

**The use of spatiotemporal analytical tools to inform  
decisions and policy in One Health scenarios**

A dissertation submitted to the faculty of the University of Minnesota

By

**Kaushi Supun Tharindi Kanankege**

In partial fulfilment of the requirements for the degree of

Doctor of Philosophy

**Andres M. Perez (Advisor)**

**Nicholas B.D. Phelps (co-Advisor)**

February 2019

© Copyright by Kaushi ST Kanankege 2019  
All Rights Reserved

## Acknowledgements

It is my pleasure to express my sincere gratitude to the College of Veterinary Medicine of the University of Minnesota and those who guided, encouraged and supported me along the way in my PhD journey. First and foremost, I would like to thank my advisor, Prof. Andres M. Perez who is a tremendous mentor for believing in me and granting me an opportunity to become his student. Working with Prof. Perez has been a rewarding experience and his guidance and support throughout my graduate program have contributed greatly to my research and professional development. I would extend my sincere gratitude to Dr. Nick Phelps, who is my co-advisor, for being a continuous support throughout the journey and encouraging me to explore the concepts of decision-making and policy related to the thesis work. I would also like to extend my sincere acknowledgement to my advisor in minor in Epidemiology; Prof. Jeff Bender, for providing his expert advice towards the completion of the degree. Moreover, I am immensely grateful to my committee member, Dr. Julio Alvarez for his guidance and support in critical thinking and encouraging me to see the research questions with diverse perspectives. I am also grateful to Dr. Lin Zhang, for her expert guidance and support in improving my research work. My special thanks go to late Prof. Bob Morrison who was the chair of my thesis committee before he left us all due to an unfortunate accident. I am grateful for the opportunity to know him, learn from, and work with.

Many thanks to the Minnesota Discovery, Research, and Innovation Economy (MnDRIVE) program of the Office of the Vice President for Research (OVPR) of the University of Minnesota for funding this research. Furthermore, I wish to express my sincere appreciation of Minnesota Department of Natural Resources, Minnesota Board of Animal Health, and Minnesota dairy Herd Improvement Association for providing us with data and their expert opinion.

I wish to express a debt of gratitude owed to the guidance of Dr. Thomas Molitor, Dr. Randall Singer, Dr. Scott Wells, Dr. Kaylee Errecaborde, and Dr. Mark Rutherford for encouraging me to improve my skills in both research and teaching. I would like to extend my gratitude to Lisa Hubinger and Kate Barry, coordinators of the Veterinary Medicine Program (VMED) of the College of Veterinary Medicine.

I would tender my cordial acknowledge to the support of my colleagues and fellow team members: Dr. Amy Kinsley, Dr. Catalina Picasso, Dr. Kimberley VanderWaal, Dr. Moh

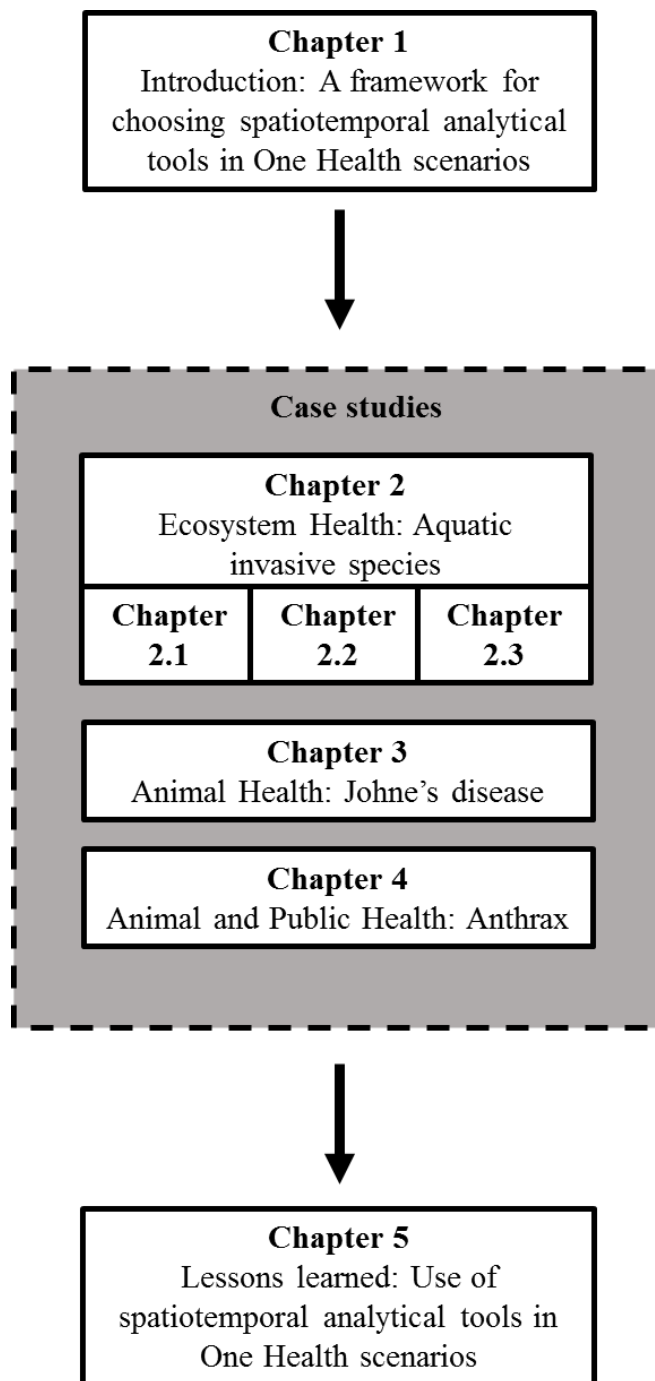
Alkhamis, Dr. Gustavo Machado, Dr. MariaSol Perez, Dr. Shivdeep Hayer, and the members of former Science, Technology, Engineering, and Mathematics for Minnesota Advancement (STEMMA) group. I am also thankful to my VMED colleagues Dr. Irene Bueno and Haejin Hwang for their kind support. I would also extend sincere gratitude to my mentors at the University of Wisconsin-Madison and University of Peradeniya, Sri Lanka for paving the path towards accomplishing my research goals in veterinary epidemiology.

Most importantly, I express my heartfelt gratitude to my wonderful husband, Maciek Kazula for his endearing love and support in sickness and in health and being my greatest strength. I shall not forget to extend my warmest thanks to my parents, Rani Wickramaratne and Kusumasiri Kanankege, for their unwavering support. I am fully in debt to my parents for shaping me to be the person who I am today. I am also grateful to my brother, my in-laws, and my friends and family for providing me with encouragement and support.

**Dissertation abstract****The use of spatiotemporal analytical tools to inform decisions and policy in  
One Health scenarios**

Kaushi ST Kanankege

The use of spatiotemporal analytical tools to generate risk maps and risk scores that facilitate early detection of health and environmental threats is increasingly popular in many countries and international organizations around the world. The traditional approach of spatial epidemiology focuses on mapping and conducting tests for detection of spatial aggregation of cases, referred to as “clusters”, to determine visual and geographical relational clues, and then ecologic approaches to recognize etiologic signs of disease distribution in relation to explanatory factors. The advances in spatial epidemiology are focused on the application of spatiotemporal findings to inform mitigation measures, use of big data to improve the validity and reliability of case-data based analyses, and eventually to provide risk estimates in a timely manner to support decision and policy in preventive and control measures, while supporting the improvement of existing data collection processes. This study provided a framework for choosing spatiotemporal analytical tools, summarizing the features of tools commonly used in spatial analysis, and discussing their potential use when informing decisions related to One Health scenarios. To this end, three case studies addressing endemic conditions affecting ecosystem health, animal health, and public health in Minnesota were compared. A risk score; an estimate/characterization of the disease spread, and suggestions on risk zones were introduced, using spatiotemporal analytical tools, addressing aquatic invasive species in Minnesota waters, Johne’s disease in dairy cattle, and Anthrax, affecting wildlife, livestock, and humans, respectively. The One Health concept promotes a collaborative approach, through effective communication and cooperation across disciplines and sectors, to solve complex problems that intersect animal, human and environmental health. An essential component in the process is understanding the stakeholder perspectives of the problem. Therefore, the comparison between the case studies focused on the lessons learned through the researcher-stakeholder interactions and identification of the opportunities and challenges in the process. Overall, the work presented through this dissertation, serves as precedent for establishing a protocol of “good practices” when promoting the use of spatiotemporal analytical tools to inform the implementation of scientifically driven risk management and policy solutions to One Health scenarios.

**Organization of the chapters in this thesis**

## Table of Contents

Acknowledgements .....	i
Dissertation abstract.....	iii
Organization of the chapters.....	iv
Table of contents .....	v
List of tables .....	vi
List of figures.....	x
List of abbreviations .....	xiii
<b>Chapter 1 (Introduction). A framework for choosing spatiotemporal analytical tools in One Health scenarios .....</b>	<b>1</b>
1.1 Overview .....	2
1.2 Introduction .....	3
1.3 Data and Methods .....	4
1.3.1 A framework for choosing spatiotemporal epidemiological tools .....	4
1.3.2 Commonly used spatiotemporal analytical tools (SAT) .....	8
1.3.3 Evaluating the performance of spatiotemporal analytical tools.....	13
1.3.4 Available software tools facilitating SAT .....	14
1.3.5 Other spatiotemporal analytical tools.....	14
1.3.6 Factors influencing the choice of spatiotemporal analytical tools.....	16
1.4 Discussion and Conclusions.....	22
1.5 Acknowledgements.....	23
1.6 References .....	23
<b>Chapter 2 (Ecosystem Health). The use of spatial cluster analysis, co-kriging, and maximum entropy ecological niche modeling as tools to inform decisions in aquatic invasive species management in Minnesota .....</b>	<b>38</b>
<b>Chapter 2.1. Recognizing Zebra mussels and Eurasian watermilfoil reporting patterns in Minnesota using spatial cluster analysis .....</b>	<b>39</b>
2.1 Overview .....	40
2.2 Introduction .....	41
2.3 Data and Methods .....	43
2.4 Results.....	50
2.5 Discussion.....	56
2.6 Conclusions .....	59
2.7 Acknowledgements.....	60
2.8 References .....	60
<b>Chapter 2.2. A probability co-kriging model to account for reporting bias and recognize areas at high risk for Zebra Mussels and Eurasian Watermilfoil invasions in Minnesota.....</b>	<b>64</b>
2.9 Overview .....	65
2.10 Introduction.....	66
2.11 Data and Methods .....	68

2.12 Results.....	74
2.13 Discussion .....	79
2.14 Conclusions.....	82
2.15 Acknowledgements .....	83
2.16 References.....	83
2.17 Supplement Tables and Figures .....	89
<b>Chapter 2.3. Developing a risk scoring system to estimate to recognize areas at high risk for Zebra Mussels and Eurasian Watermilfoil invasions in Minnesota: Combined interpretation of Maxent and co-kriging outputs.....</b>	
2.18 Overview .....	93
2.19 Introduction.....	94
2.20 Data and Methods .....	95
2.21 Results .....	101
2.22 Discussion .....	108
2.23 Conclusions.....	110
2.24 Acknowledgements .....	111
2.25 References.....	111
2.26 Supplement Tables and Figures.....	117
<b>Chapter 3 (Animal Health). The use of Getis Ord and Conditional autoregression as tools to study the spatial epidemiology of Johne’s disease in Minnesota: A cross-sectional study based on a voluntary testing program.....</b>	
3.1 Overview .....	119
3.2 Introduction .....	120
3.3 Data and Methods .....	121
3.4 Results .....	126
3.5 Discussion.....	132
3.6 Conclusions .....	134
3.7 Acknowledgements.....	134
3.7 References .....	134
3.8 Supplement tables and figures .....	142
<b>Chapter 4 (Animal and Public Health). The use of spatiotemporal cluster analysis as a tool to recognize historic patterns of Anthrax in endemic areas: A comparison between Minnesota and Kazakhstan.....</b>	
4.1 Overview .....	146
4.2 Introduction .....	147
4.3 Data and Methods .....	148
4.4 Results .....	151
4.5 Discussion.....	160
4.6 Conclusions .....	164
4.7 Acknowledgements.....	164
4.8 References .....	165



<b>4.9 Supplement tables and figures .....</b>	<b>170</b>
<b>Chapter 5 (Lessons learned). Risk based management of One Health scenarios using spatiotemporal analytical tools: Lessons learned.....</b>	<b>174</b>
<b>5.1 Overview .....</b>	<b>175</b>
<b>5.2 Introduction .....</b>	<b>176</b>
<b>5.3 Data and Methods .....</b>	<b>181</b>
<b>5.3.1 Description of the three examples/case studies .....</b>	<b>182</b>
<b>5.3.2 How, when, and by whom the risk maps may be used in One Health scenarios: Discussions with stakeholders related to the three case studies .....</b>	<b>183</b>
<b>5.3.3 Ecosystem health: Aquatic invasive species in Minnesota .....</b>	<b>183</b>
<b>5.3.4 Animal health: Paratuberculosis in cattle .....</b>	<b>185</b>
<b>5.3.5 Animal and public health: Anthrax in an endemic areas .....</b>	<b>188</b>
<b>5.4 Discussion and conclusions .....</b>	<b>192</b>
<b>5.5 Acknowledgements.....</b>	<b>195</b>
<b>5.6 References .....</b>	<b>195</b>
<b>Comprehensive reference list .....</b>	<b>201</b>

## List of Tables

Table 1.1 A summary of types of common spatial analytical tools and the purpose.....	6
Table 2.1. Spatial clusters resulted from the multivariate multinomial model of scan static test for zebra mussels (ZM) and Eurasian watermilfoil (EWM) invasions, adjusted by the human population density ( $p < 0.05$ ). Observed to expected ratio (O/E#). Categories of population density (Pop. Den. Category. ‡) include: (1) <10; (2) between 10 to 100; and (3) >100 people per square kilometre (Km)*.....	51
Table 2.2. Summary of the directionality tests for zebra mussels (ZM) and Eurasian watermilfoil (EWM). .....	54
Table 2.3. Number of waterbodies with the characteristic of each variable by 2010 and 2015.. ..	69
Table 2.4. Pearson correlation coefficient for the six waterbody specific variables used in the study.....	75
Table 2.5. Summary of co-kriging model validations for the probability of zebra mussel and Eurasian watermilfoil introductions in Minnesota. Cross validation was done using the k fold test ( $k=5$ ). True validation was done by fitting models for invasions as of 2010 and validating using the invasions reported between 2011 and 2015. Area under the receiver operating characteristic curve (AUC), Sensitivity and Specificity at the threshold risk are summarized.....	76
Supplement Table S2.1. Parameters used in the co-kriging models for Eurasian watermilfoil (EWM) and zebra mussels (ZM).....	91
Table 2.6. Predictors used in the final Maxent models for zebra mussels (ZM#) and Eurasian watermilfoil (EWM‡). .....	102
Table 2.7. Summary of model validations for both Maxent, cokriging and combined interpretation of the models using risk matrix. Cross validation was done using the K fold test. True validation was done by fitting models for invasions as of 2010 and validating using the invasions reported after 2010. Area under the receiver operating characteristic curve (AUC), Sensitivity (Se) and Specificity (Sp) at the risk rank $\geq 3$ for Maxent and cokriging, and combined interpretation using the risk matrix are listed below. Combined interpretation was only subjected to true validation.. .....	104
Table 2.8. Risk matrices for (1) zebra mussels (ZM)* and (2) Eurasian Watermilfoil (EWM)# invasion risk using the Maxent and co-kriging model outputs. Each model output was categorized into 10 classes where 1 represents the lowest risk, and 10 represent the highest risk. The number of waterbodies which has each of the risk class combination was illustrated in each cell of the matrices. Percentage** are the marginal percentages i.e. the number of waterbodies in each category compared to the total number of waterbodies ( $n = 18,411$ ). Combined risk of invasion is represented by the color scheme, where red and dark-green represent very high-risk and negligible risk of invasion, respectively.....	107

Supplement Table S2.2. Environmental variables used in the final MaxENT models predictors used in the final Maxent models for zebra mussels (ZM#) and Eurasian watermilfoil (EWM <sup>†</sup> ). .....	117
Table 3.1. Herd demographic factors and environmental factors associated with JD in North American dairy cattle, according to the published literature.....	122
Table 3.2. Minimum sample sizes to be collected required to estimate freedom from Johne's Disease (JD) at the herd level, using an imperfect test and adjusting for a finite population calculated using the AusVet EpiTool Epidemiological calculator (URL: <a href="http://epitools.ausvet.com.au">http://epitools.ausvet.com.au</a> ).. .....	130
Table 3.3. Odds ratios, coefficients, and p-values of the association between epidemiological factors and herd-level Johne's Disease status, based on ELISA assays performed on individual milk samples in 600 herds in Minnesota.....	131
Supplement Table S3.1. Environmental factors assessed for association with Johne's disease status in Minnesota dairy herds participating in the voluntary testing program conducted by Minnesota Dairy Herd Improvement Association. ....	143
Supplement Table S3.2. Akaike information criterion (AIC) values for the model fits.....	144
Table 4.1. Summary of spatiotemporal clusters of Anthrax detected in Minnesota (between 1912 and 2014) and Kazakhstan (between 1933 and 2014), using the spatiotemporal permutation model of the scan statistics. The spatial and temporal windows of the scan statistic were set to 10% and 5% respectively.....	158
Table 4.2. Summary of the similarities and differences of spatiotemporal patterns in Minnesota and Kazakhstan. ....	159
Supplement Table S4.1 Summary of the historic events in Minnesota related to the patterns how the trigger events i.e., climatic, anthropogenic, agricultural, and environmental changes, which may have led to the recognized spatiotemporal clusters of Anthrax.....	180
Supplement Table S.4.2 Summary of the historic events in Kazakhstan related to the patterns how the trigger events i.e., climatic, anthropogenic, agricultural, and environmental changes, which may have led to the recognized spatiotemporal clusters of Anthrax.....	181
Table 5.1. Comparison between the three case studies.....	182
Table 5.2. Summary of the assessment of researcher-stakeholder interactions related to Ecosystem health (Aquatic invasive species, AIS), animal health (Johne's disease), and public health (Anthrax) case studies: Lessons learned.....	191

## List of Figures

- Fig. 1.1 Schematic illustration of a framework for choosing spatiotemporal analytical tools (SAT). The research questions/objectives are identified with Q1:Q4. The specific SATs under the relevant categories, i.e. T1:T4, are listed in Table 1.1. .... 5
- Fig. 1.2 Factors influencing the choice of spatiotemporal analytical tools (SAT)..... 16
- Fig. 2.1 The distribution of each invasive species at certain time intervals. Panels a, b, and c represents Zebra mussels (ZM) distribution between 1989 – 2015. Eurasian watermilfoil (EWM) distribution between 1987 - 2015 is illustrated in panels d, e, and f..... 44
- Fig. 2.2 Reported and confirmed new invasions and accumulated number of cases of zebra mussels (ZM) and Eurasian watermilfoil (EWM) in Minnesota waterbodies by year (Based on the Minnesota Department of Natural Resources database: MNDNR: <http://www.dnr.state.mn.us/invasives/ais/infested.html>. Cited June 22, 2015).. .... 45
- Fig. 2.3 Clusters of zebra mussels (ZM) and Eurasian watermilfoil (EWM) detected by the purely spatial multivariate multinomial cluster analysis, stratified by the human population density of Minnesota, based on 2007 census data (Bright et al., 2013). The numbered circles represent the significant clusters stratified by the categories of human population density per square kilometer..... 51
- Fig. 2.4 Directionality of the zebra mussel and Eurasian watermilfoil reporting resulted from the standard deviation ellipse analysis showed a northwestern or southeastern trend with 111o and 128o rotations ZM and EWM invasions, respectively. Std. Dev. 1 and 2 refer to the first and second standard deviations encompass approximately 68% and 95% of all input case locations. The major roads and the major river centerlines are indicated in the map to facilitate recognition of potential means of dispersal..... 55
- Fig. 2.5 Co-kriging model outputs illustrating the probability of introduction of zebra mussels (ZM) and Eurasian watermilfoil (EWM) to Minnesota waterbodies, for the invasions as of 2010. The risk classes 1 through 5 indicates the intensity of the probability of introduction, where, class 5 represent a high probability of ZM or EWM introduction. The number of waterbodies under each category and as a percentage of the total waterbodies (n=18,411) is listed..... 77
- Fig. 2.6. Co-kriging model outputs illustrating the probability of introduction of zebra mussels (ZM) and Eurasian watermilfoil (EWM) to Minnesota waterbodies, for the invasion status of 2015. The risk classes 1 through 5 indicates the intensity of the probability of introduction, where, class 5 represent a high probability of ZM or EWM introduction. The number of waterbodies under each category and as a percentage of the total waterbodies (n=18,411) is listed..... 78
- Supplement Fig. S 2.1 The boater traffic between waterbodies based on the Watercraft Inspection Program conducted by Minnesota Department of Natural Resources. The data from year 2013 are illustrated. The panel A of figure represents the movement of boaters from

previously visited waterbody-to-waterbody where the survey data were collected. The panel B represents the movement of boaters from waterbody where the survey data were collected-to-the waterbody where they plan to visit next. .... 98

