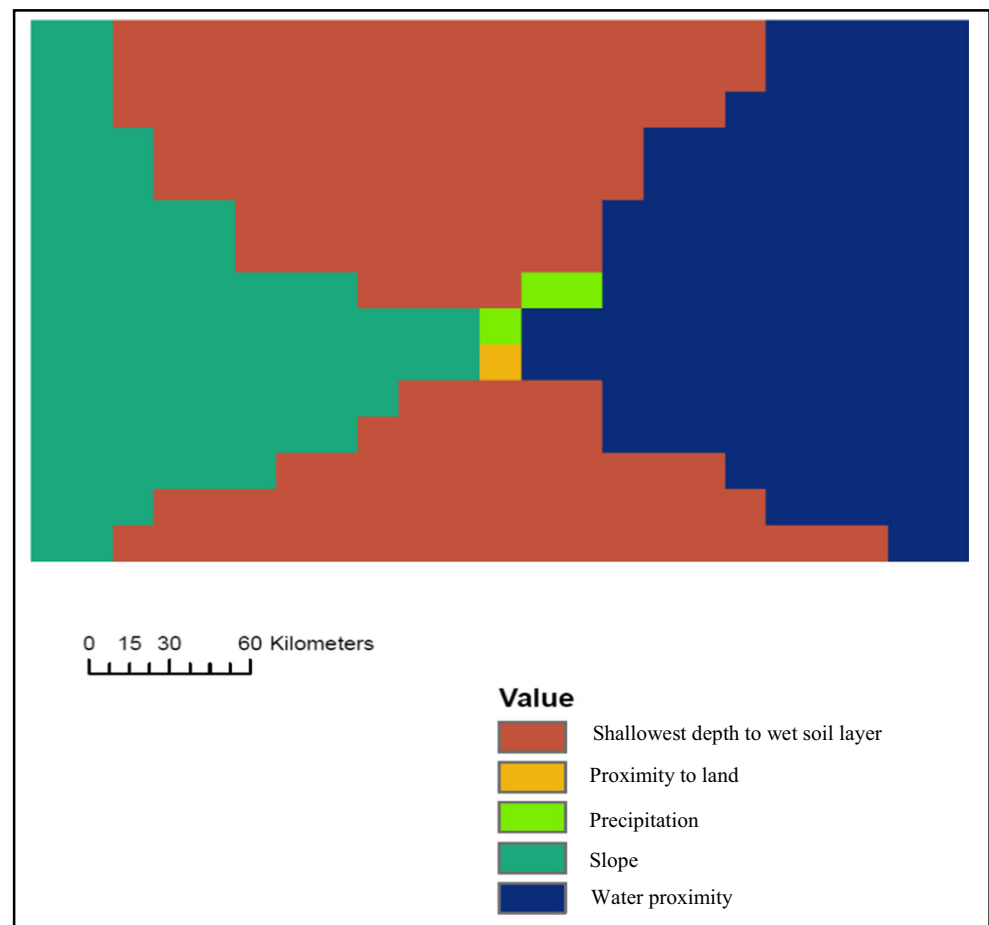
**Fig. 6** Results of sensitivity analysis

a wet soil layer) and `water_prox` (proximity to water), while the slope values are quite low in this area. The areas of high uncertainty were explained by individual weights alone (with low weight interaction effects) while the areas of high average suitability and low uncertainty have high weight interactions and this suggests that these interactions have a diminishing effect on the uncertainty of habitat suitability.

`Wtdepannmi` (shallowest depth to wet soil layer), `forest_pro` (proximity to forest), `ph1to1h2o_` (soil pH), and `slope` have a positive spatial linear correlation, where the distribution of high values of input habitat criteria is matched by the distribution of high values of their respective sensitivities ( $S$  values). A negative correlation between inputs and the equivalent sensitivities was observed for `elev` (elevation), `urban_prox` (proximity to urban development), `pasture_pr` (proximity to pasture), `PRCP_0` (precipitation), `TMIN_0` (minimum temperature), and `water_prox` (proximity to water). The distribution of high values of input habitat criteria does not match with the distribution of high values of their respective sensitivities ( $S$  values).