Supplement Fig. S2.2 An illustration of the Eigen vector centrality for the waterbodies in the boater traffic network created using the surveys of Watercraft Inspection Program conducted by Minnesota Department of Natural Resources. The data from year 2013 are illustrated. . . 99

Fig. 2.7. MaxENT model outputs illustrating the habitat suitability of Minnesota for zebra mussels (ZM) and Eurasian watermilfoil (EWM). The predicted risk is for waterbodies only. The risk classes 1 through 10 indicates the intensity of the suitability of the locations, where, class 10 represent most suitable areas. The number of waterbodies under each category and as a percentage of the total waterbodies (n=18,411) is listed. The polygons represent counties of the state of Minnesota..... 105

Fig. 2.8. Co-kriging model outputs illustrating the risk of introducing zebra mussels (ZM) and Eurasian watermilfoil (EWM) to Minnesota waterbodies. The risk classes 1 through 10 indicates the intensity of the probability of introduction, where, class 10 represent a high probability of ZM or EWM introduction. The number of waterbodies under each category and as a percentage of the total waterbodies (n=18,411) is listed. The polygons represent counties of the state of Minnesota..... 106

Fig. 3.1. The participation of study herds as a percentage of the ideal sample size, by county, is summarised with the graduated symbols. The Minnesota Dairy Herd Improvement Association testing laboratories are illustrated with triangles. The background indicates the dairy cattle farm density by county, based on 2012 census of the National Agricultural Statistics Service (NASS, 2012). . . 127

Fig. 3.2. Spatial patterns of Johne's disease status and the associated covariates. Points represent the location of study herds (n=600). Panels a, b, and c illustrates the results of the Getis Ord  $G_i^*$  local test where herds with high value of the variable next to herds with high values of the variable are represented in red (high-high clusters), herds with low value of the variable next to herds with low values of the variable represented in blue (low-low clusters), and non-matching pairs in yellow (high-low, or low-high values) (Getis and Ord, 1992; Ord and Getis, 1995). Panels e, f, g, and h simply depict the distribution of covariates... 129

Fig. 4.1. Epidemic curves illustrating the number of animal Anthrax outbreaks. Outbreaks were reported between 1912 and 2014 in Minnesota and between 1933 and 2014 in Kazakhstan (Bars indicate annual number of reported cases, left axis, and lines indicate cumulative number of cases, right axis)... 152

Fig. 4.2. Number of animal Anthrax cases by month in Minnesota (1912-2014) and Kazakhstan (1933-2014). The number of cases included in the monthly analysis (n) were 289 and 2,027 for Minnesota and Kazakhstan (53.6% of the total records of Kazakhstan) respectively.... 153

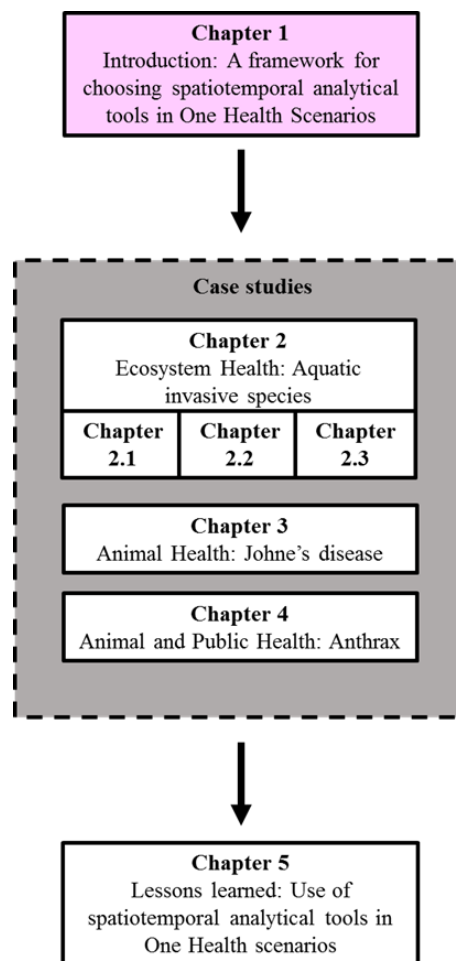
Fig. 4.3. Maps illustrating the distribution of animal Anthrax at certain time intervals. Panels a, b, and c represents cases in Minnesota between 1912 and 2014. Panels d, e, and, f represents cases in Kazakhstan between 1933 and 2014. According to the spatiotemporal directionality

test performed using the relative time matrix, a northeastern directionality ( $42^\circ$ ) was found in Minnesota and a southwestern directionality for Kazakhstan ( $260^\circ$ ) with low ( $\sim 0.1$ ) though significant ( $p < 0.05$ ) AC values in both locations..... 154

- Fig. 4.4. Spatiotemporal clustering of animal Anthrax cases in Minnesota (time: 1912-2014). Clusters were detected using the space-time permutation model of the spatial scan statistic with the spatial-window set to 10% and 5%-time window. Clusters are numbered according to the order of the start date of the outbreak... ..... 156
- Fig. 4.5. Spatiotemporal clustering of animal Anthrax cases in Kazakhstan (time: 1933-2014). Clusters were detected using the space-time permutation model of the spatial scan statistic with the spatial- and time-windows set to 10% and 5%, respectively. Clusters are numbered according to the order of the start date of the outbreak..... 157
- Supplement Fig S4.1 Sensitivity analysis of the spatiotemporal window sizes used in the space-time permutation model of scan statistics... ..... 172
- Supplement Fig S4.1 Frequency distribution of the spatiotemporal cluster radii of historic animal Anthrax in Minnesota, USA and the Republic of Kazakhstan. Spatiotemporal clusters were detected using the space-time permutation model of the spatial scan statistic with the spatial- and time-windows set to 10% and 5%, respectively..... 173
- Fig. 5.1 The use of spatiotemporal eco-epidemiological tools in the process of risk based management of One Health scenarios. The steps labelled 1: 5 are discussed in this paper using three case studies representing One Health scenarios.. ..... 180

**List of abbreviations**

AIS	Aquatic invasive species
AUC	Area under the Receiver operating characteristics curve
cAUC	Calibrated AUC related to spatial sorting bias in Maxent
ENM	Ecological Niche Modeling
EWM	Eurasian watermilfoil
JD	Johne's disease
MNBAH	Minnesota Board of Animal Health
GIS	Geographic Informatics Systems
KDE	Kernel density estimation
Maxent	Maximum entropy ecological niche modeling
MNDHIA	Minnesota Dairy Herd Improvement Association
MNDNR	Minnesota Department of Natural Resources
O/E	Observed-to-expected ratio of scan statistics
SAT	Spatiotemporal analytical tools
ZM	Zebra mussels



## Chapter 1: Introduction

# *A framework for choosing spatiotemporal analytical tools in One Health scenarios*

This work has not been published/in-review elsewhere



## 1.1 Overview

Spatiotemporal analytical tools (SAT) are becoming increasingly popular in risk-based monitoring/management of adverse health events affecting humans, animals, and ecosystems across the geographical space. Different disciplines use different SAT to address similar research questions. Bringing these different techniques together under the same umbrella may support an interdisciplinary approach of shared knowledge. Here, we are introducing a framework for choosing SAT when addressing common research questions across disciplines focused on human, animal, and ecosystem health. This framework is comprised of three stages: a) pre-hypothesis testing where hypotheses regarding the spatial dependence of events are generated; b) primary hypothesis testing stage where spatial dependence and patterns are tested; and c) secondary-hypothesis testing and spatial modeling stage, where the predictions and inferences using the spatial dependence and the associated covariates are made. First, research questions were simplified into five main questions. Answers to the questions lead to the choice of SATs from four different categories including: T1) Visualization, descriptive analysis and test for spatial dependence; T2) Spatial/Spatiotemporal pattern recognition; T3) Spatial smoothing and interpolation; and T4) Spatial correlation models. Furthermore, we propose seven factors/criteria, which can help guiding the choice of SAT. The seven factors include epidemiological characteristics of the disease, the study design, the level of spatial explicitness of data, data quality and availability, the research question and hypothesis, stakeholder involvement in research, and existence of regulations, resources, and policy related to mitigation of the disease or the adverse health effect. Through this framework we intend to facilitate the choice between available SATs, and promote their cross-disciplinary use to support improving human, animal, and ecosystem health, which is essentially a One Health approach.

## 1.2 Introduction

### 1.2.1 Spatiotemporal analytical tools (SAT)

Spatial epidemiology is defined as “the description and analysis of geographic variations in disease with respect to demographic, environmental, behavioral, socioeconomic, and infectious risk factors” (Elliott and Wartenberg 2004). The integration of epidemiological concepts, spatial analysis, geographic information science (GIS), and statistics can lead to the accomplishment of the objectives of spatial epidemiology in understanding and modeling spatiotemporally explicit health risks (Ward and Carpenter 2000; Beale et al. 2008). Traditionally, spatial epidemiology focused on two major concepts: a) mapping and spatial pattern analysis, such as cluster analysis, to determine visual and geographical relational cues, and b) using ecologic approaches to recognize etiologic clues of disease spread and explanatory factors (Lawson 2006). Combined approaches of health geography, i.e. application of geographical analysis to health issues (Kirby et al., 2017), as well as the concepts of eco-epidemiology (Susser, 2004; March and Susser, 2006) provide a set of powerful tools to improve spatial epidemiology. The term Eco-epidemiology refers to engagement of a richer and a broader framework into the risk-factor epidemiology while accounting for multiple levels of causation, the interplay between genetic and environmental factors, and temporal characteristics (Susser, 2004; March and Susser, 2006). These advances in spatial epidemiology are intended to understand the disease distribution and underlying risk factors, improve the quality of epidemiological data, and eventually provide risk estimates in a timely manner to support decision and policy in preventive and control measures (Cromley and McLafferty, 2002; Wagner Robb et al., 2016; Pfeiffer and Stevens, 2015).

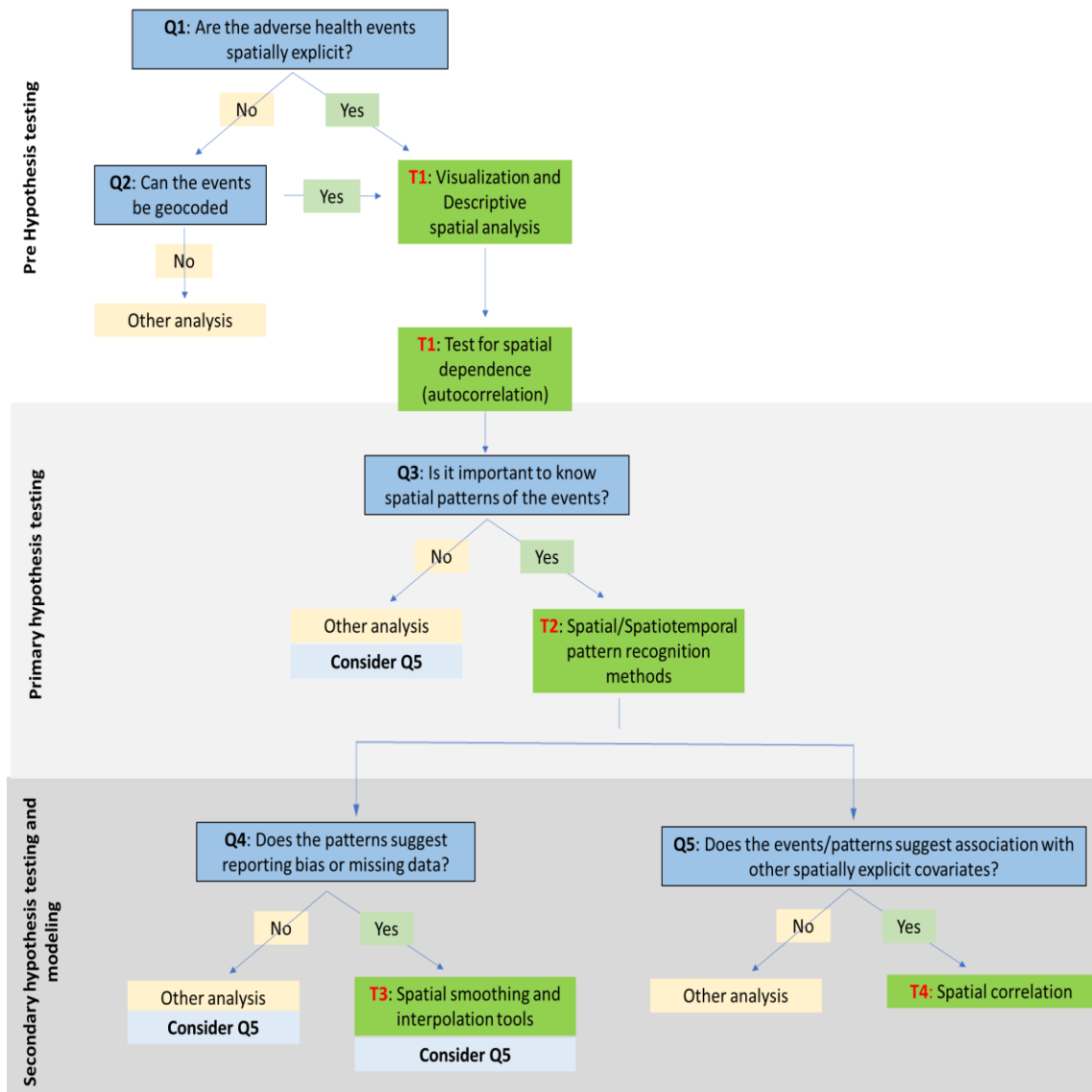
## 1.2.2 Objective

A plethora of SATs, especially geostatistical tools, have been published and used in the field of spatial epidemiology (Cromley and McLafferty, 2002; Banerjee et al., 2014). However, selecting an appropriate analytical tool is often a challenging decision. This is because the choice of SAT depends on variety of factors/criteria. The majority of the published reviews and books on SAT are focused on describing the characteristics of the methods and do not consider these factors. Thus, understanding these criteria and developing a framework that enables the researchers to analyze existing epidemiological data, draw inferences, and plan future research in spatiotemporal epidemiology is an essential step. The objective of this work is to introduce a framework that facilitate choosing SAT and discuss the choice of SAT in relation to seven major factors.

## 1.3 Data and Methods

### 1.3.1 A framework for choosing spatiotemporal epidemiological tools

Here, we are introducing a framework for choosing SAT (Fig. 1.1). The framework is classified into three stages: a) pre-hypothesis testing/ hypothesis generating stage; b) primary hypothesis testing stage; and c) secondary-hypothesis testing and spatial modeling stage where the predictions and inferences are made. Here, the primary hypothesis refers to the existence of spatial dependence and spatial patterns in the distribution of adverse health events, while the secondary hypotheses involve the association of the events with risk factors/covariates. The different types of SAT are broadly classified into four categories: T1) Visualization, descriptive analysis and test for spatial dependence; T2) Spatial/Spatiotemporal pattern recognition; T3) Spatial smoothing and interpolation; and T4) Spatial regression. The framework seeks to suggest a suitable category of the SAT among the four, based on the stage of the research/stakeholder question. The types of SAT that are commonly used in epidemiological studies are listed under each category (T1:T4) in Table 1.1 and discussed briefly in the consecutive text. The usage of tools are further discussed in relation to the seven major factors/criteria that influence the choice of SAT. It is important to note, however that this is not a detailed review on the existing SAT. Reviews on SAT can be found elsewhere (Ward and Carpenter 2000; Beale et al. 2008; Carroll et al., 2014; Kirby et al., 2017), as well as, a glossary of commonly used terms and their definitions in spatial epidemiology is found in Rezaeian *et al.* (2007).



**Fig. 1.1.** Schematic illustration of a framework for choosing spatiotemporal analytical tools (SAT). The research questions/objectives are identified with Q1:Q5. The specific SATs under the relevant categories, i.e. T1:T4, are listed in Table 1.1.

**Table 1.1.** A summary of types of common spatial analytical tools and their purpose.

	Purpose	Measure	Commonly used techniques	Reference
<b>T1: Visualization, descriptive analysis and test for spatial dependence</b>	Transformation of locational information into geographic coordinates	Geocoding	By street address, postal code, or administrative divisions	Zandbergen, 2008; Hart and Zandbergen, 2013
	Visualization and description of the spatial distribution	Exploratory spatial data analysis	Mean center	Isaaks and Srivastava, 1989
			Median center	Isaaks and Srivastava, 1989
			Standard deviation (weighted by attributes)	Isaaks and Srivastava, 1989
			Directional mean and variance	Levefer, 1926
			Moran scatter plot	Anselin, 1996
	Test whether there is spatial dependence in the event data	Spatial autocorrelation	Mantel test	Mantel, 1967
			Global Moran's I	Moran, 1950
			Geary's C	Geary, 1954
			Geti's ord	Getis and Ord, 1992; Ord and Getis, 1995
		Distance analysis	Nearest neighbor analysis	Clark and Evans, 1954
			Ripley's K	Ripley, 1977; Dixon, 2002
	Measurement of uneven distribution of the populations and risk factors	Local or stratified spatial heterogeneity	Getis Ord Gi*	Getis and Ord, 1992; Ord and Getis, 1995
			K-Means clustering	MacQueen, 1967
			Anselin's local Moran (L-Moran)	Anselin, 1995
Spatial stratified heterogeneity test			Wang et al., 2016	
Measure the spatial dependence while accounting for background population		Oden's Ipop	Oden, 1995; User manual, 2012: <a href="https://www.biomedware.com">https://www.biomedware.com</a>	
<b>T2: Spatiotemporal pattern recognition</b>	Test spatial trends	Testing for first-order effects	Trend analysis	Lawson, 2006
	Test whether there is any spatial clustering in the data	Global cluster detection	Nearest neighbor test	Clark and Evans, 1954
			Cuzick and Edward's test (case-control data)	Cuzick and Edwards, 1990
			Local Indicators of spatial association (LISA)	Anselin, 1995
	Locate the clusters and the statistical	Purely spatial local cluster detection	Spatial scan statistics	Kulldorff and Nagarwalla, 1995; Kulldorff, 2009
Turnbull's test			Turnbull et al., 1990	

	significance of the clustering		Besag and Newell's test	Besag and Newell, 1991	
	Test whether there is space and time clustering in the data	Spatiotemporal cluster detection	Knox test	Knox, 1964	
			Mantel test	Mantel, 1967	
			Space-time permutation scan statistic	Kulldorff et al., 2005	
			Eder-Myer-Mantel test	Stark and Mantel, 1967	
	Detecting the direction of progression of an adverse event over time	Spatiotemporal directionality	Spatiotemporal directionality test	Jacquez et al., 2002; User manual, 2012: <a href="https://www.biomedware.com">https://www.biomedware.com</a>	
			Spatiotemporal anisotropy parameter	Bilonick, 1988; Snepvangers et al. 2003	
	<b>T3: Spatial smoothing and interpolation</b>	Quantifying spatial variations in event intensity: spatial point pattern (SPP) intensity	Density based point pattern recognition	Univariate Kernel density estimation (KDE)	Bithell, 1990; Kelsall and Diggle 1995, Hazelton, 2016
				Multidimensional KDE	Silverman, 1986; Wand and Jones, 1995
				Empirical Bayes smoothing (EBS)	Clayton and Kaldor, 1987
Characterization of the clustering/spatial patterns		Spatial interpolation and spatial regression	Neighborhood matrices		
			Headbang smoothing	Kafidar, 1996; Mungiole et al., 1999; Gelman et al., 2000	
			Triangulated Irregular Network (TIN)	Sanson et al., 1991	
			Inverse Distance Estimation (IDW)	Keckler, 1995; Tomczak, 1998	
			Kriging	Matheron, 1971; Isaaks and Srivastava, 1989	
			Spline regression models	MacNab and Gustafson, 2007	
			Trend Surface Interpolation	Lusting, 1969; Davis, 1973; Agterberg, 1984; Mitas and Mitasova 1999	
			Multivariate spatial interpolation (Predict the probability of disease using correlated and highly sampled auxiliary variables)	Co-kriging	Isaaks and Srivastava, 1989
				Regression kriging	Odeh et al., 1995; Goovaerts, 1997; Eldeiry and Garcia, 2010
			Spatiotemporal interpolation	Spatiotemporal interpolation	Space-time kriging
Two-stage Bayesian approach		Lawson et al., 2012			
Autoregressive spatial smoothing and temporal Spline smoothing		MacNab and Dean, 2004			

<b>T4: Spatial correlation</b>	Predict the probability of disease using explanatory variables (When regression residuals have spatial autocorrelation)	Spatial and spatiotemporal autoregression models	Simultaneous autoregressive (SAR) models	Cressie, 1993; Haining, 2003; Banerjee et al., 2014
			Geographically weighted regression (GWR)	Brunsdon et al., 1996; Nakaya et al., 2005
			Purely spatial: Conditional autoregressive (CAR) prior; Besag model; Besag, York-Mollie (BYM) model	Besag et al., 1991; Banerjee et al., 2014
			Spatiotemporal CAR models	Mariella and Tarantino, 2010; Rushworth et al., 2017;
	Predict probabilities when the risk surfaces are heterogeneous or discontinuous	Latent structure models		Cai et al., 2012; Hossain and Lawson, 2016
	Analysis of spatially explicit time-to-event data	Spatial survival models	Spatial cure rate model	Banerjee, 2016
			Frailty models	Banerjee, 2016
	Predict the probability of disease when the disease occurrence is correlated with environmental variables	Environmental / Ecological niche modeling	Maximum Entropy Ecological Niche modelling (Maxent)	Elith et al., 2011; Phillips et al. 2006; Merow et al., 2013
			Genetic Algorithm for Rule Set Production (GARP)	Stockwell and Peters, 1999; Stockwell and Peterson, 2002; Blackburn et al., 2007
			Machine learning techniques: Random forest	Breiman, 2001; Mi et al., 2016

### 1.3.2 Commonly used spatiotemporal analytical tools (SAT)

#### T1 Visualization, descriptive analysis and test for spatial dependence

Spatial data visualization is one of the key steps in understanding and generating hypothesis on the spatial distribution of events. Global Navigation Satellite Systems (GNSS), such as Global Positioning System (GPS); Global Navigation Satellite System (GLONASS); Galileo; Navigation Indian Constellation (NavIC); and BeiDou, provide the ability to position the exact geo-spatial locations during the data collection phase. In the absence of GNSS based data, geocoding plays a major role to generate spatially explicit databases (Zandbergen, 2007; Hart and Zandbergen, 2013).