To determine the habitat factor with the maximum sensitivity value ( $S$ ) on a cell-by-cell basis, all the  $S$ -maps (first-order maps) were overlaid. All the space was partitioned into regions of dominating weights, and as a result, came up with a weight dominance map shown in Fig. 6. The weights for `wtdepannmi` (shallowest depth to wet soil layer), `water_prox` (proximity to water) and `slope` are responsible for the majority of uncertainty associated with high suitability scores. From Fig. 7, it is apparent that `wtdepannmi` and `water_prox` were the most influential habitat factors individually, while `PCRCP` (precipitation) and `Land_prox` (proximity to urban development) also contribute to this uncertainty, although not vastly. Elevation, `forest_pro` (proximity to forests), `pasture_pr` (proximity to pasture), `pH`, and `TMIN_0` (minimum temperature) do not completely influence this uncertainty. `Wtdepannmi` (shallowest depth to a wet soil layer annually) and proximity to water (`water_prox`) are the dominating habitat factors that determine the suitability habitat for the *L. monocytogenes* pathogen because they had the highest  $S$  values. When the difference between the first-order ( $S$ ) and total-effect ( $ST$ ) sensitivity indices of each habitat factor expressed as  $STi-Si$

**Fig. 7** Areas of dominant weight sensitivities ( $S_i$ )



was considered, slope had the highest value which meant that it was highly involved in interactions with other habitat factors. Other habitat factors like precipitation, proximity to water, and proximity to urban development in this order are influential when involved in interactions with other factors. Also, to note is the fact that even though the slope factor has a relatively low value of ST as shown in Fig. 6 (right side), it could influence the model output through its interactions with other factors. Figure 7 shows areas of dominant weight sensitivities and the three habitat factors that stand out are wtdepanmi (shallowest depth to a wet soil layer), slope and water\_prox (proximity to water), while land\_prox (proximity urban development) and precipitation have the least score.

These results concur with studies on the *L. monocytogenes* pathogen that have been carried out before. For example, according to Jiang et al. (2002); Sjogren (1994); Entry et al. (2005), pathogen survival time in the soil normally varies from 4 to over 250 days, and this strongly reveals the pathogen's ability to survive under adverse conditions. The obligate parasites (cannot complete their life cycle without exploiting a suitable host) usually only live for a few minutes outside their host, whereas many enteric pathogens can survive in ground water and soil for months (Vinten et al. 2004; Guber et al. 2005; Unc and Goss 2006; Entry et al. 2005). Furthermore, runoff and groundwater from agricultural land shows that these enteric bacteria increase in spring flows and decrease in the dry period (Spackman et al. 2003; Berry and Miller 2005; Guber et al. 2005; Unc and Goss 2006). Also according to Entry et al. (2000a) and Entry et al. (2000b), survival of bacterial pathogens in soil increases in moist warm soil. Studies by Lehnert (1960) and Welshimer (1960) have also shown that survival of *Listeria* in soil is influenced by moisture content. In addition to the results obtained in this study, it is evident that when the slope gradient interacts with other factors, its contribution also becomes significant.

This therefore explains the pathogen increase in spring flows and decrease in dry periods (Spackman et al. 2003; Berry and Miller 2005; Guber et al. 2005; Unc and Goss 2006). The S-maps are quantitatively different, if one weight has a high ST value, the other weights score lower. This is because every ST map renders the fractional contribution of a particular weight to the total unconditional variance of the average suitability map (Ligmann-Zielinska et al. 2012).

## Conclusions

The most important habitat factors were obtained using CART analysis, in which the ten most influential habitat factors for the survival of the *L. monocytogenes* pathogen were the shallowest depth to a wet soil layer (measured annually), proximity to urban development, minimum temperature on the day samples

are collected, proximity to forests, proximity to pasture, proximity to water, elevation, slope, soil pH and precipitation.

From MCE, it was evident that the areas closest to water and those very close to a wet soil layer were the most suitable for the survival of the *L. monocytogenes* pathogen. But because of the uncertainties that surround HSMs, uncertainty and sensitivity analysis confirmed that low values of shallowest depth to a wet soil layer and proximity to water singly contribute the most to the survival of the *L. monocytogenes* pathogen. The slope gradient on the other hand greatly contributes to the distribution of *L. monocytogenes* when involved in interactions with other habitat factors.

## Recommendations

This spatial model showing the hot spots for *Listeria monocytogenes* should be applied in New York State to improve disease surveillance and also ensure that measures are taken to prevent the transfer of this pathogen to food and drinks. The fact that the shallowest depth to water and proximity to water are the most important habitat factors that singly contribute the most to the survival of the *L. monocytogenes* pathogen implies that the soil water content is very important and therefore, using suitable treatment strategies to maintain the quality of lakes, ponds, and streams is essential in keeping them free of this pathogen. It is also important to endeavor to control water runoff by for example growing trees in areas with steep slopes (areas with high elevation values).

There is need for collaboration between various stakeholders to ensure that control measures for the spread of *L. monocytogenes* in New York State are put in practice. These measures will require combined efforts among the ecologists, epidemiologists, public health scientists and policy makers among others. Such collaborations help in ensuring that more research in this area of spatial model development for pathogen hotspots is advanced.

With the growing number and severity of zoonoses, it is important to put the time factor into perspective. Future research could therefore consider habitat suitability modeling in real time to predict pathogen presence and related impacts. This model can also be transformed to suite any other disease and this will help to produce predictive HSMs for various disease pathogens and as a result, public health will be improved locally, in Africa and worldwide.

There is need to build capacity in spatial disease modelling especially in Africa. This will require setting up of training centers and purchase of high performance computing servers that aid in processing large datasets and in turn ensure that accurate predictions are made in the disease control sector. This will also aid in improving spatial epidemiology for better disease control and prevention strategies.