According to Walter Tobler's First Law of Geography, "everything is related to everything else, but near things are more related than distant things (Tobler 1970)." This phenomenon, otherwise known as spatial autocorrelation or spatial dependence, is a key component of spatial epidemiology. The majority of the descriptive techniques are focused on determining the extent to which data are spatially autocorrelated and performing hypothesis tests after accounting for spatial autocorrelation (Dormann et al., 2007). Assumptions involved in descriptive analytics include the spatial stationarity, isotropic spatial autocorrelation, and spatial continuity (Dormann et al., 2007). In simpler terms these assumptions imply that events (infectious diseases in animals for example) of the considered spatial process are homogeneously distributed across homogeneously distributed across the region regardless of geographical directions or barriers. However, understanding the violations of these assumptions, i.e. detecting patterns of non-stationarity or anisotropy, is paired with the descriptive analytics (Isaaks and Srivastava 1989). Moran's I (Moran 1950), Geary's C (Geary, 1954), Mantel test (Mantel 1967), and Getis Ord (Getis and Ord 1992; Ord and Getis 1995), which often referred to as "global spatial autocorrelation indices", are the commonly used techniques in spatial analysis.

Measurement of spatial heterogeneity, i.e. uneven distribution of the populations and risk factors across the geographical space, is another important component in the One Health process. Spatial heterogeneity measures could be either 1) local where we measure whether an attribute at one site is different from its surrounding or 2) stratified where the attributes are stratified within strata, such as Agro-ecological zones or land use categories in which the spatial variance between strata was measured. An example of local measures of spatial heterogeneity is Getis Ord  $G_i^*$  (i.e. hot-pot/cold spot analysis) (Getis and Ord 1992; Ord and Getis 1995). Other techniques such as G-statistics are increasingly available facilitating the measurement of stratified spatial heterogeneity (Chen and Wang, 2016). The indices of spatial heterogeneity provide opportunity to quantitatively measure the differences and compare the landscape patterns of populations and risk factors.

## **T2 Spatial/Spatiotemporal pattern recognition**

A spatial cluster is an excess of events or measurements in geographic space, compared to the null expectation of spatial randomness (Jacquez, 2008). The cluster analysis generally is aimed at detecting if there is any clustering in the spatial data (i.e. Global cluster analysis), and detecting and locating the clusters (local cluster analysis and focused cluster analysis). In general,



the cluster analysis provides information about the cluster morphology, including the magnitude of the excess/deficit feature, geographic size, shape, and the locations of spatial clusters.

Detecting first-order adjacencies such as Local Indicators of Spatial Autocorrelation (LISA) statistics (Anselin 1995, Ord and Getis 1995) and nearest-neighbors relationships such as used in Cuzick and Edward's (1990) test can be considered as global cluster detection techniques. Most local cluster-detection techniques employ circular scanning windows such as the scan statistic (Kulldorff and Nagarwalla 1995), Turnbull's test (Turnbull et al. 1990) and Besag and Newell's (1991) test. Some of these local cluster analyses such as scan statistics have been incorporated into widely used software such as SaTScan that enable temporal, spatial, and spatiotemporal cluster analysis in a user-friendly manner. It is essential to realize that spatial variation and hence cluster morphology is complex, and may not be well described by the circular cluster window approaches (Jacques, 2008). Therefore, alternative approaches that are flexible for the cluster shape such as Upper Level Set Scan statistics (Patil and Taillie, 2004) and B-statistics (Jacques et al., 2006) have been introduced. A detailed description on the spatial pattern recognition and cluster analytical techniques are found elsewhere (Jacques 2008).

### **T3 Spatial smoothing and interpolation**

Many research studies apply spatial smoothing and interpolation techniques to improve estimation. Smoothing techniques facilitate visualization of the intensity of events (Bithell, 1990), allow accounting for background spatial distribution of the population at risk (Diggle et al., 2005), and generate tolerance contours (i.e. confidence regions) for which the relative risk of a disease is significantly high (Kelsall and Diggle 1995). Smoothing techniques are used to reduce noise by shrinking values towards the adjacent observations, which is applicable to both homogenous and heterogeneous point processes. In a heterogeneous point process where the intensity of the spatially varying event varies within the study area, the estimation of the event intensity using either parametric or non-parametric methods is called the smoothing (Bithell, 1990; Kelsall and Diggle 1995; Hazelton, 2016). Spatial smoothing techniques uses a moving weighted function to reduce the noise component, where the differences in the values on a surface are accentuated resulting in spatially continuous map. The commonly used spatial smoothing techniques include kernel density estimation (KDE) (Bithell, 1990; Kelsall and Diggle 1995; Rushton, 1997; Rushton et al., 2004) and headbanging (Kafidar 1996; Mungiole et al., 1999; Gelman et al. 2000), which are considered as alternatives of detecting circumscribing clusters of varying shapes in lieu of circular clusters (Kelsall and Diggle 1995; Jacques 2008).

Spatial interpolation techniques are used to estimate/predict values at the unknown locations using the available/known data points (Isaaks & Srivastava 1989). The resulting interpolated surfaces i.e. statistical surfaces are raster layers and often can be considered as risk maps in epidemiological analyses. There are multiple spatial interpolation techniques including Inverse distance estimation (IDW) (Keckler, 1995; Tomczak, 1998), Triangulated Irregular Network (TIN) (Sanson et al., 1991), Kriging as well as its variations such as Co-kriging (Isaaks and Srivastava 1989), and Trend Surface Interpolation (Lusting, 1969; Davis, 1973; Agterberg, 1984; Mitas and Mitasova 1999). TIN allows construction of 3-dimensional surfaces based on a secondary variable of a researcher's choice, which, for example, the prevalence of a disease in a farm location. Kriging can be understood as a two-step process, where, step 1 is fitting the spatial variogram or likelihood for the data observed at the sampled points; and step 2 involves the interpolation of values for unsampled points or blocks using the weights derived from this covariance structure (Isaaks and Srivastava 1989). In situations where disease events are biased or undersampled, co-kriging can be used to enhance the accuracy of the estimation using a highly sampled auxiliary variable. Trend surface interpolation facilitates mapping variables while allowing for the local fluctuations. Therefore, trend surface analysis may reflect the regional distribution, trend, and the local variation of the mapped disease (Huanxin and Nan, 1991). Interpolation techniques, their model assumptions, and usage are discussed extensively, elsewhere (Isaaks and Srivastava 1989; Mitas and Mitasova 1999).

Spatiotemporal interpolation techniques are used to predict variables in-between and beyond observation times (Cressie and Wickle, 2011; Biondi, 2013). In space-time kriging, the spatial, temporal, and spatio-temporal dependence structures are modeled using spatiotemporal variograms (Biondi, 2013). Modeling the spatial and temporal components independently is one of the drawbacks in most of the spatiotemporal interpolation techniques (Susanto et al., 2016). A detailed discussion on the spatiotemporal interpolation techniques used in the environmental modeling is found elsewhere (Susanto et al., 2016).

#### **T4 Spatial correlation models**

In geographic correlation studies in epidemiology, spatial regression analysis is commonly used to examine the effects of certain risk factors/covariates on disease incidence while accounting for the spatial autocorrelation/dependence in the residuals (Frome and Checkoway, 1985; Auchincloss et al., 2013; Banerjee, Carlin, Gelfand, 2014). This is because the standard regression models assume that observations are independent, an assumption that is not

met when spatially dependent data are analyzed. Both Frequentist and Bayesian spatial regression techniques have been extensively used in epidemiological analyses. Spatial regression models vary by their computational complexity, capacity of capturing spatial heterogeneity, and the quantification of uncertainty associated with parameter estimates ([Auchincloss et al., 2013](#)).

Spatial autoregressive models include: Simultaneous autoregressive (SAR) models ([Cressie, 1993](#); [Haining, 2003](#); [Banerjee et al., 2014](#)), Geographically weighted regression ([Fotheringham et al., 1998](#)), and Conditional autoregressive models (CAR) ([Besag et al., 1991](#)). Due to sampling and reporting variabilities of disease incidences and risk factors, borrowing strength from neighboring regions to get more reliable estimates is the motivation behind these spatially dependent regression models (e.g. closer neighbors might receive higher weights). This strategy of borrowing information from neighbors is applicable in autoregressive models, where the spatial or spatiotemporal structure is modeled via sets of autocorrelated random effects ([Besag et al., 1991](#); [Lawson 2013](#); [Banerjee et al., 2014](#)). In addition to accounting for the spatial dependency, multiple spatiotemporal regression models have been used in epidemiological studies that enable the researchers to analyze the influence of spatial and temporal dependence of disease events and risk factors ([Banerjee et al., 2014](#); [Lee et al., 2018](#)). Detailed descriptions on spatial and spatiotemporal autoregressive models can be found elsewhere ([Banerjee et al., 2014](#); [Lee et al., 2018](#)).

Latent structure models accounts for the heterogeneity or the discontinuity in risk surface such that homogenous areas can be grouped together while discriminating for the risk levels ([Hossain and Lawson, 2016](#)).

Ecological niche modeling (ENM) approaches are widely used to characterize the complexity and heterogeneity of the landscapes in research related to epidemiologically relevant vector and parasite-reservoir distributions ([Peterson, 2006](#); [Peterson et al., 2011](#); [Escobar and Craft, 2016](#)). In addition to the characterization of the areas where disease is distributed, ENM is used to identify potential distributional areas in response to the likely geographic shifts in distributional areas of species or phenomena under scenarios of climate change or changing land use ([Peterson et al., 2005](#)). Genetic Algorithm for Rule Set Production (GARP) ([Stockwell and Peterson, 2002](#); [Blackburn et al., 2007](#)); Maximum Entropy Ecological Niche modelling (Maxent) ([Phillips et al. 2004](#), [Elith et al. 2011](#)); and Machine Learning Techniques such as random forest ([Breiman, 2001](#); [Mi et al., 2016](#)) are the commonly used ENM algorithms in epidemiology. Most studies use presence-only data for the analyses. Further details regarding GARP, Maxent, and

other ENM algorithms are found elsewhere ([Stockwell and Peters, 1999](#); [Stockwell and Peterson, 2002](#); [Phillips et al. 2004](#), [Elith et al. 2011](#)).

### **1.3.3 Evaluating the performance of spatiotemporal analytical tools**

The performance of SAT, especially the predictive modeling-and regression models (listed under T3 and T4 of the Table 1), are evaluated for their performance because the predictions would have no merit if the accuracy of the models cannot be assessed using independent data ([Fielding and Bell, 1997](#)). A variety of model evaluation and validation techniques can be used to assess the performance of predictive SATs (Listed under T3 and T4 of Table 1). These include the sensitivity and specificity measures (i.e. the number of correctly classified cases), area under the receiver operating characteristics (ROC) curves, evaluation of penalized-likelihood criteria such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) for regression based models evaluating associations between disease outcomes and risk factors ([Akaike, 1977](#)). The sensitivity of a spatial model in disease mapping can be defined as the model's ability to correctly predict the high-risk areas/locations, whereas, the specificity of the model would be its ability to correctly identify low-risk areas. Data partitioning techniques such as bootstrapping ([Buckland and Elston, 1993](#); [Verbyla and Litaitis 1989](#)); randomization ([Capen et al., 1986](#)); Prospective sampling ([Capen et al., 1986](#); [Fielding and Haworth, 1995](#)); and k-fold partitioning ([Stockwell, 1992](#)) are commonly used methods. Cross validation, i.e. partitioning the data into several subsets and each fitting the model excluding one subset and validating the fitted model's ability to correctly predict the risk areas using the excluded subset of data, is one of the common practices in spatial model validation. Similarly, the true validation, i.e. dividing the data over time and fitting model using early data/incidents and validation of the model predictions using recent disease occurrences is another possible approach. True validation is also achieved through the prospective sampling where new cases are evaluated against already built models from a different region or from a different time ([Fielding and Bell, 1997](#)).

### **1.3.4 Available software tools facilitating SAT**

Multiple free and proprietary software tools are available facilitating the spatiotemporal analytical studies. However, there is no quality control over to assess the accuracy, reliability, and

sustainability of majority of those non-proprietary software. Some software have become successful commercial products that are widely in use (Beale *et al.* 2008), while others are abandoned due to less popularity and irregular maintenance. Geocoding can be implemented using either commercial GIS software or online that are developed by governmental (Ex. [USGS map locator: https://store.usgs.gov/map-locator](https://store.usgs.gov/map-locator)), private ([ArcGIS Online Geocoding Service](https://geocode.arcgis.com/arcgis/) by Esri (<https://geocode.arcgis.com/arcgis/>); QGIS Geocoding Plugins (<https://plugins.qgis.org/plugins/GeoCoding/>); Geocoding using Google maps (<https://cloud.google.com/maps-platform>), or through educational organizations (e.g. [TAMU Geocoding Services of the University of Texas A & M: http://geoservices.tamu.edu/](http://geoservices.tamu.edu/)). These software enable both batch geocoding where multiple addresses are submitted at once for geocoding, and reverse geocoding, i.e. determining the nearest street address based on given coordinates.

The commonly used user-friendly software in the spatiotemporal analysis that are capable of performing the descriptive analysis, spatial pattern recognition, smoothing/interpolation, and/or spatial modeling are ArcGIS (ESRI, 2018), QGIS (QGIS Development Team, 2018), GRASS (GRASS Development Team, 2017), GeoDa (Anselin *et al.*, 2005; <http://geodacenter.github.io/index.html>), Clusterseer (User manual, 2012: <https://www.biomedware.com/>), SaTScan (<http://www.satscan.org/> version 9.6), and CrimeStat (Levine, 2010). Similarly, there are multiple toolboxes relevant to spatiotemporal analysis that can be used through following software: R statistical software (R core team, 2017), [SAS \(SAS 14.1 User Guide, 2015\)](#), STATA (StataCorp, 2017), and [Matlab \(Matlab: https://www.mathworks.com\)](https://www.mathworks.com) platforms that are specifically developed for handling geospatial analysis.

### 1.3.5 Other spatiotemporal analytical tools

In addition to the SAT discussed here, there are non-geostatistical spatial analytical tools such as cellular automata (CA) (Batty, 1997) and Agent-based modeling (ABM) (Crooks and Heppenstall, 2012) that are in use among the researchers in eco-epidemiological studies. The mathematical and geostatistical SAT utilizes a set of equations to model the associations within and between variables, while, the CA and ABM use graphical networks of nodes and edges to represent the associations (Abdou *et al.*, 2012). Both CA and ABM are grid based modeling approaches where a set of cells, i.e. “neighborhood”, is defined relative to a given cell. In CA, the spread of adverse health events are modeled using the state transition rules across discrete space

and time steps depending on the neighborhood (Batty, 1997; Abdou et al., 2012). In contrast, in ABM, an agent is defined and their spread is modeled using the characteristics of agents (perception, performance, memory, and policy) and their interaction with each other and the space/environment (Crooks and Heppenstall, 2012). Agents may represent any type of autonomous entity including people, animals, insects, or infectious agent that are capable of following the defined characteristics. The stochastic nature and flexibility in modeling is one of the key reasons why spatiotemporal epidemiological studies are increasingly interested in ABM.

Other types of models such as gravity models (Barrios et al., 2012; Holhe, 2016), and integrated approaches where spatial analysis is hybridized with surface/terrain analysis (Dister et al., 1997) and network analysis (Emch et al., 2012; Kanankege et al., 2018) are also being used in eco-epidemiological studies to capture flow of adverse health events or populations across geographical space. It is important to note that the choice of the SAT and other tools is entirely a researcher driven decision, however, there are certain factors/criteria associated with the decision of choosing one method over the other (Fig. 1.2). The factors include: 1) characteristics of the disease/adverse event; 2) study design; 3) spatial explicitness of data; 4) data quality and availability; 5) research question and hypothesis; 6) stakeholder involvement; and 7) existence of resources, policy, and regulations for the mitigation of events.



**Fig. 1.2.** Factors influencing the choice of spatiotemporal analytical tools (SAT).

### 1.3.6 Factors influencing the choice of spatiotemporal analytical tools

#### 1.3.6.1 Characteristics of the disease/adverse event

Interaction between host, pathogen, and environment, i.e. the epidemiological triad, results in disease transmission in populations. The epidemiological characteristics of the disease transmission, which are determined by triad components, including acute or chronic nature of onset, endemicity, emerging/re-emerging nature, and mode of transmission are to be considered when choosing a SAT. Often chronic and endemic diseases, such as Tuberculosis and Johne's disease, are needed to be evaluated considering both space and time, and therefore require spatiotemporal SATs. The mode of transmission: direct, airborne, or vector borne transmission often results in a secondary layer of complexity into the spatial distribution of diseases. Similarly, the human mediated transmission of animal diseases and other health or environmental threats, for example dispersal of invasive species, invariably increases the complexity of the spatial characteristics (Lindhal and Grace, 2015; Kanankege et al., 2018). Therefore, when choosing SAT to address such complex systems, the use of hybrid approaches, by bringing several SATS

to complement one another, would facilitate addressing the multifaceted nature of the transmission.. An example is the use of ENM to determine the ecological suitability for invasive species or suitability for mosquito vectors and combining the outputs of ENM using other tools listed under T3 and T4 to determine the high-risk areas for invasions or vector-borne diseases.

Characteristics of the susceptible population is another factor to consider when choosing SAT. For example, when analyzing spatial characteristics of farmed animals compared to wildlife, the human mediated transmission and historical background of the disease may play an important role than the environment itself. In such situations where susceptible populations are pre-defined by human involvement, tools such as Ecological Niche modeling may not be entirely suitable to assess the spatial characteristics of a disease. Modern risk factor epidemiology is often criticized for being focused on individual risk factors and being disconnected from examination of the broader historical and social/environmental forces that determine population disease risk (Wagner Robb et al., 2016). The influence of environmental factors makes the disease spread more complexed. It is possible that, even though there is no evident spatial dependence of pattern of the disease, the presence of the disease is associated with environmental characteristics in the location, such as soil characteristics and long-term survival of the pathogen in favorable soils.

### **1.3.6.2 Study design**

The evidence for risk-based or evidence-based decisions are gathered through epidemiological study designs, which are broadly categorized into experimental and observational in nature. Observational studies such as cohort, case-control, cross-sectional, survival studies, and ecological studies often results in spatially explicit data on adverse health events. Geo-statistical techniques described under spatial predictions (T3 and T4) are commonly used when evaluating data originated from cohort, case-control, and cross-sectional study designs. Data originated from ecological studies are likely to have only the disease cases (i.e. presence-only data), and no information on the confirmed locations where disease was not found. This characteristic of “presence-only data” poses a challenge in choosing an appropriate SAT to study the exposure to risk factors, because there is no ideal control group for comparison. ENM are models that use presence only data to parametrize the association between the adverse events and environmental risk factors. However, the choice of ENM is restricted by the factors including the susceptible population, unit of analysis, and spatial explicitness of event data. Moreover, the



hybrid approaches of SAT such as ENM and spatial regression may become useful when addressing data from various observational studies.

### **1.3.6.3 Spatial explicitness of data**

The characteristics of spatially explicit data defines the choice of the model. The common types of models include: 1) point-level models where the variable was measured at certain locations over continuous space (Ex. Temperature); 2) point process models where the locations of event occurrence are random (Ex. Disease occurrence) (Diggle, 2007); and 3) areal models where the events are aggregated by predefined polygonal areas such as administrative divisions (Ex. Disease prevalence by country) (Banerjee, Carlin, and Gelfand 2014). Reduced spatial explicitness may lead to the choice of epidemiological unit of analysis from individual farm level to administrative divisions and thus confine the researcher to use SATs comply with aggregated data such as conditional autoregressive models (listed under T4 of Table 1.1). It is also important to note that the aggregated data by administrative divisions are subjected to the modifiable areal unit problem (MAUP), where the unit-aggregated values (e.g. prevalence, densities) are influenced by both the shape and scale of the aggregation unit (Openshaw, 1984; Jelinski and Wu, 1996). MAUP is known to affect disease mapping (Morris and Munasinghe, 1997), spatial cluster detection (Waller and Turnbull, 1993), and correlation studies (Fotheringham, 1991). The availability of exact location details facilitates more explicit and reliable analysis.