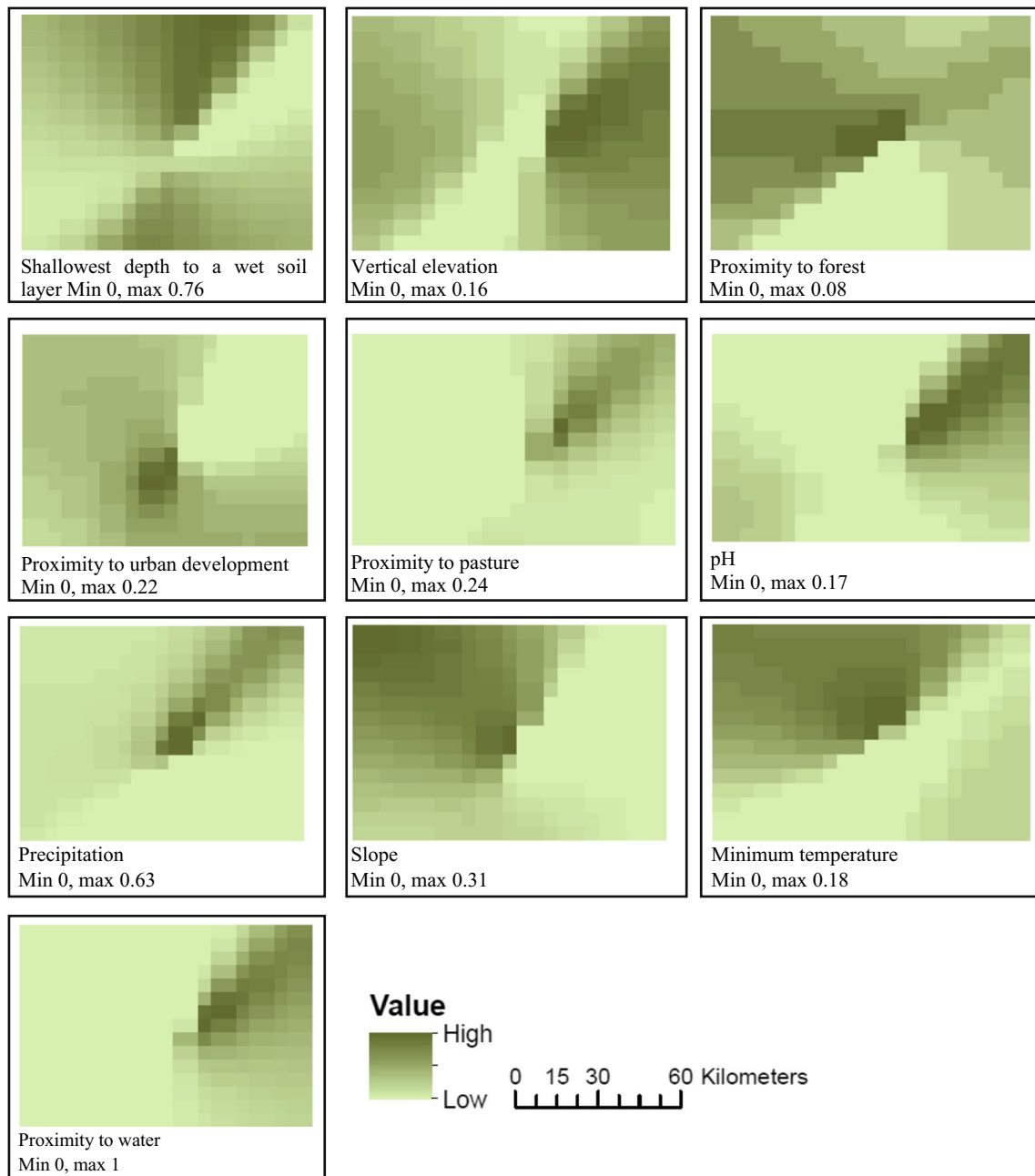
## Appendix

**Table 4** Habitat factors, their descriptions, and units of measurements

Habitat factor	Description	Unit
Aws025wta	Volume of water that the soil to a depth of 25 cm can store	cm
Aws050wta	Volume of water that the soil to a depth of 50 cm can store	cm
Aws0100wta	Volume of water that the soil to a depth of 100 cm can store	cm
Aws0150wta	Volume of water that the soil to a depth of 150 cm can store	cm
Om_r	Total organic matter content for the soil layer	% dry weight
water_prox	Proximity to water	m
Urban_prox	Proximity to urban development	m
Pasture_pr	Proximity to pasture	m
Forest_pro	Proximity to forest	m
TMIN_0	Minimum temperature on the specified day, t0	°F
TMIN_1	As for TMIN_0, but 1 day before day, t1	°F
TMIN_2	As for TMIN_0, but 2 days before day, t2	°F
TMIN_3	As for TMIN_0, but 3 days before day, t3	°F
TMIN_0_1	Average minimum temperature for the specified time period, t0–t1	°F
TMIN_0_2	As for TMIN_0_1, but t0–t2	°F
TMIN_0_3	As for TMIN_0_1, but t0–t3	°F
TMIN_0_4	As for TMIN_0_1, but t0–t4	°F
TMIN_0_5	As for TMIN_0_1, but t0–t5	°F
TMIN_0_6	As for TMIN_0_1, but t0–t6	°F
TMIN_0_7	As for TMIN_0_1, but t0–t7	°F
TMIN_0_8	As for TMIN_0_1, but t0–t8	°F
TMIN_0_9	As for TMIN_0_1, but t0–t9	°F
TMIN_0_10	As for TMIN_0_1, but t0–t10	°F
TMAX_0	Maximum temperature on the specified day, t0	°F
TMAX_1	As for TMAX_0, but 1 day before day, t1	°F
TMAX_2	As for TMAX_0, but 2 days before day, t2	°F
TMAX_3	As for TMAX_0, but 3 days before day, t3	°F
TMAX_0_1	Average maximum temperature for the specified time period, t0–t1	°F
TMAX_0_2	As for TMAX_0_1, but t0–t2	°F
TMAX_0_3	As for TMAX_0_1, but t0–t3	°F
TMAX_0_4	As for TMAX_0_1, but t0–t4	°F
TMAX_0_5	As for TMAX_0_1, but t0–t5	°F
TMAX_0_6	As for TMAX_0_1, but t0–t6	°F
TMAX_0_7	As for TMAX_0_1, but t0–t7	°F
TMAX_0_8	As for TMAX_0_1, but t0–t8	°F
TMAX_0_9	As for TMAX_0_1, but t0–t9	°F
TMAX_0_10	As for TMAX_0_1, but t0–t10	°F
PRCP_0	Amount of rain on the specified day of collection	mm
PRCP_1	As for PRCP_0, but 1 day before day, t1	mm
PRCP_2	Amount of rain 2 days before day of collection, t2	mm
PRCP_3	As PRCP_2, but 3 days before day, t3	mm
PRCP_0_2	Average precipitation for the specified time period, t0–t2	mm
PRCP_0_3	As for PRCP_0_2, but t0–t3	mm
PRCP_0_4	As for PRCP_0_2, but t0–t4	mm