Among the SATs discussed here, the spatial regression techniques such as conditional autoregressive models are commonly applied on the areal/aggregated data, although, the neighborhood matrices can be defined based on distance (E.g. presence of a farm within 10 km radius) and applicable on point level or point process data as well (Kanankege et al., in review BMC). ENM are commonly applicable on “presence only” individual level data i.e. individual animal or entity, for example, presence of a species or a wild animal with Anthrax. Spatial smoothing and interpolation methods are applicable to point level data, while, areal data are usually modeled using spatial regression methods such as CAR modeling.

### **1.3.6.4 Data quality and availability**

Depending on the source of data, the quality of data varies drastically. Research studies that are planned and executed often result in better quality data. However, ecological studies that are

based on active or passive surveillance systems are prone to biases (selection and information biases). Prior to choosing an appropriate SAT for predictive modeling, testing for the biases and the representativeness of data is an essential step. SATs such as spatial cluster analysis (listed under T2 of Table 1.1) becomes useful tools when testing for biasedness of the data and generating hypothesis on biasedness, such as hypothesizing human population density as a confounding factor influencing the reporting as well as the spread of invasive species (Kanankege et al., 2018). However, the use of auxiliary variables to account for underreporting/under-sampling and the use of spatial interpolation techniques such as co-kriging (T3 of Table 1.1) in such data scarce situation may be useful as alternative approaches (Kanankege et al., 2018). Another study which used a hybrid approach where spatial smoothing techniques and spatial pattern recognition method to account for data scarcity is found in Perez et al. (2005).

Data availability drives the analyses to be confined to administrative borders which could give rise to edge effect, where the areas along the edge of a study area gets misclassified because the events outside the borders are not included in the analysis. When the neighboring areas are exempt of population at risk, such as presence of sea, the edge effect is controlled. However, in the situation where the bordering area contains a susceptible population, the risk estimates are bound to be under- or over- estimated (Griffith and Amrhein, 1983; Perez et al., 2005). Although the choice of SAT may not permit resolving the edge effect, several techniques allowing the corrections are discussed in the published literature including generalized least squares solution, the dummy variable solution, and the missing data solution (Griffith and Amrhein, 1983; Griffith, 1985).

### **1.3.6.5 Research question and hypothesis**

Research questions and decisions on preventing health threats that involve spatial relationships have remained a minority in many of the disciplines. The most common research questions answered using spatial analysis include: calculations of proximity, estimation of summary measures across geographic areas, spatial pattern recognition such as clustering, estimation of the features at under sampled locations, and assessment of spatial association within the health events and association between health event and underlying risk factors (Auchincloss et al., 2012). Moreover, SAT are useful in hypothesis generating and testing processes where different tools are applicable at each stage (Mayer, 1983). The SATs listed under T2 of Table 1.1,

which support spatial and spatiotemporal pattern recognition, play a major role in the primary hypothesis generating stage. Spatiotemporal patterns includes: concentration (or dispersion), eccentricity, clusters, and measuring geometric aspects of pattern such as distance to source of the potential primary source of infection (Williams and Wentz, 2008).

The framework introduced here encourages the understanding of the spatial data at pre hypothesis testing and primary hypothesis testing stages prior to the modeling or secondary hypothesis testing. This approach may support narrowing down the research question while generating and testing novel hypotheses on variable associations.

### **1.3.6.6 Stakeholder involvement**

Stakeholders are “Any person or group who has an interest in the research topic and/or who stands to gain or lose from a possible policy change that, directly or indirectly, might be influenced by the research findings.” (Slunge et al. 2017). Involvement of stakeholders, especially during the planning stage is often beneficial in formulating decision-driven research studies (Slunge et al. 2017). This includes deciding on the unit of analysis, the level of detail required for making decisions, and understanding the characteristics of the threat for the public and the environment health. Similarly, communicating with the stakeholders throughout and at the end of the research is essential to refine analysis and understand the criteria of prioritizing and perceive of the risk based on their experience. For example, a stakeholder is primarily interested in defining the size of an area, when planning a surveillance or vaccination campaign against natural Anthrax in animal populations, rather than understanding the strength of association between occurrences of Anthrax with the alkalinity of soil. In such situations, the solution may be reached within the spatial/spatiotemporal pattern recognition category of SAT instead of further modeling (Kanankege et al., in review PLoS ONE).

The thresholds of risk associations and risk maps, in both qualitatively and quantitatively, are often based on arbitrary decisions of risk thresholds or cut-off values that are choices made upon agreements between researchers and stakeholders. Stakeholder suggestions on the thresholds may often influenced by the existing resources, policy, and regulations (discussed below). Therefore, having standards for risk thresholds becomes a challenging task. Moreover, when comparing risk maps on infectious diseases with multiple hosts or vectors, it is likely that there may be disagreements (Loth et al. 2011). The differences between risk maps may largely be attributed to the distribution of the population at risk, influence of changing environmental factors, and the data collection discrepancies (Loth et al. 2011).

### 1.3.6.7 Resources, policy, and regulations for the mitigation of the events

The existence of resources for prevention and control, policy, and regulations can influence the preparedness for mitigating strategies. Policy and regulations are often associated with the severity and the economic impact of the consequences of a certain health threat. Thus, when choosing a SAT to address a condition with existing policy or regulations, the likelihood of having a specific research question to answer is more likely to be a priority rather than pure exploratory approaches. For example, Anthrax is a reportable and vaccine preventable disease and a surveillance or vaccination program for Anthrax would require an optimal zone radius, which is achieved by SATs listed under T2, such as spatiotemporal cluster analysis ([Kanankege et al., in review- PLOS ONE](#)). Furthermore, it is likely that the data quality, availability, and technical capacity is richer in the reportable scenarios compared to the non-regulated conditions.

### 1.3.7 How to use the framework to choose SAT: An example

When connecting the seven factors with the framework introduced here, spatial explicitness of data, data quality and availability, and research question and hypothesis directly influences the five questions illustrated in the framework here (Fig 1.1). For example let's assume a researcher is primarily interested in understanding epidemiological characteristics of natural Anthrax and use that information to plan a surveillance program in an endemic area. Anthrax is a reportable disease. Among the seven factors, the characteristics of Anthrax is central to the other six factors because the pathogenic nature remains unchanged. Anthrax, which caused by the spore-forming bacterium *Bacillus anthracis*, is characterized by the prolonged survival of the spores on soil and wide range of hosts including wildlife, livestock, and human ([Sternbach, 2003](#); [Driks, 2009](#)). Therefore, the observational study designs on Anthrax are likely to be retrospective based on reported cases ([Kanankege et al., in review- PLOS ONE](#)). If the data are not geocoded and case reports contain location descriptions, the tools of geocoding listed under T1 of the Table 1.1 comes to play a role. The spatial explicitness of data is determined by the details in the case reports, which is also related to the data quality and the reliability. The type of spatial data likely to derive are point process data (presence of the disease), point level data (number of animals affected), and areal data (number of cases by an administrative division such as counties). Reduced spatial explicitness may lead to aggregation of the events by administrative divisions and non-availability of the temporal details would confine the researcher to use purely spatial

tools for the analysis. The availability of exact location details and the time of the case supports more explicit and reliable analysis.

After the step of testing for spatial dependence (using T1 tools), it is important to consider doing both spatial as well as spatiotemporal pattern detection on the data (T2 tools), considering the prolonged survival of the pathogen. The spatiotemporal pattern detection may lead to the refinement of research question and hypothesis and therefore, to choose a predictive modeling tool among T3 and T4 categories. Because the pathogen is invariably dependent upon the distribution of susceptible species and environmental characteristics such as soil pH, the choice of predictive modeling using correlated factors is ideal. However, it is important to recognize that, the ideal analysis for Anthrax would be a spatiotemporal correlation modeling approach instead of purely spatial analysis such as ecological niche modeling (Mullins *et al.*, 2013).

#### **1.4 Discussion and Conclusions**

This study introduced framework for choosing for choosing spatiotemporal analytical tools for eco-epidemiological studies; summarized commonly used SAT; and discussed the choice of SAT in relation to seven factors/criteria. Use of SATs improves the qualitative description of an eco-epidemiological investigation by adding precision, facilitating the comparison of distributions by means of quantitative criteria, and capturing risk factors and characteristics that are unlikely to be detected by visual inspection or analyzing data without the spatial component (Ward and Carpenter, 2000). Therefore, SAT outcomes may serve as estimates of the effects of “real” exposures to human, animal, and environmental health threats and facilitate recognizing the effect size at more vulnerable locations and time periods.

Common weaknesses associated with the spatial analysis and risk mapping are related to shortcomings in the accuracy and mapping of data in the geographical space, choice of the analytical/ modeling tools and relevant assumptions, and the decisions related to the representation of the risk maps to the end users (Ocaña-Riola, 2010; Loth *et al.* 2011). When spatial analytics and models are conducted based on available and potentially biased data, the resulting risk maps are invariably subjected to the negative impact of the data quality. However, we emphasize the use of existing data, bringing several databases together, and the spatiotemporal analytical tools can support initiating the process of improving data quality.

The choice of SAT, as discussed, varies with multiple factors. Inevitably, all analytical tools and models involve certain assumptions on statistical properties of variables and often these

assumptions are violated in natural environments. In other words, none of the SAT are precise matches for any particular situation (Susanto et al., 2016). For example, spatial continuity of risk is a common assumption in risk-mapping process while there can be natural (Ex. Mountain range acting as a physical barrier) or infrastructural barriers (Ex. Urban vs. rural neighborhoods) that violate the continuity assumption resulting in step changes of risk between adjacent areas (Rushworth et al, 2017). Therefore, clarity on the choice of SAT, underlying assumptions, and the seven factors/criteria is essential when choosing SAT to address eco-epidemiological problems.

## 1.5 Acknowledgements

This study was funded in part by the Minnesota Discovery, Research, and Innovation Economy (MnDRIVE) program and Office of the Vice President for Research (OVPR) of the University of Minnesota.

## 1.6 References

1. Abdou M, Hamill L, Gilbert N. 2012. Designing and Building an Agent-Based Model. Eds. Heppenstall AJ, Crooks AT, See LM, Batty M. Springer Dordrecht Heidelberg, NY. DOI 10.1007/978-90-481-8927-4.
2. Agterberg FP. 1984. Trend surface analysis. Spatial statistics and models, Gaile, G. L. and C. J. Willmott, eds., Reidel, Dordrecht, Holland, 147–171.
3. Akaike, H. 1977. On entropy maximization principle. Applications of Statistics, P. R. Krishnaiah, ed., North-Holland Publishing Co., Amsterdam, Holland, 27-41.
4. Alkhamis M, Moore BR, Perez AM. 2015. Phylodynamics of H5N1 highly pathogenic avian influenza in Europe, 2005-2010.: potential for molecular surveillance of new outbreaks. Viruses. 7: 3310-28. SoI: 10.3390/v7062773
5. Alkhamis MA, Hijmans RJ, Al-Enezi A, Martínez-López B, Perez AM. 2016. The use of spatial and spatiotemporal modeling for surveillance of H5N1 highly pathogenic avian influenza in poultry in the middle east. Bio One. Avian diseases. 60:146-155.

6. Anselin L. 1996. The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In *Spatial Analytical Perspective on GIS*, 111–25, edited by M. Fischer, H. Scholten, and D. Unwin. London: Taylor & Francis.
7. Anselin L. 1995. Local indicators of spatial association – LISA. *Geographical Analysis*. 27(2):93-115. doi: 10.1111/j.1538-4632.1995.tb00338.x. PubMed PMID: WOS:A1995QU19800001.
8. Arbia G. 2006, *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*, Springer-Verlag, Berlin.
9. Auchincloss AH, Gebreab SY, Mair C, Roux AVD. 2012. A Review of Spatial Methods in Epidemiology, 2000-2010. *Annual Review of Public Health*, Vol 33. 33:107-+. doi: 10.1146/annurev-publhealth-031811-124655. PubMed PMID: WOS:000304202700008.
10. Banerjee S, Carlin BP, Gelfand AE. 2014. *Hierarchical Modeling and Analysis for Spatial Data*, Second Edition, Monographs on Statistics and Applied Probability (2nd ed.), Chapman and Hall/CRC, ISBN 9781439819173
11. Banerjee S. 2016. Spatial survival models. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
12. Barrios JM, Verstraeten WW, Maes P, Aerts JM, Farifteh J, Coppin P. 2012. Using the Gravity Model to Estimate the Spatial Spread of Vector-Borne Diseases. *Int. J. Environ. Res. Public Health*. 9, 4346-4364.
13. Batty M. 1997. Cellular Automata and Urban Form: A Primer. *Journal of the American Planning Association*, 63, 266-274.
14. Beale L, Abellan JJ, Hodgson S, Jarup L. 2008. Methodologic issues and approaches to spatial epidemiology. *Environmental Health Perspectives*. Aug; 116:1105-1110.
15. Benhamou S. 2011. Dynamic approach to space and habitat use based on biased random bridges. *PLoS ONE* 6(1): e14592. Doi:10.1371/journal.pone.0014592

16. Besag J, York J and Mollié A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*. 43(1): 1–20.
17. Besag, J, Newell, J. 1991. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society A*, (154), 154, 143–55.
18. Biondi F. 2013. Space-time kriging extension of precipitation variability at 12 km spacing from tree-ring chronologies and its implications for drought analysis. *Hydrology and Earth System Sciences Discussions*, 10:4301–4335, 2013. URL: <http://dx.doi.org/10.5194/hessd-10-4301-2013>.
19. Blackburn JK, McNyset KM, Curtis A, Hugh-Jones ME. 2007. Modeling the geographic distribution of *Bacillus anthracis*, the causative agent of anthrax disease, for the contiguous United States using predictive ecologic niche modeling. *American Journal of Tropical Medicine and Hygiene* 77:1103-1110
20. Breiman L. 2001. Random forests. *Machine Learning* 45(1):5–32. DOI 10.1023/A:1010933404324
21. Brunson C, Fotheringham AS, Charlton ME. 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*. 28(4):281-98. doi: 10.1111/j.1538-4632.1996.tb00936.x. PubMed PMID: WOS:A1996VL03500001.
22. Buckland, S.T. & Elston, D.A. (1993) Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology* 30: 478–95.
23. Cai B, Lawson AB, Hossain MM, Choi J. 2012. Bayesian latent structure models with space-time dependent covariates. *Statistical Modelling* 12(2), 145–164.
24. Capen DE, Fenwick JW, Inkley DB, Boynton AC. 1986. Multivariate models of songbird habitat in New England forests. In: *Wildlife 2000: Modelling Habitat Relationships of Terrestrial Vertebrates*, ed. J.A. Verner, M.L. Morrison and C.J. Ralph, pp. 171–75. Madison, WI, USA: University of Wisconsin Press.
25. Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. 2014. Visualization and analytics tools for infectious disease epidemiology: A systematic review. *Journal of*



- Biomedical Informatics.51:287-98. doi: 10.1016/j.jbi.2014.04.006. PubMed PMID: WOS:000343362800028.
26. Chen X, Wang K. 2017. Geographic area-based rate as a novel indicator to enhance research and precision intervention for more effective HIV/AIDS control. *Preventive Medicine Reports*. 5:301–7.
  27. Clayton D, Kaldor J. 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. 43:671–681. doi: 10.2307/2532003.
  28. Clements ACA, Pfeiffer DU. 2009. Emerging viral zoonoses: Frameworks for spatial and spatiotemporal risk assessment and resource planning. *Veterinary Journal*, 182(1), 21-30. doi:10.1016/j.tvjl.2008.05.010
  29. Connelly NA, O'Neill CR, Knuth BA, Brown TL. 2007. Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. *Environ. Management*. 40, 105-112.
  30. CRARM: The Presidential/ Congressional Commission on Risk Assessment and Risk Management. 1997. Framework for Environmental Health Risk Management. Commission on Risk Assessment and Management, Washington DC. Vol.
  31. Cressie N, Wikle CK. 2011. Statistics for spatio-temporal data. *Wiley Series in Probability and Statistics*. ISBN: 978-0-471-69274-4
  32. Cressie NAC. 1993. Statistics for spatial data. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York.
  33. Cromley EK, McLafferty SL. 2002. GIS and Public Health. The Guilford Publications Inc., New York. ISBN: 1-57230-707-2
  34. Crooks AT, Heppenstall AJ. 2012. Agent-based models of geographical systems. Eds. Heppenstall AJ, Crooks AT, See LM, Batty M. Springer Dordrecht Heidelberg, NY. DOI 10.1007/978-90-481-8927-4.

35. Cuzick J, Edwards R. 1990. Spatial clustering for inhomogenous populations. *Journal of the Royal Statistical Society. Series B.* 52, 73-104.
36. Davis JC. 1973. *Statistics and Data Analysis in Geology*. New York: John Wiley & Sons, 550p.
37. Diggle P, Rowlingson B, Su T. 2005. Point process methodology for on-line spatiotemporal disease surveillance. *Environmetrics* 16, 423–434.
38. Diggle P. 2007. Spatio-temporal point processes: methods and applications. In Finkenstadt B., L. Held, and V. Isham (eds.), *Statistical Methods for Spatio-Temporal Systems*, pp. 1–45. London: CRC Press.
39. Dixon PM. 2002. Ripley’s K function. *Encyclopedia of Environments*. Eds. El-Shaarawi AH and Piegorisch WW. 3: 1796-1803. ISBN: 0471899976.
40. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. 36(1):27-46. doi: 10.1111/j.1600-0587.2012.07348.x. PubMed PMID: WOS:000315892600003.
41. Driks A. 2009. The Bacillus anthracis spore. *Molecular Aspects of Medicine*. 30(6):368-73. doi: 10.1016/j.mam.2009.08.001. PubMed PMID: WOS:000272755000004.
42. Eldeiry AA, Garcia LA. 2010. Comparison of Ordinary Kriging, Regression Kriging, and Cokriging Techniques to Estimate Soil Salinity Using LANDSAT Images. *Journal of Irrigation and Drainage Engineering*. 136(6):355-64. doi: 10.1061/(asce)ir.1943-4774.0000208. PubMed PMID: WOS:000277745200001.
43. Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*; 17: 43-57.
44. Elliott P, Wartenberg D. 2004. Spatial epidemiology: Current approaches and future challenges. *Environmental Health Perspectives* 112:998-1006.

45. ESRI: ArcMap Version 10.6. 2018. Environmental Research Institute, Inc. Redlands, CA, USA.
46. Fielding AH, Haworth PF. 1995. Testing the generality of bird-habitat models. *Conservation Biology* 9: 1466–81.
47. Fielding AL, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38-49
48. Fotheringham AS, Charlton ME, Brunsdon C. 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ. Plann. A.* 30:1905–1927.
49. Fotheringham AS, Wong DWS. 1991. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environ Plan A.* 23:1025–1044
50. Frome EL, Checkoway H. 1985. Use of Poisson Regression Models in Estimating Incidence Rates and Ratios. *Am J Epidemiol.* 121:309–323
51. Geary RC. 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician* 5(3): 115–145.
52. Gelman A, Price PN, Lin C. 2000. A method for quantifying artefacts in mapping methods illustrated by application to headbanging. *Statistics in Medicine* 19: 2309-20
53. Getis A, Ord JK. 1992. The analysis of spatial association by distance statistics. *Geogr. Anal.* 24 (3), 189–206
54. Goovaerts P. 1997. *Geostatistics for natural resource evaluation*, Oxford University Press, New York, New York.
55. GRASS Development Team. 2017. *Geographic Resources Analysis Support System (GRASS) Software, Version 7.2*. Open Source Geospatial Foundation. Electronic document: <http://grass.osgeo.org>

56. Griffith DA, Amrhein CG. 1983. An Evaluation of Correction Techniques for Boundary Effects in Spatial Statistical Analysis: Traditional Methods. *Geographical Analysis*, Vol. 15(4): 352–360.
57. Griffith DA. 1985. An Evaluation of Correction Techniques for Boundary Effects in Spatial Statistical Analysis: Contemporary methods. *Geographical Analysis*. 17(1): 81-88.
58. Haining R. 2003. *Spatial data analysis: theory and practice*. Cambridge University Press, Cambridge.
59. Hart TC, Zandbergen PA. 2013. Reference data and geocoding quality Examining completeness and positional accuracy of street geocoded crime incidents. *Policing-an International Journal of Police Strategies & Management*. 36(2):263-94. doi: 10.1108/13639511311329705. PubMed PMID: WOS:000320741900003.
60. Hazelton ML. 2016. Kernel smoothing methods. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
61. Hernán MA, Robins JM. 2006. Estimating causal effects from epidemiological data, *J Epidemiol Community Health*. vol. 60 7(pg. 578-586)
62. Hohle M. 2016. Infectious Disease Modeling. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
63. Huanxin W, Nan H. 1991. Trend-surface analysis was applied to the study of the geographic distribution for LBWI in China. *Chinese J Hlth Stat*, 8, 12
64. Jacquez GM, Greiling D, Durbeck H, Estberg L, Do E, Long E, Rommel B. 2002. *ClusterSeer. User Guide 2: Software for Identifying Disease Clusters*. TerraSeer Press, Ann Arbor, MI. 316 pp

65. Jacquez GM. 2008. Spatial Cluster Analysis (Chapter 22). *The Handbook of Geographic Information Science*. S. Fotheringham and J. Wilson (Eds.). Blackwell Publishing, pages 395-416
66. Jensen BB, Ersboll AK, Korsholm H, Skall HF, Olesen NJ. 2014. Spatio-temporal risk factors for viral haemorrhagic septicaemia (VHS) in Danish aquaculture. *Diseases of Aquatic Organisms*, 109(2), 87-97. doi:10.3354/dao02706.
67. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. 2008. Global trends in emerging infectious diseases. *Nature*. 451(7181):990-U4. doi: 10.1038/nature06536. PubMed PMID: WOS:000253313100048.
68. Kafadar K. 1996. Smoothing geographical data, particularly rates of disease. *Statistics in Medicine*. 15:2539–2560
69. Kanankege KST, Alkhamis MA, Phelps NBD, Perez AM. 2018. A probability co-kriging model to account for reporting bias and recognize areas at high risk for zebra mussels and Eurasian watermilfoil invasions in Minnesota. *Frontiers in Veterinary Science*. doi: 10.3389/fvets.2017.00231
70. Keckler D. 1995. *The Surfer Manual*. Golden, CO.: Golden Software, Inc.
71. Kelsall JE, Diggle PJ. 1995. Kernel estimation of relative risk. *Bernoulli*. 1(1–2), 3–16.
72. Kirby RS, Delmelle E, Eberth JM. 2017. Advances in spatial epidemiology and geographic information systems. *Annals of Epidemiology*. 27(1):1-9. doi: 10.1016/j.annepidem.2016.12.001. PubMed PMID: WOS:000393268100001.
73. Knox EG. 1964. The detection of space-time interactions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 13, 25-30.
74. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. 2005. A space-time permutation scan statistic for disease outbreak detection. *Plos Medicine*. 2(3):216-24. doi: 10.1371/journal.pmed.0020059. PubMed PMID: WOS:000228382500013.