PRCP_0_5	As for PRCP_0_2, but t0–t5	mm
PRCP_0_6	As for PRCP_0_2, but t0–t6	mm
PRCP_0_7	As for PRCP_0_2, but t0–t7	mm
PRCP_0_8	As for PRCP_0_2, but t0–t8	mm
PRCP_0_9	As for PRCP_0_2, but t0–t9	mm
PRCP_0_10	As for PRCP_0_2, but t0–t10	mm
FT_0	No. of freeze/thaw cycles on the specified day, t0	NA
FT_1	As for FT_0, but 1 day before day, t1	NA
FT_2	As for FT_0, but 2 days before day, t2	NA
FT_3	As for FT_0, but 3 days before day, t3	NA
FT_0_1	No. Freeze/thaw cycles for the specified time period, t0–t1	NA
FT_0_2	As for FT_0_1, but t0–t2	NA
FT_0_3	As for FT_0_1, but t0–t3	NA
FT_0_4	As for FT_0_1, but t0–t4	NA
FT_0_5	As for FT_0_1, but t0–t5	NA
FT_0_6	As for FT_0_1, but t0–t6	NA
FT_0_7	As for FT_0_1, but t0–t7	NA
FT_0_8	As for FT_0_1, but t0–t8	NA
FT_0_9	As for FT_0_1, but t0–t9	NA
FT_0_10	As for FT_0_1, but t0–t10	NA
Wtdepanmi	Shallowest depth to a wet soil layer (annual)	cm
Wtdepaprju	Shallowest depth to a wet soil layer (April to June)	cm
ph1to1h2o_	Surface soil pH	NA
Slope	Difference in elevation between two points, expressed as a percentage	%
Elev	Vertical elevation	m

t0 denotes the day of sample collection, t1 the day before and so on until 10 days before sample collection (t10)

NA not applicable



**Fig. 8** S-maps (first-order sensitivity index maps) for the habitat factors

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Alexander A,K, Lewis L,B, Marathe M, Eubank S, Blackburn K,J (2012) Modeling of wildlife-associated zoonoses: applications and caveats. *Vector-borne and zoonotic diseases* 12(12):1005–1018
- Baatz, M. and Schäpe, A. (2000) Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. *Proceedings of the Angewandte Geographische Informationsverarbeitung XII*. In: Strobl J, Blaschke T, Griesebner G, et al. (eds) *Beiträge zum AGIT Symposium Salzburg, 12–23, 2000*
- Berry E,D, Miller D,N (2005) Cattle feedlot soil moisture and manure content: II. Impact on *Escherichia coli* O157. *J Environ Qual* 34(2): 656–663
- CDC (2013) Listeriosis—statistics. Available at: <http://www.cdc.gov/listeria/statistics.html>. Accessed 2nd October 2013
- De Smith M, Longley P, Goodchild M (2013) *Geospatial analysis: a comprehensive guide to principles, techniques and software tools. 4th Edition, the Winchelsea Press 2006–2013* Dessai, S. and Hulme

- M. 2004. —Does climate adaptation policy need probabilities. *Clim Pol* 4:107–128
- Entry J,A, Hubbard R,K, Thies J,E, Fuhmann J,J (2000a) The influence of vegetation in riparian filterstrips on coliform bacteria I. Movement and survival in surface flow and groundwater. *J Environ Qual* 29:1206–1214
- Entry J,A, Hubbard R,K, Thies J,E, Fuhmann J,J (2000b) The influence of vegetation in riparian filterstrips on coliform bacteria: II. Survival in soils. *J Environ Qual* 29(4):1215–1224
- Entry J,A, Leytem A,B, Verwey S (2005) Influence of solid dairymanure and compost with and without alumon survival of indicator bacteria in soil and on potato. *Environ Pollut* 138(2):212–218
- Food Safety Authority of Ireland (FSAI) (2005) Report of the control and management of *Listeria monocytogenes* contamination of food. Abbey Court, Lower Abbey Street, Dublin
- Gillingham, P., M. and Johnson, C., J (2004) Final report to the Muskwa-Kechika trust fund: uncertainty and sensitivity in habitat suitability models, ecosystem and science management program, University of Northern British Columbia
- Guber A,K, Shelton D,R, Pachepsky Y,A (2005) Effect of manure on *Escherichia coli* attachment to soil. *J Environ Qual* 34(6):2086–2090
- Hirzel A,H, Lay G,L (2008) Habitat suitability modelling and niche theory. *J Appl Ecol* 45:1372–1381
- Hongoh V, Hoen G,A, Aenishaenslin C, Waaub J, Bélanger D, Michel P (2011) Spatially explicit multi-criteria decision analysis for managing vector-borne diseases. *Int J Health Geogr* 10(70):1–9
- Jiang X, Morgan J, Doyle M (2002) Fate of *Escherichia coli* O157:H7 in manure-amended soil. *Appl Environ Microbiol* 68(5):2605–2609
- Kienberger S, Lang S, Zeil P (2009) Spatial vulnerability units—expert-based spatial modelling of socio-economic vulnerability in the Salzach catchment, Austria. *Nat Hazards Earth Syst Sci* 9:767–778
- Lang S, Albrecht F, Blaschke T (2006) OBIA-tutorial: introduction to object-based image analysis, centre for geoinformatics. Salzburg University, Salzburg
- Lehnert C (1960) Die Tenazität von *Listeria monocytogenes* in der Aussenwelt. *Zentralbl. Bakteriol. Parasitenk. Abt1 Originale* 180: 350–356
- Ligmann-Zielinska A, Jankowski P (2014) Spatially-explicit integrated uncertainty and sensitivity analysis of criteria weights in multicriteria land suitability evaluation. *Environ Model Softw* 57: 235–247
- Ligmann-Zielinska, A., Jankowski, P., Watkins, J. (2012). Spatial uncertainty and sensitivity analysis for multiple criteria land suitability evaluation. Available at: < [http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CBwQFjAA&url=http%3A%2F%2Fwww.giscience.org%2Fpast%2F2012%2Fproceedings%2Fabstracts%2Fgiscience2012\\_paper\\_134.pdf&ei=GZPPUmcB8So0QXnu4GACQ&usq=AFQjCNGpWiNfw5OB6zUSKFeEmpg3HcceQ&sig2=82bx9kSKEqb0LaY1kR2hIg&bvm=bv.71667212,d.bGQ](http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CBwQFjAA&url=http%3A%2F%2Fwww.giscience.org%2Fpast%2F2012%2Fproceedings%2Fabstracts%2Fgiscience2012_paper_134.pdf&ei=GZPPUmcB8So0QXnu4GACQ&usq=AFQjCNGpWiNfw5OB6zUSKFeEmpg3HcceQ&sig2=82bx9kSKEqb0LaY1kR2hIg&bvm=bv.71667212,d.bGQ) > [Accessed 23 July 2014]