75. Kulldorff M, Nagarwalla N. 1995. Spatial disease clusters and inference. *Stat. Med.* 14, 1. 799–810.
76. Kulldorff M. 1997. A spatial scan statistic. *Commun. Stat.: Theory Methods* 26, 1481–1496.
77. Kulldorff M. 2009. SaTScanTMv9.4.1: software for the spatial and space-time scan statistics. <http://www.satscan.org/>.
78. Lawson AB, Bohning D, Biggeri A, et al, eds. 1999. Disease mapping and its uses. disease mapping and risk assessment for public health. Chichester: Wiley.
79. Lawson AB, Choi J, Cai B, Hossain M, Kirby RS, Liu JH. 2012. Bayesian 2-Stage Space-Time Mixture Modeling With Spatial Misalignment of the Exposure in Small Area Health Data. *Journal of Agricultural Biological and Environmental Statistics.* Sep;17:417-441.
80. Lawson AB. 2006. *Statistical methods in Spatial Epidemiology.* 2nd Ed. John Wiley & Sons Ltd., The Atrium, Southern Gate, Chichester, West Sussex, England. ISBN-13: 978-0-470-01484-4
81. Lawson AB. 2013. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology.* Chapman and Hall/CRC press.
82. Leroux BG, Lei X, Breslow N. 2000. Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials.* Springer. pp. 179–191.
83. Levefer DW. 1926. Measuring geographic concentration by means of the standard deviation ellipse. *Am. J. Sociol.* 32, 88–94.
84. Levine N. 2010. *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.3).* Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. July.
85. Lindahl JF, Grace D. 2015. The consequences of human actions on risks for infectious diseases: a review. *Infect Ecol Epidemiol* ;5:30048. doi: 10.3402/iee.v5.30048. eCollection 2015.

86. Loth L, Gilbert M, Wu J, Czarnecki C, Hidayat M, Xiao X. 2011. Identifying risk factors of highly pathogenic avian influenza (H5N1 subtype) in Indonesia. *Preventive Veterinary Medicine*. Volume 102, Issue 1, 1 October 2011, Pages 50–58.
87. Lusting LK. 1969. Trend-surface analysis of the Basin and Range Province, and some geomorphic implications, U.S. Geol. Survey Prof. Paper 500-D, 70p.
88. MacNab YC, Dean CB. 2004. Autoregressive Spatial Smoothing and Temporal Spline Smoothing for Mapping Rates. *Biometrics*. 57(3): 949-956. doi.org/10.1111/j.0006-341X.2001.00949.x.
89. MacNab YC, Gustafson P. 2007. Regression B-spline smoothing in Bayesian disease mapping: With an application to patient safety surveillance. *Statistics in Medicine*. 26(24):4455-74. doi: 10.1002/sim.2868. PubMed PMID: WOS:000250158100007.
90. MacQueen JB. 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Symposium on Math, Statistics, and Probability* (pp. 281–297). Berkeley, CA: University of California Press.
91. Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*. 27 (2): 209–220.
92. March D, Susser E. 2006. The eco- in eco-epidemiology. *International Journal of Epidemiology*. 35(6):1379-83. doi: 10.1093/ije/dyl249. PubMed PMID: WOS:000243806400002.
93. Mariella L, Tarantino M. 2010. Spatial Temporal Conditional Auto-Regressive Model: A New Autoregressive Matrix. *Australian Journal of Statistics*. 39(3):223-244.
94. Matheron G. 1971. *The Theory of the Regionalized Variables and its Applications; Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau No. 5; Published by the École Nationale Supérieure des Mines de Paris*, p 211.
95. MATLAB and Statistics Toolbox TM Release. 2012. The MathWorks, Inc., Natick, Massachusetts, United States.

96. Mayer JD. 1983. The role of spatial analysis and geographic data in the detection of disease causation. Pergamon Press Ltd. Soc. Sci. Med. Vol.17: (16) 1213-1221
97. Merow C, Smith MJ, Silander Jr. JA. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*; 36: 001-012.
98. Mi C, Huettmann F, Guo Y, Han X, Wen L. 2016. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* 5:e2849. DOI 10.7717/peerj.2849
99. Mitas L, Mitasova H. 1999. Spatial Interpolation. In: P.Longley, M.F. Goodchild, D.J. Maguire, D.W.Rhind (Eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications*, Wiley
100. Moore DA, Carpenter TE. 1999. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev.* 21(2):143-6
101. Moran PAP. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37, 17–23.
102. Morris RD, Munasinghe RL. 1993. Aggregation of existing geographic regions to diminish spurious variability of disease rates. *Stat Med.* 12:1915–1929.
103. Mullins JC, Garofolo G, Van Ert M, Fasanella A, Lukhnova L, Hugh-Jones ME, et al. 2013. Ecological Niche Modeling of *Bacillus anthracis* on Three Continents: Evidence for Genetic-Ecological Divergence? *Plos One.* 8(8):8. doi: 10.1371/journal.pone.0072451. PubMed PMID: WOS:000323425700170.
104. Mungiole M, Pickle LW, Simonson KH. 1999. Application of a weighted headbanging algorithm to mortality data maps. *Statistics in Medicine.* 18:3201–3209.
105. Nakaya T, Fotheringham AS, Brunson C, Charlton M. 2005. Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine.* 24(17):2695-717. doi: 10.1002/sim.2129. PubMed PMID: WOS:000231816900010.

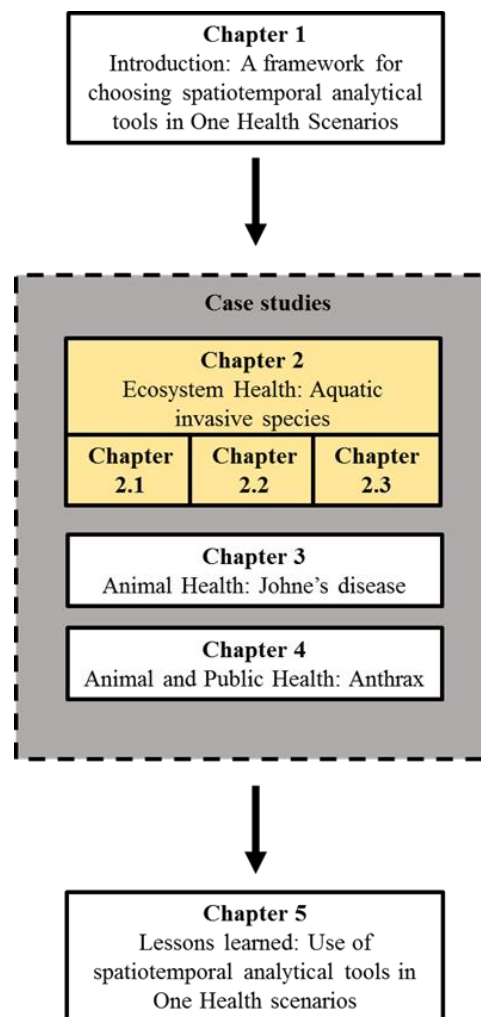


106. Oden N. 1995. Adjusting Moran's I for population density. *Statistics in Medicine* 14: 17-26.
107. Openshaw S. 1984. *The Modifiable Areal Unit Problem. Concepts and Techniques in Modern Geography*. Norwich, UK: Geo Books.
108. Ord JK, Getis A. 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, 27:287
109. Perez AM, Thurmond MC, Grant PW, Carpenter TE. 2005. Use of the scan statistic on disaggregated province-based data: Foot-and-mouth disease in Iran. *Preventive Veterinary Medicine*. 71(3-4): 197-207. DOI: 10.1016/j.prevetmed.2005.07.005.
110. Peterson AT, Sanchez-Cordero V, Ben Beard C, Ramsey JM. 2002. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. *Emerging Infectious Diseases*. 8(7):662-7. doi: 10.3201/eid0807.010454. PubMed PMID: WOS:000176394200004.
111. Peterson AT, Tian H, Martínez-Meyer E, Soberón J, Sánchez-Cordero V, Huntley B. 2005. Modeling distributional shifts of individual species and biomes. In: Lovejoy TE, Hannah L, editors. *Climate change and biodiversity*. New Haven (CT): Yale University Press. p. 211–28.
112. Pfeiffer DU, Otte MJ, Roland-Holst D, Zilberman D. 2013. A one health perspective on HPAI H5N1 in the Greater Mekong sub-region. *Comparative Immunology Microbiology and Infectious Diseases*, 36(3), 309-319. doi:10.1016/j.cimid.2012.11.005
113. Pfeiffer DU, Stevens KB. 2015. Spatial and temporal epidemiological analysis in the Big Data era. *Preventive Veterinary Medicine*, 122(1-2), 213-220. doi:10.1016/j.prevetmed.2015.05.012.
114. Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259
115. QGIS Development Team. 2018. QGIS Geographic Information System. Open Source Geospatial Foundation. URL <http://qgis.org>

116. R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
117. Ricardo Ocaña-Riola. 2010. Common errors in disease mapping. *Geospatial Health* 4(2): 139-154
118. Ripley BD. 1977. Modelling spatial patterns. *J. R. Stat. Soc. Series B Stat. Methodol.*, 39, pp. 172-192
119. Sanson RL, Pfeiffer DU, Morris RS. 1991. Geographic information systems: their application in animal disease control. *Rev. sci. tech. Off. int. Epiz.* 10 (1), 179-195
120. SAS Institute Inc. 2015. SAS/STAT® 14.1 User's Guide. Cary, NC: SAS Institute Inc.
121. Silverman BW. 1986. *Density estimation for statistics and data analysis*, Boca Raton, FL: Chapman & Hall/CRC.
122. Slunge D, Drakenberg O, Ekblom A, Göthberg M, Knaggård A, Sahlin U. 2017. Stakeholder interaction in research process - A guide for researcher and research groups. Available online at: [https://gmv.gu.se/digitalAssets/1619/1619929\\_stakeholder-interaction-in-research-processes---guide---final-march-15-2017.pdf](https://gmv.gu.se/digitalAssets/1619/1619929_stakeholder-interaction-in-research-processes---guide---final-march-15-2017.pdf)
123. Song TM, Ryu S. 2015. Big data analysis framework for healthcare and social sectors in Korea. *Healthc Inform Res.* 21:3–9.10.4258/hir.2015.21.1.3
124. Stark CR, Mantel N. 1967. Lack of seasonal or temporal spatial clustering of Down's Syndrome births in Michigan. *American Journal of Epidemiology* 86: 199-213.
125. StataCorp. 2017. *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC
126. Sternbach G. 2003. The history of anthrax. *Journal of Emergency Medicine.* 24(4):463-7. doi: 10.1016/s0736-4679(03)00079-9. PubMed PMID: WOS:000182902600019.
127. Stevens KB, Pfeiffer DU. 2016. The Role of Spatial Analysis in Risk-Based Animal Disease Management. *Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman &

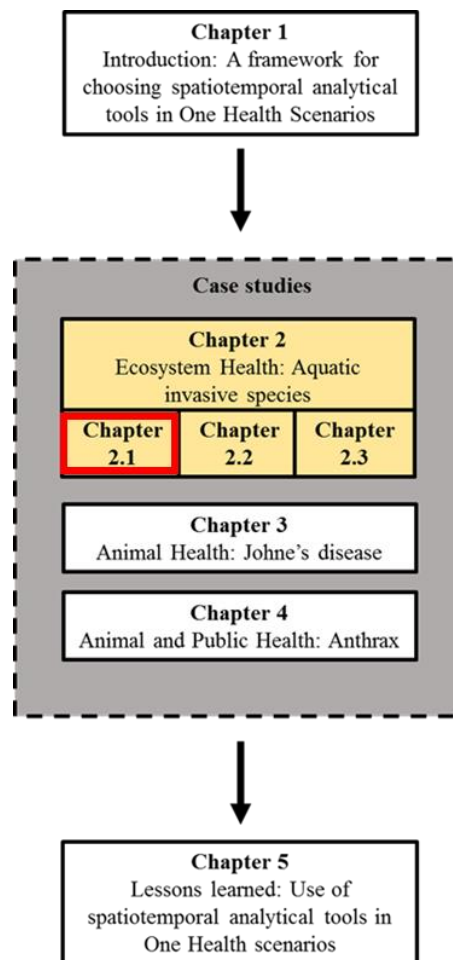
- Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
128. Stockwell D, Peters D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *Int J Geogr Inf Sci* 13: 143–158.
  129. Stockwell DRB, Peterson AT. 2002. Effects of sample size on accuracy of species distribution models. *Ecol Modell* 148: 1–13.
  130. Stockwell DRB. 1992. Machine learning and the problem of prediction and explanation in ecological modelling. Ph.D. Thesis, Australian National University.
  131. Susanto F, de Souza P, He J. 2016. Spatiotemporal Interpolation for Environmental Modelling. Reindl L. Eds. *Sensors (Basel)*. 16(8): 1245. doi: 10.3390/s16081245
  132. Susser E. 2004. Eco-epidemiology: Thinking outside the black box. *Epidemiology*. 15(5):519-20. doi: 10.1097/01.ede.0000135911.42282.b4. PubMed PMID: WOS:000223382600004.
  133. Tobler WR. 1970. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* 46: 234 -240.
  134. Tomczak M. 1998. Spatial Interpolation and its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW) - Cross-Validation/Jackknife Approach. *Journal of Geographic Information and Decision Analysis*, vol. 2, no. 2, pp. 18-30.
  135. Turnbull BW, Iwano EJ, Burnett W, Howe HL, Clark LC. 1990. Monitoring for clusters of disease: Applications to leukaemia incidence in upstate New York. *American Journal of Epidemiology* 132, S136–S143.
  136. Verbyla DL, Litaitis JA. 1989. Resampling methods for evaluating classification accuracy of wildlife habitat models. *Environmental Management* 13: 783–7.
  137. Wagner Robb S, Bauer SE, Vena JE. 2016. Integration of Different Epidemiologic Perspectives and Applications to Spatial Epidemiology. *Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Eds: Lawson AB, Banerjee S, Haining RP,

- Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
138. Wagner RS, Bauer SE, Vena JE. 2016. Integration of Different Epidemiologic Perspectives and Applications to Spatial Epidemiology. Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
139. Wall MM. 2004. A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, pp. 311-324.
140. Waller LA, Turnbull BW. 1993. The Effects of Scale on Tests for Disease Clustering. *Stat Med*. 12:1869–1884.
141. Wand MP, Jones MC. 1995. Kernel smoothing, Boca Raton, FL: Chapman & Hall/CRC.
142. Ward MP, Carpenter TE. 2000. Analysis of time-space clustering in veterinary epidemiology. *Prev Vet Med*. 43:225–37.10.1016/S0167-5877(99)00111-7.
143. Zandbergen PA. 2008. A comparison of address point, parcel and street geocoding techniques. *Computers Environment and Urban Systems*. 32(3):214-32. doi: 10.1016/j.compenvurbsys.2007.11.006. PubMed PMID: WOS:000257017100005.



## Chapter 2: Ecosystem Health

*The use of spatial cluster analysis, co-kriging, and maximum entropy ecological niche modeling as tools to inform decisions in aquatic invasive species management in Minnesota*



## Chapter 2.1

### *Recognizing Zebra mussels and Eurasian watermilfoil reporting patterns in Minnesota using spatial cluster analysis*

Published in Journal of Great lakes research (2018):

Kanankege KST, Alkhamis MA, Perez AM, Phelps NBD. 2018. Zebra mussels and Eurasian watermilfoil reporting patterns in Minnesota. Journal of Great Lakes Research. doi:

10.1016/j.jglr.2018.02.008

## 2.1 Overview

Recognizing common reporting patterns of aquatic invasive Zebra mussels (*Dreissena polymorpha*, ZM) and Eurasian watermilfoil (*Myriophyllum spicatum*, EWM) helps to better understand invasions. We hypothesize that confirmed invasions may be confounded by human population density, leading to overrepresentation of invasions in highly populated areas and underrepresentation in less populated areas. This work recognizes dispersal patterns of confirmed ZM and EWM invasions in Minnesota, USA, using spatial clustering and directionality tests, while adjusting for human density. By 2015, 125 (0.68%) and 304 (1.65%) of 18,411 Minnesota waterbodies were reported to have ZM and EWM, respectively. A multivariate multinomial model of the scan test was used to identify spatial clustering of invasions. The resulting 23 clusters included 13 with either or both ZM and EWM, and most clusters (11/13) occurred in areas with >10 people per square kilometer. Whereas, among the 10 clusters without invasion, nine were from less populated areas. The standard deviation ellipse and the spatiotemporal directionality tests indicated a northwestern trend of invasions, which is in the same direction as the I-94 interstate highway connecting urban centers. Results suggested that confirmed ZM and EWM invasions are potentially confounded by human densities, which is explained by varying human impact on either or both dispersal and reporting of invasions. Considering this impact of human density, we suggest a combination of passive and targeted surveillance where the magnitude of efforts are stratified by the human densities may provide insight into the true invasion status and its progression in the Great Lakes region.

## 2.2 Introduction

Aquatic invasive species (AIS) disrupt the stability of ecosystems and are difficult to eradicate (Pysek and Richardson, 2010). The state of Minnesota has experienced numerous AIS incursions and spends over 10 million dollars each year on activities intended to prevent, control, or manage AIS (Invasive Species Program 2016; MN Statute 477A.18, 2016). Eurasian watermilfoil (*Myriophyllum spicatum*) and Zebra mussels (*Dreissena polymorpha*) are AIS of concern for Minnesota and were first reported in the state in 1987 and 1989, respectively. The official records of confirmed aquatic invasions are publicly available from the Minnesota Department of Natural Resources (MNDNR). Despite efforts focused on control and prevention, new cases of zebra mussel and Eurasian watermilfoil invasions keep accumulating every year (<http://www.dnr.state.mn.us/invasives/ais/infested.html>. Cited January 21, 2018).

The initial introduction of zebra mussel into the Great Lakes region of North America has been attributed to the unintentional transport in the ballast water of transatlantic ships (Carlton, 2008; Brown and Stepien et al., 2010). According to genetic diversity analyses, the ancestral locations of zebra mussels may have varied across Eurasia suggesting multiple independent translocations of zebra mussels into North America (Brown and Stepien et al., 2010). Zebra mussels are rapidly propagating bivalves that disrupt the stability of the food web in aquatic ecosystems, affecting both pelagic and benthic species (Karatayev et al., 2015). Removal of zebra mussels colonizing public water supplies and industrial facilities pipes is expensive (Connelly et al., 2007).

Eurasian watermilfoil, an invasive aquatic macrophyte, was likely introduced into North America through the aquarium trade (Les and Merhoff, 1999). Asia is the likely source of Eurasian watermilfoil based on genetic analysis (Moody et al., 2016). Eurasian watermilfoil grows rapidly, limiting the ability to implement effective removal or control strategies upon establishment in a waterbody (Lodge et al., 2006; Roley and Newman, 2008). Dense vegetation of Eurasian watermilfoil outcompetes native macrophytes and interrupts recreational activities (Cheruvelli et al., 2002). Understanding the dispersal patterns common to both zebra mussels and Eurasian watermilfoil, such as the clustering of confirmed invasions, may support recognizing the drivers behind both the spread of the invasions and potential bias in public reporting.

Public observation of new AIS invasions plays a major role in data collection in Minnesota ([http://www.dnr.state.mn.us/invasives/report\\_invasives.html](http://www.dnr.state.mn.us/invasives/report_invasives.html). Cited January 22, 2018). There is no active systematic surveillance for AIS with the objective of early detection, although,



some of the invasions were discovered during surveys by the MNDNR. In the absence of active surveillance, AIS invasions are primarily recorded based on public reports and subsequent confirmation by the MNDNR ([http://www.dnr.state.mn.us/invasives/report\\_invasives.html](http://www.dnr.state.mn.us/invasives/report_invasives.html) Cited January 22, 2018). We hypothesize that the confirmed invasions may be confounded by human population density, leading to overrepresentation of invasions in highly populated areas and underrepresentation in less populated areas. Briefly, a confounder is a factor associated with both the exposure and the output of interest, which may distort the magnitude of the relationship between exposure and output (Jolley *et al.*, 1992; Szklo and Nieto, 2007). For example, consider the exposure of interest as the presence of invaded waterbodies in the area and the output as new invasions in neighboring waterbodies. Having a large human population may result in high frequency boater traffic between waters, increasing the exposure levels, as well as the possibility of higher frequency observations due to the increased traffic. A similar phenomenon related to a terrestrial invasive species has been explained by Aikio *et al.* (2010). Therefore, it is important to adjust for confounding by human population density when analyzing the publicly reported data.

Thus, the objective of this study was to identify clustering patterns of confirmed zebra mussels and Eurasian watermilfoil invasions, while accounting for the potential confounding by human population density. To that end, estimates of spatial or spatiotemporal clustering patterns and the directionality of AIS reporting using geostatistical methods would help to quantify the dispersal dynamics. Recognition of the common dispersal patterns would contribute to better understanding of the invasions and generating hypotheses on the mechanisms of spread and underlying risk factors, comparable to the idea behind infectious disease spatial dynamics in epidemiological studies (Ward and Carpenter, 2000; Elliott and Wakefield, 2001). Understanding the human influence on reported invasions would guide the implementation of active surveillance efforts and encourage passive surveillance and public reporting through educational activities in the Great Lakes region.

## 2.3 Data and Methods

### 2.3.1 Data

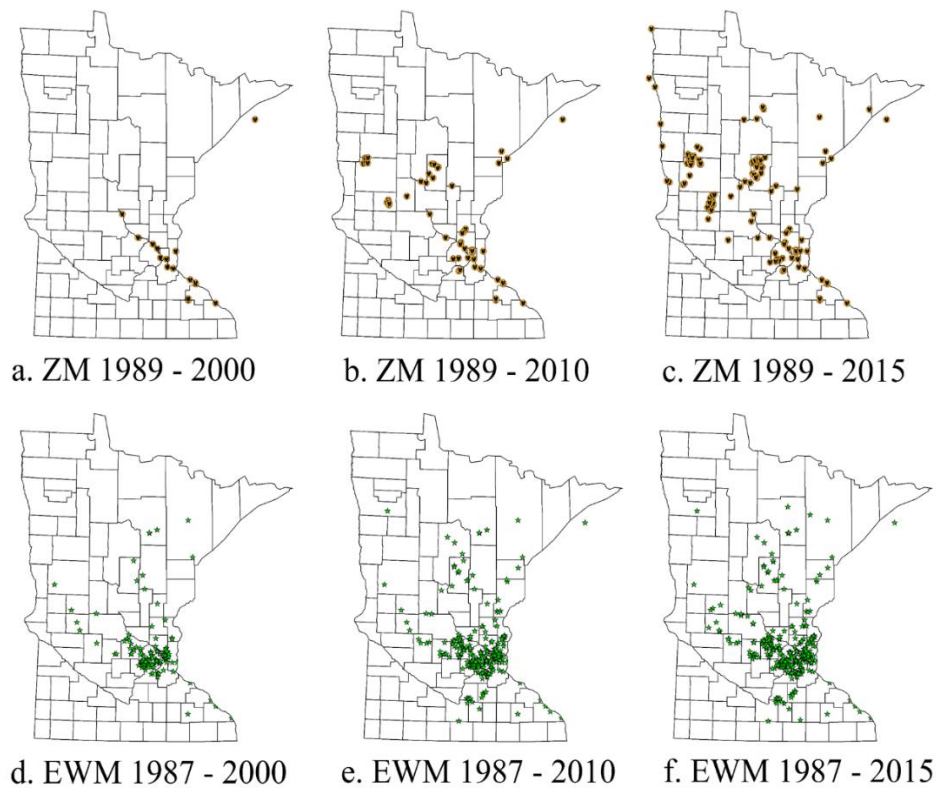
The units of analysis here were waterbodies (n=18,411), mainly lakes and ponds (n=18,253) and a few river locations (n=148). Locational data for waterbodies were extracted from the spatial layer referred to as “MNDNR Hydrography,” which is available from the Minnesota GIS Commons ([MNGSC Hydro, 2015](#)). Waterbodies were represented by their centroids, regardless of the size.

AIS presence data were obtained from the MNDNR’s publicly available online database. The invasions are reported to the MNDNR by the public and MNDNR confirms the newly reported invasions prior to adding them to their online database

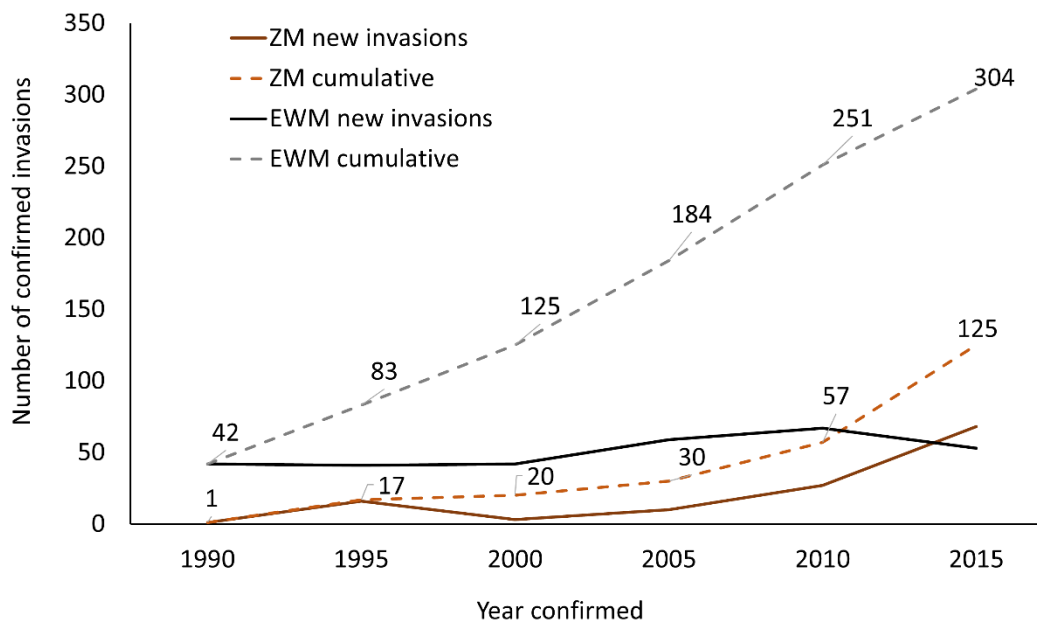
(<http://www.dnr.state.mn.us/invasives/ais/infested.html>. Cited June 22, 2015). When long rivers were confirmed with invasions, several Minnesota counties were included in the data records.

Therefore, invaded riverine locations were represented by the rivers’ midpoint within each county, according to the data records. The number of new cases and the cumulative number of cases confirmed by the MNDNR were used for the analysis (Fig. 2.1). The temporal data only includes the year in which the invasion was confirmed. By the end of 2015, the number of waterbodies invaded by zebra mussels and Eurasian watermilfoil were 125/18,411 (0.68%) and 304/18,411 (1.65%), respectively (Fig. 2.2). The confirmed presence of the AIS was used here to define a case regardless of the magnitude of infestation. This study considers all lakes, ponds, and certain riverine locations, as the waterbodies at risk for invasions regardless of their size, because the invaded waterbodies ranged between 0.006 km<sup>2</sup> to ~ 82,000 km<sup>2</sup> (i.e. Lake Superior).

Geographical locations were mapped using ArcMap 10.3 ([ESRI, 2016](#)). The waterbodies were categorized based on the invasion status into four invasion categories: 1) reported with both zebra mussels and Eurasian watermilfoil (EWM+/ZM+; n=21), 2) reported only with zebra mussels (EWM-/ZM+; n=104), 3) reported with Eurasian watermilfoil only (EWM+/ZM-; n=283), and 4) as of 2015, not reported with either of the AIS (ZM-/EWM-; n=18,003).



**Fig. 2.1.** The distribution of each invasive species at certain time intervals. Panels a, b, and c represents Zebra mussels (ZM) distribution between 1989 – 2015. Eurasian watermilfoil (EWM) distribution between 1987 - 2015 is illustrated in panels d, e, and f.



**Fig. 2.2.** Reported and confirmed new invasions and accumulated number of cases of zebra mussels (ZM) and Eurasian watermilfoil (EWM) in Minnesota waterbodies by year (Based on the Minnesota Department of Natural Resources database: MNDNR: <http://www.dnr.state.mn.us/invasives/ais/infested.html>. Cited June 22, 2015).

Human population density data were obtained from the LandScan Global Population Database of 2011 census, assuming 2011 is a representative year for population density between the study period 1987 to 2015 (Bright et al., 2013). To facilitate interpretation of the results, maps of selected major roads and major river centerlines were incorporated at the completion of analysis. The road map of 2012, available through Minnesota Geospatial Commons and originated from the Minnesota Department of Transportation was used to extract major roads (MNGSC Roads, 2012). As defined in the metadata of the spatial layer, U.S. road classes including interstate highways, freeways, arterials, and major collectors were considered as major roads in the analysis (MNGSC Roads, 2012). The major river centerlines were extracted from the “MNDNR Hydrography” spatial layer available from Minnesota Geospatial Commons (MNGSC Hydro, 2015). The rivers longer than 200 kilometers were considered as the major rivers.

### 2.3.2 Methods

Techniques for detection of directionality and spatial clustering (Levefer, 1926; Kulldorff, 2009; Jung et al., 2010; Wang et al., 2015) were used to quantify the pattern of confirmed AIS dispersal in Minnesota. The multivariate multinomial model of the scan statistics test was used to detect spatial clusters. To further examine the hypothesis that known invasions are confounded by the urban centers, directionality of the confirmed invasions was detected using the standard deviation ellipse and the spatiotemporal directionality tests as explained below.

### 2.3.3 Statistical Analysis

#### *Global cluster analysis*

The intent of clustering analysis is to test whether the invaded locations are spatially grouped compared to the null hypothesis of random distribution across the geographical space. The global cluster analysis only detects the tendency to cluster and does not point out where the clusters are located. The global cluster analysis based on the Cuzick and Edwards test was performed using the ClusterSeer v.2.05. software for each AIS data set separately (Cuzick and Edward, 1990; Jacquez et al., 2012; <https://www.biomedware.com/>). The Cuzick and Edwards's test compares observed number of AIS invaded locations whose nearest neighbor is also a case, among the  $k$  nearest neighbors, with the number of paired cases that would be expected under the null hypothesis of random spatial distribution (Cuzick and Edward, 1990). The test was run from the first up to tenth  $k$  neighborhood level. Significance of the clustering was determined at the  $p$  value 0.05. A detailed description regarding the cluster analysis using the Cuzick and Edwards test is described elsewhere (Cuzick and Edward, 1990).

#### *Multivariate multinomial local cluster analysis*

To locate statistically significant local clusters, a purely spatial multinomial model of the scan statistics was performed using the SaTScan software (Kulldorff et al., 2007; Kulldorff, 2009; Jung et al., 2010). The purely spatial multinomial cluster analysis detects any spatial aggregation of the four invasion categories (i.e. EWM-/ZM-; EWM+/ZM-; EWM-/ZM+; EWM+/ZM+) regardless of the time (Jung et al., 2010). Spatial scan statistics was based on a likelihood ratio test comparing potential cluster with the remaining area. A circular scanning window of varying sizes that moves across the study area was used to compare the ratio of the cases and the control

locations (Kulldorff and Nagarwalla, 1995). Then the test statistics for each window was calculated, and the windows that maximize the likelihood ratio were recognized as the most likely clusters.

To adjust for the potential confounding related to high human population density and the invasions in certain areas, clusters were analyzed by adjusting for the population density making the analysis a multivariate variation of the scan statistics. The adjustment for the human population density was comparable to a stratification of data by a confounding variable as a way of correcting for the confounding effects in statistical and epidemiological analyses (Jolley et al., 1992; Szklo and Nieto, 2007). The multivariate multinomial model of the scan statistics test incorporates two datasets and simultaneously searches for clusters in both data sets (Kulldorff et al., 2007; Jung et al., 2010). The two datasets used in the analysis here were the four categories of invasions and the three categories of human population densities, as explained below.

Human population density data obtained from LandScan Global Population database does not contain values for areas such as waterbodies, however, the point locations used in the analysis were the centroids of waterbodies. When used the original Landscan data itself, 10,793/18,411(58.6%) of the point locations in our study did not acquire a population density value. Therefore, to obtain a continuous value across space, which facilitate to have human population density values for the centroids of the waterbodies, the LandScan data were subject to kernel density smoothing, using a Gaussian model with a smoothing factor of 0.3 and a radius of 0.2Km. The kernel smoothed human population density values were extracted for each waterbody. For Lake Superior, the shoreline population density of St. Louis County, 0.023 per square kilometer, was used. In the kernel smoothed data layer, the human population density varied between 4 to 1,410 per square kilometer, with a mean of 32.12 and a standard deviation of 107.04 people per square kilometer. The population density was categorized into three classifications, based on the geometric intervals: 1) low population density (<10 people per square kilometer), 2) intermediate population density (between 10 to 100 people per square kilometer), and 3) high population density (>100 people per square kilometer). Geometric intervals classification minimizes the sum of squares of the number of elements in each class while ensuring that the change between intervals is consistent, a geometric progression method where the coefficient was calculated by dividing the previous interval by the current interval (Protter and Protter, 1988; ESRI, 2007). Based on the kernel smoothed population density data, the number of waterbodies in each population density category were as follows: 1) low population density (n= 11,177), 2) intermediate population density (n=5,638), and 3) high

population density (n=1,595). The categorization of the human population density was done to measure the impact of different levels of human densities on the AIS reporting. This is because, when understanding the impact of population density, an increase of one person per square kilometer does not mean the same in less populated areas compared to the highly populated areas. Similar approaches and reasoning for categorization of continuous variables and the associated pros and cons are described elsewhere (Szklo and Nieto, 2007). Since we are using two variables stratified into categories, the analysis is called the multivariate multinomial model of the scan statistic test.

The multivariate multinomial model of the scan statistic test detects clusters regardless of the time and adjusts the data by population density. The maximum spatial window size of the cluster analysis was set to include up to 5% of the waterbodies at risk, which enables detection of smaller clusters. The invasion category (n=4) and the population density category (n=3) were the two variables used in the analysis. A parameter called the observed to expected ratio (O/E) was used to determine the likelihood ratio of the observed number of events within the recognized cluster compared to the expected number of events, under the null hypothesis of random occurrence of the events (Kulldorff et al., 2007; Jung et al., 2010). The multivariate multinomial scan statistics used here generates O/E values for each category of the two variables. Seven O/E values were calculated for each cluster, four and three for invasion and population density respectively. The category of each variable with the highest O/E were considered as the prominent categories for the clusters. For example, a cluster was considered to (a) have both zebra mussel and Eurasian water milfoil (EWM+/ZM+) invasions and (b) have a population density of >100 people per square Kilometer when the O/E values were highest for those two categories. In addition to the highest O/E, we have recorded the second highest categories with O/E >1 for better understanding of the characteristics of the clusters. The Monte Carlo based *p* value is determined to rank the significance of clusters among all data sets (Kulldorff et al., 2007; Jung et al., 2010). A detailed description of multivariate, multinomial scan statistics is available elsewhere (Kulldorff et al., 2007; Jung et al., 2010).

### ***Analysis of the directional trend***

The directional trend of waterbodies and the confirmed cases, was evaluated using the standard deviation ellipse (SDE) analysis for each AIS separately, using ArcMap 10.3 (ESRI, 2016). All the waterbodies (n=18,411) and the waterbodies invaded by zebra mussels (n=125) and Eurasian watermilfoil (n=304) were used as the cases in the directionality test. SDE for all

the waterbodies was computed to compare whether the observed directionality of the reported invasions is not merely representing the directionality where the waterbodies are located. The method calculates the standard deviation of the coordinates compared to the mean center of the invaded locations, providing the sense of overall orientation of the cases (Levefer, 1926; Wang et al., 2015). The analysis was conducted up to two standard deviations. Assuming the invasions follow a spatially normal distribution, the first and second standard deviations encompass approximately 68% and 95% of all input case locations (Levefer, 1926; Wang et al., 2015). The major and minor axes of the ellipse represent the magnitude of the minimum and maximum dispersal of the cases from their mean center (Gesler and Albert, 2000). The standard deviation ellipse provides a sense of the bidirectional trend of dispersal (Ex. northwest or southeast direction) rather than confirming possibility of a unidirectional trend.

In addition to the standard deviation ellipse, the spatiotemporal directionality of the AIS reporting was also analyzed to determine whether the confirmed cases have a unidirectional trend over time, using the Clusterseer v.2.05. software. The spatiotemporal directionality test uses a relative time connection matrix to recognize the average direction of the AIS dispersal over time. The average direction of confirmed invasions over time was estimated using a vector whose direction was the average direction of lines that connect each invasion with all subsequent invasions (Jacquez, 1996). The magnitude, i.e. the angular concentration, of the average direction of the invasion was the angular variance of those connecting lines (Jacquez, 1996). If the confirmed AIS cases followed a consistent spatiotemporal trend towards a given direction, then the angular variance is expected to be small while the angular concentration, a value between 0 and 1, would be large. The measure of angular concentration indicates whether the cases were reported one after the other in a similar direction or whether the cases were randomly reported from all directions. The resulting cluster characteristics and directionality information supports hypothesis generation and identifying possible risk factors associated with the AIS dispersal.



## 2.4 Results

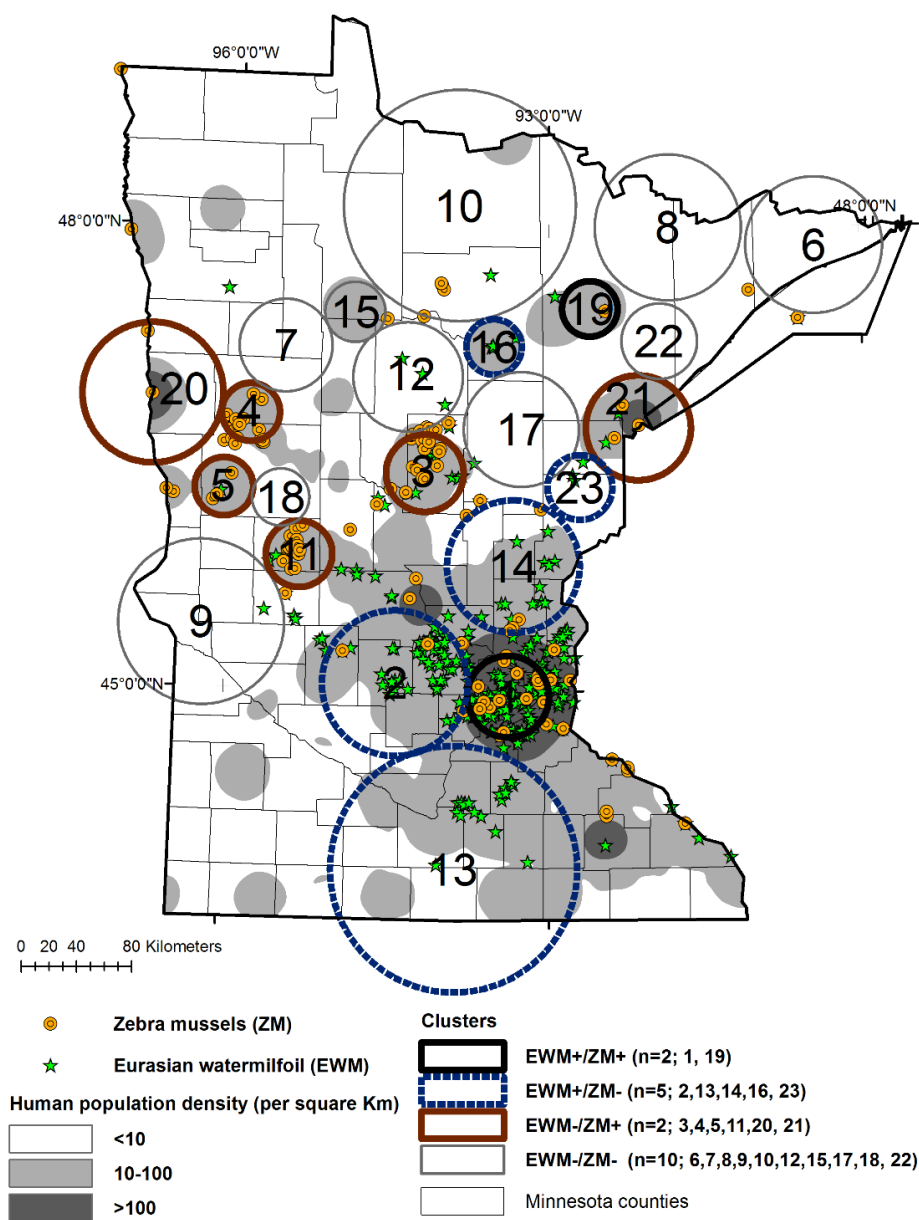
### *Cluster analysis*

Evidence of highly significant global clustering of both zebra mussel and Eurasian watermilfoil was detected at every assessed level of neighborhood (i.e. First through tenth neighborhood levels, at  $p < 0.05$ ), using the Cuzick Edwards test. The significant result in the Cuzick Edwards test confirmed that there is potential for clustering of both AIS on geographical space.