- Lilburne L, Tarantola S (2009) Sensitivity analysis of spatial models. *Int J Geogr Inf Sci* 23:151–168
- Lloyd-Smith Lab (n.d.) Our research. Available at: <https://www.eeb.ucla.edu/Faculty/lloydsmith/research/>
- Malczewski J (1999) GIS and multicriteria decision analysis. John Wiley and Sons, Inc., New York
- Mazimwe, A. (2013) Spatial vulnerability assessment to malaria in Uganda: The socio-economic dimension, Master's thesis, University of Salzburg, Salzburg
- Naeem S (2009) Ecology: Gini in the bottle. *Nature* 458:579–580
- Pushkareva VI, Ermolaeva SA (2010) *Listeria monocytogenes* virulence factor Listeriolysin O favors bacterial growth in co-culture with the ciliate *Tetrahymena pyriformis*, causes protozoan encystment and promotes bacterial survival inside cysts. *Biomed Central Microbiology* 10(26)
- Regan H,M, Colyvan M, Burgman M,A (2002) A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecol Appl* 12:618–628
- Sachs J,D et al (2009) Biodiversity conservation and the millennium development goals. *Science* 325:1502–1503
- Saltelli A, Annoni P (2010) How to avoid a perfunctory sensitivity analysis. *Environ Model Softw* 25:1508–1517
- Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) Global sensitivity analysis: the primer. Wiley-Interscience, Chichester, England
- Sjogren R,E (1994) Prolonged survival of an environmental *Escherichia coli* in laboratory soil microsomes. *Water Air Soil Pollut* 75(3–4): 389–403
- Spackman R, Entry J,A, Sojka R,E, Ellsworth J,W (2003) Polyacrylamide for coliform bacteria removal from agricultural wastewater. *J Soil Water Conserv* 58(5):276–283
- Store R, Kangas J (2001) Integrating spatial multi-criteria evaluation and expert knowledge for GIS-based habitat suitability modelling. *Landscape Urban Plan* 55:79–93
- Unc A, Goss M,J (2006) Culturable *Escherichia coli* in soil mixed with two types of manure. *Soil Sci Soc Am J* 70(3):763–769
- Va'zquez-Boland J,A,M, Kuhn P, Berche T, Chakraborty G, Domínguez-Bernal W, Goebel B, Gonzalez-Zorn J, Wehland, Kreft J (2001) *Listeria* pathogenesis and molecular virulence determinants. *Clinical Microbiological Review* 14:584–640
- Vinten A,J,A, Douglas J,T, Lewis D,R, Aitken M,N, Fenlon D,R (2004) Relative risk of surface water pollution by *E.coli* derived from faeces of grazing animals compared to slurry application. *Soil Use Manag* 20(1):13–22
- Welshimer H,J (1960) Survival of *Listeria monocytogenes* in soil. *J Bacteriol* 80:316–320
- WHO (2015) Neglected Zoonotic Diseases. Available at: [http://www.who.int/neglected\\_diseases/diseases/zoonoses/en/](http://www.who.int/neglected_diseases/diseases/zoonoses/en/).
- Wood N,L,J, Leach M, Waldman L, MacGregor H et al (2012) A framework for the study of zoonotic disease emergence and its drivers: spillover of bat pathogens as a case study. *Philos Trans R Soc* 367: 2881–2892
- Woolhouse M,E, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. *Emerg Infect Dis* 11:1842–1847