The multivariate multinomial model of the scan static test detected 23 highly significant clusters ( $p < 0.05$ ) (Table. 2.1 and Fig. 2.3.). These clusters included two that were simultaneously invaded with both zebra mussels and Eurasian watermilfoil (clusters #1 and #19); five that were invaded mainly by Eurasian watermilfoil (clusters #2, #13, #14, #16, and #23); six that were mainly invaded by zebra mussels (clusters #3, #4, #5, #11, #20, #21). Therefore, the sum of the clusters with either or both invasions was thirteen. The majority (11/13) of these clusters invaded by zebra mussels, Eurasian watermilfoil, or concurrent invasions were reported from areas with intermediate or high human population densities, except for the clusters #20 and #23 which were reported in an area with low human population density (i.e., <10 people per square kilometer). Ten of the 23 clusters were not reported with invasions by either of the two AIS (clusters #6, #7, #8, #9, #10, #12, #15, #17, #18, #22). Among the ten clusters that were not reported with zebra mussels or Eurasian watermilfoil (i.e., EWM-/ZM-), nine were in areas with low human population density, except for the cluster #15 which was in an area with intermediate human population density.

**Table 2.1.** Spatial clusters resulted from the multivariate multinomial model of scan static test for zebra mussels (ZM) and Eurasian watermilfoil (EWM) invasions, adjusted by the human population density ( $p < 0.05$ ). Observed to expected ratio (O/E<sup>#</sup>). Categories of population density (Pop. Den. Category. †) include: (1) <10; (2) between 10 to 100; and (3) >100 people per square kilometre (Km)\*.

Prominent Invasion Category	Cluster ID	Radius (Km)*	Waterbodies	O/E# of the Prominent Invasion Category	Secondary invasion category and O/E#	Prominent Pop. Den. Category
EWM+/ZM+	1	29.52	920	8.58	EWM+/ZM- (7.28)	3
	19	20.07	112	7.83		2
EWM+/ZM-	2	52.73	920	3.32	EWM-/ZM+ (1.71)	2
	13	89.35	591	1.98		2
	14	47.98	415	2.04		2
	16	20.08	191	1.36		2
	23	23.45	105	1.86		1
EWM-/ZM+	3	28.02	516	4.8	EWM+/ZM+ (1.7)	2
	4	21.00	446	4.76		2
	5	21.00	426	1.25		2
	11	24.05	374	7.57		2
	20	50.87	252	1.40		1
	21	37.98	149	3.56		2
EWM-/ZM-	6	49.43	920	1.02		1
	7	33.59	920	1.02		1
	8	52.82	920	1.02		1
	9	59.93	920	1.02		1
	10	83.78	920	1.02		1
	12	39.59	920	1.02		1
	15	21.77	210	1.02		2
	17	41.62	311	1.02		1
	18	20.71	308	1.02		1
	22	27.33	148	1.02		1



**Fig. 2.3** Clusters of zebra mussels (ZM) and Eurasian watermilfoil (EWM) detected by the purely spatial multivariate multinomial cluster analysis, stratified by the human population density of Minnesota, based on 2007 census data (Bright et al., 2013). The numbered circles represent the significant clusters stratified by the categories of human population density per square kilometer.

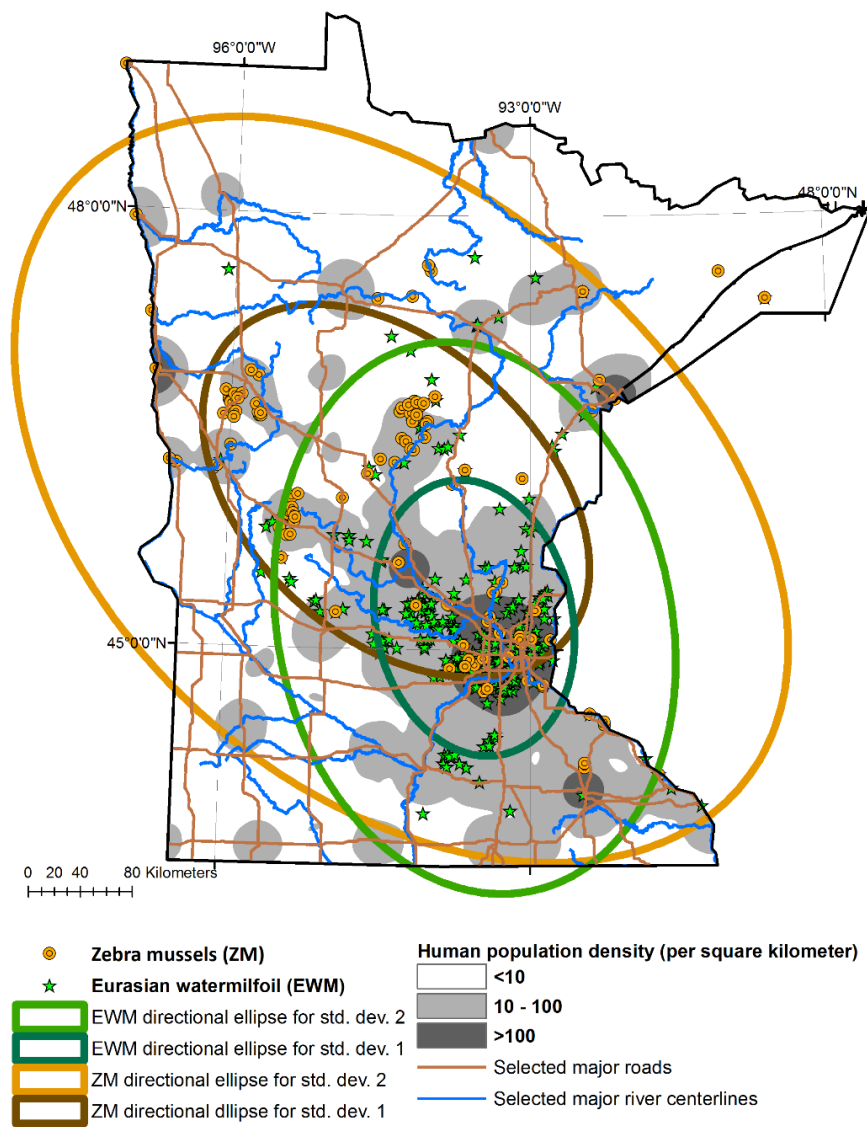
### ***Directionality***

There was a northwestern and southeastern bidirectional trend of cases with  $111^{\circ}$  and  $128^{\circ}$  rotations for the confirmed zebra mussels and Eurasian watermilfoil invasions, respectively (Table 2.2 and Fig. 2.4). The values for the two standard deviation ellipses represent the level of dispersion of each invasive species (Fig. 2.4, Table 2.2). The spatiotemporal directionality test indicated a northwestern unidirectional trend of detecting both zebra mussels and Eurasian watermilfoil invasions (Table 2.2). The average angle for the zebra mussel invasions was  $135^{\circ}$  with an angular concentration value of 0.33. The average angle for Eurasian watermilfoil was  $121^{\circ}$  with an angular concentration value of 0.11.

**Table 2.2.** Summary of the directionality tests for zebra mussels (ZM) and Eurasian watermilfoil (EWM).

Directionality test	Zebra mussel				Eurasian watermilfoil			
	Std. Dev.*	Major semi axis	Minor semi axis	Rotation	Std. Dev.*	Major semi axis	Minor semi axis	Rotation
Standard deviation ellipse (p <0.05)	Std. Dev. 1	2	1	111°	Std. Dev. 1	1	0.9	128°
	Std. Dev. 2	4	2.2	111°	Std. Dev. 2	2	1.8	128°
Spatiotemporal directionality trend (p <0.05)	Time measure	Average angle	Concentration		Time measure	Average angle	Concentration	
	Relative	135°	0.33		Relative	121°	0.11	

Std. Dev\* is the standard deviation of the coordinates compared to the mean center of the invaded locations which indicate the overall orientation of the confirmed invasions.



**Fig. 2.4.** Directionality of the zebra mussel and Eurasian watermilfoil reporting resulted from the standard deviation ellipse analysis showed a northwestern or southeastern trend with  $111^\circ$  and  $128^\circ$  rotations ZM and EWM invasions, respectively. Std. Dev. 1 and 2 refers to the first and second standard deviations encompass approximately 68% and 95% of all input case locations. The major roads and the major river centerlines are indicated in the map to facilitate recognition of potential means of dispersal.

## 2.5 Discussion

This study investigated the potential for human density to be a confounding variable on the dispersal patterns of the confirmed zebra mussel and Eurasian watermilfoil invasions in Minnesota. To achieve the study objectives, a multivariate multinomial model of the scan statistic test adjusting for human population density was used to identify significant spatial clusters of AIS in Minnesota waterbodies. The study also used spatial and spatiotemporal directionality tests to emphasize that the potential role of human movements along Interstate highway I-94, which connects major urban centers.

The results revealed evidence of global clustering of both zebra mussels and Eurasian watermilfoil, as detected by the Cuzick and Edwards test, suggesting that the invaded waterbodies are located closer to each other compared to the null hypothesis of spatial randomness. The significance of global clustering encouraged the search of spatial clusters locally and therefore, further analysis of the clustering pattern using multivariate multinomial analysis was performed. A total of 23 areas of statistically significant local clusters were detected ( $p < 0.05$ ).

Among the two clusters with concurrent invasions by zebra mussels and Eurasian watermilfoil, cluster #1, was located within the Minneapolis and St. Paul metropolitan area (Table 2.1 and Fig. 2.3), which is a major metropolitan area built around the Mississippi, Minnesota, and St. Croix rivers (major rivers are illustrated in Fig. 2.4). Cluster #1 had an observed to expected ratio 8.57 within 29.52km radius, indicating the area is highly likely to be reported with invasions conditional to the high population density in the area ( $>100$  people per square kilometer). The second concurrent invasion cluster, cluster #19, was detected in northern Minnesota in an area with intermediate human population density i.e. between 10 to 100 people per square kilometer. Although the human density in the surrounding areas is intermediate, the spatial extension of cluster #19 encompassed a major tourist destination near Mesabi Iron range (MNDNR: <http://www.dnr.state.mn.us/geologyrec/index.html>. Cited November 18, 2017), with proximity to St. Louis and Vermillion river drainages and major roads (Fig. 2.3 and Fig. 2.4), which would facilitate frequent visitations by boaters.

There were five significant clusters with prominent Eurasian watermilfoil invasions (Table 2.1 and Fig. 2.3). Four out of the five clusters were detected in areas with intermediate human population density and the other one, cluster #23, was detected in an area with low population density. However, cluster #23 was located in between the Minneapolis/St. Paul metropolitan area and Duluth, which are highly populated areas. The spatial extension of cluster #23 was in

proximity to the St. Croix river drainage, Moose lake state park, and ample road access for tourists (Fig. 2.4), which facilitate human-mediated spread of invasions as well as increased opportunity of public reporting (Banha et al., 2016).

Another six significant spatial clusters included zebra mussel invasions and shared similarities with the clusters mentioned above, including 1) their spatial extension encompassed popular recreational waterbodies (MNDNR: <http://www.dnr.state.mn.us/geologyrec/index.html>. Cited November 18, 2017), 2) five out of the six clusters with zebra mussel invasions were detected in areas with intermediate human population density, 3) invaded waterbodies, encompassed by cluster #6 are proximal to major roads including Interstate highway I-94, and major river centerlines. As seen in column six of Table 1.1, the invasion category of secondary importance in clusters #3 and #11 was the concurrent invasions (i.e., EWM+/ZM+), which indicate that these locations were at risk of both invasions compared to the other areas (Table 1.1). Cluster #20 with EWM-/ZM+ was detected in an area where generally population density is low. However, as seen in Fig 2.3, the cluster #20 captures Moorhead, Minnesota and Fargo, North Dakota which is a highly populated area. Therefore, the significance of cluster #20 is explained by the high population density in the middle of the cluster, which may have contributed to both dispersal and detection of AIS.

The standard deviation ellipse analysis indicated both AIS have a northwestern and southeastern bidirectional trend (Table 2.2 and Fig. 2.4). Interestingly, the standard deviation ellipse for zebra mussels seems to center along interstate highway I-94, which suggests the convenience of road access may be a useful factor when searching for potential underreported invasions (Fig. 2.4). Admittedly, within the limits of this ecological study, we are suggesting the potential association of the major road access rather than the causation of the road access to increase the reporting of invasions. Because the objective was to measure the trend of all the reporting points, analytical and visualization methods such as rose diagrams (Mackanness, 1991), which would capture the directionality of reporting compared to the first reported invasion, were not employed. The spatiotemporal directionality test was used to confirm the northwestern unidirectional trend of reporting zebra mussel and Eurasian watermilfoil invasions, eliminating the ambiguity of bidirectionality. However, the angular concentration resulting from the spatiotemporal directionality test for zebra mussels and Eurasian watermilfoil were 0.33 and 0.11, respectively (Table 2.2), which suggests that the consistency of the northwestern directionality is less for Eurasian watermilfoil compared to zebra mussels, even though the directionality is statistically significant for both AIS ( $p < 0.05$ ). The inconsistency of the directionality is explained



by invasions reported in both northeastern and southwestern directions in subsequent years. However, the fact that there is no active surveillance and the public reporting of invasions is subjected to delays and fluctuations over the years should not be overlooked when interpreting our spatiotemporal directionality test results. This observation of common directionality of dispersal of both zebra mussels and Eurasian watermilfoil further suggests presence of common drivers for AIS spread as well as reporting.

One of the limitations of the study is representation of waterbodies with centroids regardless of the size, which may lead to misrepresentation of waterbodies. Large lakes may have been underrepresented because one centroid represents the whole lake. On the contrary, invaded long rivers may have been overrepresented because a point location was assigned to the centroid of the river at each county. However, use of point locations to represent waterbodies was inevitable because, spatial statistics such as cluster analysis and the directionality tests are meant for point pattern data rather than polygons. Another limitation is that the categorization of the population density was done using geometric interval classification, which is only one of the many methods of data categorization. Inadequate categorization of a confounding variable could result in residual confounding, a problem that could result in imperfect measurement of the exposure and output ([Szklo and Nieto, 2007](#)). As mentioned in the introduction, the exposure of interest here is the presence of invaded waterbodies in the area and the output is new invasions in neighboring waterbodies. We consider that there is no residual confounding, because majority of the invasion clusters are in the areas with intermediate and high population densities.

Recognizing potential confounding by human density is important when analyzing or modeling invasions based on public reporting rather than surveillance as results can be misleading unless considerations are made for the reporting bias ([Aikio et al. 2010](#)). If the cluster analysis was conducted without human density, the clustering pattern shared by the two datasets, i.e. confirmed invasions and human density, would have been overlooked ([Kulldorff et al., 2007](#)). This is because the resulting clusters would capture misleading geographical areas that might not reflect the impact of human density on reporting of AIS. High human population density can also result in high boater traffic between waterbodies, which can lead to an increase in human mediated dispersal of the invasive species ([Banha et al., 2016](#)).

When planning invasion control measures, managers must have accurate assessments of risk. This is challenging when the actual invasion status is unknown and confounding effects are present. This study shows this importance of taking human population density into account, given that human population density would lead to overrepresentation of invasion in highly populated

areas and underrepresentation in less populated areas. Although human mediated dispersal is one of the key risk factors, there are other contributing factors leading to the dispersal of invasions, such as connectivity via streams or rivers and distance to other invaded locations leading to isolated events of translocation by animals or birds (Banha et al., 2016). Therefore, it is ideal to stratify the magnitude of active and passive surveillance efforts by the human population densities. This can be achieved by conducting active surveillance in less populated areas at regular intervals, such as annual or biannual surveys, while continuing to encourage people to report invasions through public awareness programs (Lodge et al., 2006). This combination of passive and targeted surveillance may help provide insight into the true invasion status and its progression in the Great Lakes region.

## **2.6 Conclusions**

This study provided insights into the spatial characteristics of zebra mussels and Eurasian watermilfoil reporting across waterbodies in Minnesota. Human population density was identified as a potential confounding factor with a strong role in shaping clustering and reporting patterns of the two AIS. Twenty-three significant spatial clusters ( $p$  value 0.05) were detected. Most of the clusters with zebra mussels, Eurasian watermilfoil, or concurrent invasions were reported in waterbodies with proximity to both highly and intermediately populated areas. Whereas, most clusters without the invasions were detected from less populated areas. Because highly populated areas may lead to overrepresentation of invasions, while invasions may have been underrepresented in less populated areas, we suggest conducting a combination of active and passive surveillance stratified by the human population densities. This will provide more insight into the true invasion status and its progression in the Great Lakes region.

## **2.7 Acknowledgements**

This study was funded in part by the Minnesota Discovery, Research, and Innovation Economy (MnDRIVE) program and Office of the Vice President for Research (OVPR) of the University of Minnesota. We thank the Aquatic Invasive Species Prevention program of the Minnesota Department of Natural Resources for providing the incident data and sharing their experience.

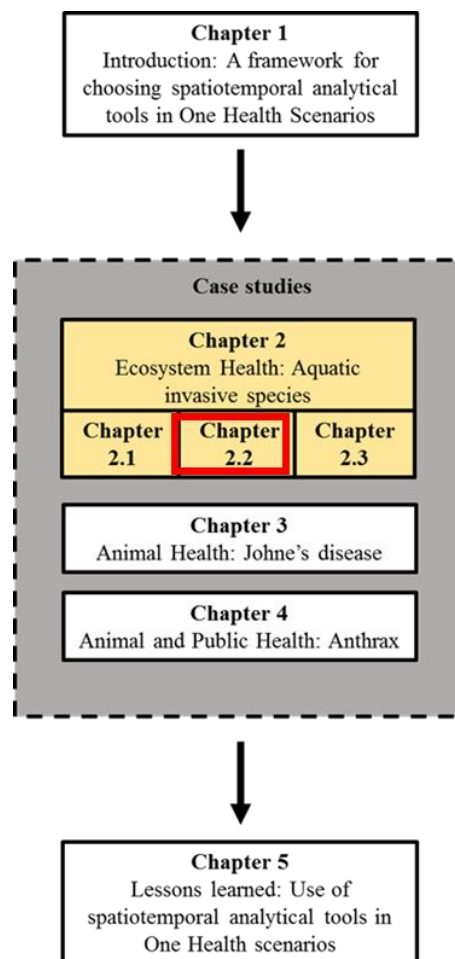
## 2.8 References

1. Aikio S, Duncan RP, Hulme PE. 2010. Herbarium records identify the role of long-distance spread in the spatial distribution of alien plants in New Zealand. *J. Biogeogr.* 37, 1740-1751.
2. Banha F, Gimeno I, Lanao M, Touya V, Duran C, Peribanez MA, et al. 2016. The role of waterfowl and fishing gear on zebra mussel larvae dispersal. *Biol. Invasions.* 18, 115-125.
3. Bright EA, Rose AN, Urban ML. 2013. LandScan 2012™. High Resolution Global Population Data. UT-Battelle, LLC. Oak Ridge National Laboratory. Contract No. DE-AC05-00OR22725. U.S. Department of Energy. <http://www.ornl.gov/landscan/>.
4. Brown JE, Stepien CA. 2010. Population genetic history of the dreissenid mussel invasions: expansion patterns across North America. *Biol. Invasions.* 12, 3687-3710.
5. Carlton JT. 2008. The zebra mussel *Dreissena polymorpha* found in North America in 1986 and 1987. *J. Great Lakes Res.* 34, 770–773.
6. Cheruvilil KS, Soranno PA, Madsen JD, Roberson MJ. 2002. Plant architecture and epiphytic macroinvertebrate communities: the role of an exotic dissected macrophyte. *J. N. Amer. Benthol. Soc.* 21, 261-277.
7. Connelly NA, O'Neill CR, Knuth BA, Brown TL. 2007. Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. *Environ. Manag.* 40, 105-112.
8. Cuzick J, Edwards R. 1990. Spatial clustering for inhomogeneous populations. *J. R. Stat. Soc. B Stat. Methodol.* 52, 73-104.
9. Elliott P, Wakefield J. 2001. Disease clusters: should they be investigated, and, if so, when and how? *J. R. Stat. Soc. A Stat. Soc.* 164, 3-12.
10. ESRI. 2016. Environmental System Research Institute. ArcGIS Desktop: Release 10.3.1. Redlands, CA, USA.

11. ESRI. 2007. Environmental System Research Institute. ArcGIS Desktop: Geometric interval. Available online at:  
[http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?topicname=geometrical\\_interval](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?topicname=geometrical_interval).
12. Gesler WM, Albert DP. 2000. How spatial analysis can be used in medical geography. In: Albert D.P., Gesler W.M., Levergood B. (Eds.), *Spatial Analysis, GIS and remote sensing applications in the health sciences*. Ann Arbor Press, Chelsea, MI, pp.11-38.
13. Invasive Species Program. 2016. *Invasive Species of Aquatic Plants and Wild Animals in Minnesota; Annual Report for 2015*, Minnesota Department of Natural Resources, St. Paul, MN. Available online at: [http://files.dnr.state.mn.us/natural\\_resources/invasives/ais-annual-report.pdf](http://files.dnr.state.mn.us/natural_resources/invasives/ais-annual-report.pdf).
14. Jacquez GM. 1996. Statistical software for the clustering of health events. *Stat. Med.* 15, 951-952.
15. Jacquez GM, Estberg L, Long A, Do E, Rommel B, Durbeck H, Greiling DA et al. 2012. Clusterseer v.2.05. *ClusterSeer User Manual 2: Software for the detection and analysis of event clusters*. Biomedware, Inc. Ann Arbor, MI.
16. Jolley DJ, Jerman B, Elliot P. 1992. Socio-economic confounding, in *Geographical and Environmental Epidemiology*. In: Elliot, P., Cuzick, J., English, D., Stern, R. (Eds.), *Methods for small-area studies*. Oxford University Press, Oxford, pp. 3-13.
17. Jung I, Kulldorff M, Richard OJ. 2010. A spatial scan statistics for multinomial data. *Statistic. Med.* 29, 1910–1918.
18. Karatayev AY, Burlakova LE, Mastitsky SE, Padilla DK. 2015. Predicting the spread of aquatic invaders: insight from 200 years of invasion by zebra mussels. *Ecol. Appl.* 25, 430-440.
19. Kulldorff M. 2009. SaTScanTM v9.4.1: Software for the spatial and space-time scan statistics. <http://www.satscan.org/>.
20. Kulldorff M, Mostashari F, Duczmal L, Katherine YW, Kleinman K, Platt R. 2007. Multivariate scan statistics for disease surveillance. *Stat. Med.* 26, 1824-1833.

21. Kulldorff M, Nagarwalla N. 1995. Spatial disease clusters and inference. *Stat. Med.* 14, 799-810.
22. Les DH, Mehrhoff L. 1999. Introduction of nonindigenous aquatic vascular plants in southern New England: A historical perspective. *Biol. Invasions* 1, 281–300.
23. Levefer DW. 1926. Measuring geographic concentration by means of the standard deviation ellipse. *Am. J. Sociol.* 32, 88-94.
24. Lodge DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA, Carlton JT, McMichael A et al. 2006. Biological invasions: Recommendations for US policy and management. *Ecol. Appl.* 16, 2035-2054.
25. Mackaness WA. 1991. Integration and Evaluation of Map Generalization. In: Buttenfield, B., McMaster, R.B. (Eds.), *Map Generalization: Making rules for knowledge representation*. . Longman, United Kingdom, pp. 217-226.
26. MNGSC Hydro: Minnesota GeoSpatial commons. 2015. MNDNR Hydrography data layer. [cited August 15, 2015]. Available online at: <https://gisdata.mn.gov/dataset/water-dnr-hydrography>.
27. MNGSC Roads: Minnesota GeoSpatial commons. 2012. Roads, Minnesota 2012 data layer. Published by Minnesota Department of Transportation (MnDOT). [cited: May 15, 2016]. Available online at: <https://gisdata.mn.gov/dataset/trans-roads-mndot-tis>. Last updated: 02/25/2016.
28. Pysek P, Richardson DM. 2010. Invasive Species, Environmental Change and Management, and Health. Pages 25-55 In: Gadgil, A., Liverman, D.M., (Eds.), *Annu. Rev. Environ. Resour.* 35, 25-55.
29. Protter MH, Protter PE. 1988. Sequences. In: *Calculus with analytic geometry*. 4th Ed. Jones and Bartlett Publishers, Inc. Boston, MA. pp 386-391.
30. Roley SS, Newman RM. 2008. Predicting Eurasian watermilfoil invasions in Minnesota. *Lake Reserv. Manag.* 24, 361-369.

31. Szklo M, Nieto FA. 2007. *Epidemiology Beyond the basics*. 3rd Ed. Jones and Bartlett Learning, LLC. Burlington, MA. pp 153-305.
32. Wang B, Shi W, Miao Z. 2015. Confidence analysis of standard deviational ellipse and its extension into higher dimensional Euclidean space. *PLOS ONE* 10, DOI: 10.1371/journal.pone.0118537.
33. Ward MP, Carpenter TE. 2000. Analysis of time-space clustering in veterinary epidemiology. *Prev. Vet. Med.* 43, 225-237.



## Chapter 2.2

***A probability co-kriging model to account for reporting bias and recognize areas at high risk for Zebra Mussels and Eurasian Watermilfoil invasions in Minnesota***

**Published in Frontiers in Veterinary Science (2018):**

Kanankege KST, Alkhamis MA, Phelps NBD, Perez AM. 2018. A probability co-kriging model to account for reporting bias and recognize areas at high risk for zebra mussels and Eurasian watermilfoil invasions in Minnesota. Frontiers in Veterinary Science. doi: 10.3389/fvets.2017.00231

## 2.9 Overview

Zebra mussels (*Dreissena polymorpha*) and Eurasian watermilfoil (*Myriophyllum spicatum*) are aggressive aquatic invasive species posing a conservation burden on Minnesota. Recognizing areas at high risk for invasion is a prerequisite for the implementation of risk-based prevention and mitigation management strategies. The early detection of invasion has been challenging, due in part to the imperfect observation process of invasions including the absence of a surveillance program, reliance on public reporting, and limited resource availability, which results in reporting bias.

To predict the areas at high risk for invasions, while accounting for underreporting, we combined network analysis and probability co-kriging to estimate the risk of zebra mussel and Eurasian watermilfoil invasions. We used network analysis to generate a waterbody-specific variable representing boater traffic, a known high risk activity for human-mediated transportation of invasive species. In addition, co-kriging was used to estimate the probability of species introduction, using waterbody specific variables.

A co-kriging model containing distance to the nearest zebra mussel infested location, boater traffic, and road access was used to recognize the areas at high risk for zebra mussel invasions (AUC=0.78). The Eurasian watermilfoil co-kriging model included distance to the nearest Eurasian watermilfoil infested location, boater traffic, and connectivity to infested waterbodies (AUC=0.76). Results suggested that, by 2016, nearly 20% of the waterbodies in Minnesota were at high risk of zebra mussel (12.45%) or Eurasian watermilfoil (12.43%) invasions, whereas only 125/18,411 (0.67%) and 304/18,411 (1.65%) are currently infested, respectively. Prediction methods presented here can support decisions related to solving the problems of imperfect detection, which subsequently improve the early detection of biological invasions.



## 2.10 Introduction

Aquatic invasive species (AIS) have the potential to affect animal, environmental, and public health (Pysek and Richardson, 2010; Havel *et al.*, 2015). The state of Minnesota in the United States has experienced numerous AIS incursions, and spend over ten million dollars each year on activities intended to prevent, control or manage AIS (Invasive Species Program 2016; MN Statute 477A.18, 2016).

Zebra mussels (*Dreissena polymorpha*) and Eurasian watermilfoil (*Myriophyllum spicatum*) are AIS of concern for Minnesota and have been reported in Minnesota since 1989 and 1987, respectively (MNDNR AIS, 2016). The first introduction of zebra mussels into North America is attributable to ballast water from transatlantic ships (Mills, 1993). Zebra mussels are rapidly propagating bivalves that disrupt the stability of the food web in aquatic ecosystems affecting both pelagic and benthic species (Karatayev *et al.*, 2015). Removal of zebra mussels colonizing public water supply pipes and pipes of industrial facilities has cost nearly \$267 million in the zebra mussel affected region in North America between 1989 to 2004 period (Connelly *et al.*, 2007). Similarly, Eurasian watermilfoil, an invasive aquatic macrophyte, was likely introduced into North America through aquarium trade (Mills, 1993). Eurasian watermilfoil proliferates rapidly impeding the effective removal or control strategies upon establishment in a waterbody (Roley and Newman, 2008). Dense vegetation of Eurasian watermilfoil outcompetes native macrophytes and interrupts recreational activities (Roley and Newman, 2008). An intensive hand harvesting project to control Eurasian watermilfoil, conducted in the upper Saranac Lake in New York, reported a labor cost of \$351,748/year in that one lake alone (Kelting and Laxon, 2010).

Aggressive and costly programs have been implemented in Minnesota to control AIS (Invasive Species Program, 2016). For example, since 2014, \$10 million per year has been allocated by the Minnesota legislature to provide resources for county-based AIS prevention activities, such as education, surveys, and watercraft inspections (MN Statute 477A.18, 2016). However, because the risk of AIS invasion had not been previously quantified, the resources were distributed proportionally to the share of boat ramps and trailer parking spaces in each county (MN Statute 477A.18, 2016). The funds are invested on prevention of the introduction or limitation of the spread of AIS within the county (Invasive Species Program, 2016, 2016; MN Statute 477A.18, 2016). Because of the high economic and conservation burden posed by the invasions, forecasting of the areas at high risk for invasions is an urgent research priority (Havel *et al.*, 2015).

The two AIS have been invading Minnesota waters for approximately 30 years, therefore, the measurement of propagule pressure, i.e., the ‘introduction effort’, needs to be focused at the local scale such as at individual waterbody (Simberloff, 2009). As a solution, previous studies have suggested using surrogate variables such as the number of boat ramps and distance to the major roads in the absence of waterbody specific data when measuring the propagule pressure (Buchan and Padilla., 2000). One of the most challenging waterbody-specific variables is the measurement of human mediated dispersal (Buchan and Padilla., 2000; Roley and Newman, 2008; Banha *et al.*, 2016). Use of human population density as a proxy for the human mediated dispersal may serve as a solution. However, densely populated areas may also tend to report the invasions more frequently, compared to less populated areas (Aikio *et al.*, 2010; Kanankege *et al.*, 2018), which may also lead to reporting bias and underreporting.

The objective of this study was to estimate the potential range expansion of zebra mussels and Eurasian watermilfoil in Minnesota, using a combination of network analysis and co-kriging, a spatial interpolation technique to account for under-reporting. The advantage of using co-kriging is that the technique enables the prediction of values for the locations without observed data, using other correlated and highly sampled variables (Isaaks and Srivastava, 1989; Rogers and Sedda, 2012). Co-kriging is commonly used in gold mining and lake and reservoir studies, and has rarely been used in veterinary epidemiological and public health studies as well (Vauclin *et al.*, 1983; Oliver *et al.*, 1998; Perez *et al.*, 2006; Martine *et. al.*, 2008). Environmental conservation studies, such as the controlling the spread of invasions, often suffer from lack of data and reporting bias because of the financial constraints on surveillance (Pysek and Richardson, 2010). In Minnesota, invasions are often reported by volunteers and the presence of the AIS may be missed in some waterbodies due to insufficient coverage, which decreases the sensitivity of the reporting. The specificity of the reporting system, instead, may be considered acceptable, given that false positive cases are unexpected. False positives are unlikely because, the Minnesota Department of Natural Resources (MNDNR) confirms newly reported invasions prior to adding them to the official online database of infested waters (MNDNR AIS, 2016). Consequently, the limitation of this passive surveillance system is the potential underreporting of the conditions. Co-kriging may also compensate for the reporting bias and underreporting by augmenting the predictive power of one variable with the support of other correlated and highly sampled variables.

Recognition of areas at high risk may act as an early warning system and help the prioritization of waterbodies for a targeted and efficient allocation of limited resources to improve

both defensive and offensive management strategies (Mandrak *et al.*, 2015; Stewart Koster *et al.*, 2015). Such risk targeted approaches certainly represent improvements over the random selection of waterbodies for surveillance and management purposes (Lodge *et al.*, 2006; Vander Zanden *et al.*, 2008). For example, current guidelines for conducting AIS early detection and baseline monitoring in lakes of Minnesota suggest that volunteers select waterbodies based on factors such as; public water access, boater traffic, tourist activity etc. (Lund *et al.*, 2015). However, selecting waterbodies based on multiple criteria is challenging and we propose that a method which take all the most relevant risk factors into account and provide a risk rank would be a better fit to guide the volunteers. Study results may inform risk-based surveillance and management of invasions (Lodge *et al.*, 2006; Mandrak and Cudmore, 2015), a process defined as making decisions for identifying, evaluating, selecting, prioritizing, and implementing control measures (CRARM, 1997). This work demonstrates the use of analytical models to estimate risk while accounting for reporting bias, with the ultimate objective of evaluating and modifying the policies and practices on biological invasions (Lodge *et al.*, 2006).

## 2.11 Data and Methods

### 2.11.1 Data

#### *Study area and AIS-presence data*

A total of 18,411 point locations representing waterbodies of Minnesota were considered as the study population in this study. Waterbodies were mainly lakes and ponds (n=18,263) and were represented by the centroids of each waterbody. In addition to the lakes, several riverine locations (n=148) from major rivers were included in the analysis. Riverine locations were identified at the rivers' midpoint within each county. The locational data for the waterbodies were extracted from the GIS layer referred to as "MNDNR Hydrography," which is available from the Minnesota GIS Commons (MNGSC Hydrography, 2015). Presence data for confirmed AIS locations were collected from the MNDNR database (MNDNR AIS, 2016). By the end of 2015, there were 125/18,411 (0.67%) zebra mussels and 304/18,411 (1.65%) Eurasian watermilfoil infested waterbodies in Minnesota (MNDNR AIS, 2016; Kanankege *et al.*, 2018). The confirmed presence of the AIS was used in the study regardless of the magnitude of infestation, because assessments on the magnitude of infestation are not available.

### *Waterbody specific variables*

Waterbody specific variables (n=6), were used as predictors in the co-kriging models. The six waterbody specific variables included (1) zebra mussels or (2) Eurasian watermilfoil invaded waterbody, (3) connectivity to another zebra mussel and (4) Eurasian watermilfoil invaded waterbody via a stream or a river, (5) boater traffic between waterbodies, and (6) inverse of the Euclidean distance to the nearest major road. Status of the invasions, i.e. confirmed presence of invasion was the primary variable for each AIS (variables 1 and 2). For the validation purposes models were fit for years 2010 and 2015, therefore, two sets of each variable were calculated. The number of waterbodies from which each variable is available varied over the time (Table 1.3). However, the same boater traffic variable was used in both 2010 and 2015 model fits because boater traffic was calculated based on a survey conducted in 2013, as described below. The Euclidean distance to the nearest major road variable was the same for both 2010 and 2015 assuming the major roads remained unchanged.

**Table 2.3.** Number of waterbodies with the characteristic of each variable by 2010 and 2015.

	<b>Number of waterbodies by 2010</b>	<b>Number of waterbodies by 2015</b>
<b>1</b> ZM invasion status *	57	125
<b>2</b> EWM invasion status *	251	304
<b>3</b> Connectivity to another ZM invaded waterbody via a river or a stream <sup>#</sup>	2,392	3,658
<b>4</b> Connectivity to another EWM invaded waterbody via a river or a stream <sup>#</sup>	3,129	3,715
<b>5</b> Eigen vector centrality of the boater traffic network	1,376	1,376
<b>6</b> Inverse of the Euclidean distance to the nearest major road	18,411	18,411

\*Presence only. <sup>#</sup>Connected waterbodies only.

Proximity and connectivity to infested waterbodies have been recognized as key risk factors leading to zebra mussel and Eurasian watermilfoil invasions (Bobeldyk *et al.*, 2005; Roley and Newman, 2008). Because of the pairwise distance calculation for the semi-variance of candidate variables in the model, the kriging process includes the distance between locations as an integral part of the algorithm (Isaaks and Srivastava, 1989). Therefore, when AIS

presence/absence is the primary variable, the spatial dependence, i.e., the distance to the nearest infested location is inherently included in the co-kriging model.

Surface water connectivity between waterbodies via a stream or a river was obtained by intersecting the map of the river and streamlines features with the polygon features representing lakes, ponds, and reservoirs using ArcGIS version 10.3.1 (ESRI, 2016). River and streamline feature data were obtained from the “Stream Routes with Kittle Numbers and Mile Measures” GIS layer available from the Minnesota GIS Commons (MNGSC Stream, 2015). Several published studies identified the potential for downstream (e.g. via downstream drift) and upstream (e.g. via watercraft) spread of zebra mussels and Eurasian watermilfoil (Bobeldyk *et al.*, 2005; Spencer and Carruthers, 2012; Osawa *et al.*, 2013). However, the distance measures denoting the extent of the spread upstream or downstream were either not studied or varied among the published literature. Therefore, for simplicity, an invasion was assumed to occur both up and down stream regardless of the flow direction. Invaded locations that were not directly intersecting a river or streamline were given a buffer distance of 100m around the point location, and the closest river or stream feature was assigned as connected because the proximity to the infested location poses the risk of invasion (Roley and Newman, 2008; Karatayev *et al.*, 2015). Rivers and streams were represented by a unique identification number referred to as “Kittle Numbers” assigned by the MNDNR (Fisheries Stream Survey Manual, 2007; MNGSC Stream, 2015). Kittle numbers consisted of an alphabetical letter, followed by a string of digits (Fisheries Stream Survey Manual, 2007). For example, if an invaded waterbody was connected to kittle number #H026, then any waterbody connected to #H026 was assigned as connected to an invaded waterbody. Connectivity networks were generated separately for zebra mussels and Eurasian watermilfoil.

Boater traffic between waterbodies may lead to human mediated dispersal of AIS (Roley and Newman, 2008; Banha *et al.*, 2016). Here, boater traffic was measured using data collected by the MNDNR Watercraft Inspection Program, a survey conducted since 1992 as a conservation measure to protect state waters (MNDNR WIP, 2014). The Watercraft Inspection Program survey is conducted at selected waterbodies. Priority for data collection is given to those that are invaded, located near an invaded waterbody, highly used, or located close to popular travel destinations (MNDNR WIP, 2014). The boaters who visit the waterbodies were interviewed regarding the previous waterbody visited and the waterbody they plan to visit next. In 2013, the Watercraft Inspection Program surveys were conducted at 240 locations, and 119 (49.6%) of those locations were invaded by either zebra mussels or Eurasian watermilfoil. Because of the

miscellaneous reporting errors, only 21% of the surveys were eligible to be used in the final Watercraft Movement Network. Based on the survey, boater traffic data were available from 1,376 unique waterbodies (7.5% of the total waterbodies). Because the analysis was focused on predicting the current risk of invasions rather than understanding the impact of boater traffic on past invasions, it was assumed that movements recorded in 2013 were representative of movement patterns observed between 1987 and 2015.

Network analysis which provides a framework to identify units that are frequently or intensely connected within the network and identify contact patterns ([Martinez Lopez \*et al.\*, 2009](#)), was applied to the Watercraft Inspection Program data from 2013. A total of 187,074 surveys were conducted between April 25<sup>th</sup> 2013 and November 30<sup>th</sup> 2013. Recreational boater movement data are not collected during the winter season ([MNDNR WIP, 2014](#)). In the analysis, network “nodes” were the waterbodies and visits between waterbodies served as “edges.” Each completed survey accounted for two edges, representing the following links: 1) between the previously visited location and the surveyed location, and 2) between surveyed location and the next stated location that the watercraft would visit. Three centrality measures, namely, the Eigenvector, Betweenness, and Degree were calculated for the network. The centrality measure that highly correlates with the status of the invasions by zebra mussel and Eurasian watermilfoil was chosen, upon calculating the Pearson correlation analysis. Eigenvector centrality was chosen as the network parameter representing the connectivity of each waterbody within the watercraft movement network. Eigenvector centrality is a representation of the relative importance of a node regarding its position and connectivity to other highly connected nodes in the network ([Martinez Lopez \*et al.\*, 2009](#)). It was assumed that highly connected nodes could play a major role in distributing AIS.

Distance to the nearest major road represents the convenience of accessibility to a waterbody. Boater traffic data are collected from limited waterbodies; however, an indirect measure of the potential visitations is the calculation of road accessibility ([Buchan and Padilla, 2000](#); [Gallardo, 2014](#)). Therefore, distance to the nearest major road from the waterbodies was calculated using the major roads map of 2012, available through the Minnesota Geospatial Commons and originated from the Department of Transportation ([MNGSC Roads, 2012](#)). As defined in the metadata of the spatial layer, road classes including interstate highways, freeways, arterials, and major collectors were considered as major roads in the analysis ([MNGSC Roads, 2012](#)). The inverse of the Euclidean distance was used as the variable when fitting the models.

### 2.11.2 Methods

#### *Co-kriging to estimate the probability of introduction*

Probability co-kriging was used to estimate the probability of zebra mussel or Eurasian watermilfoil introduction into the waterbodies, conditional to the distance between locations and other waterbody specific variables. Co-kriging is a linear weighted averaging method in which weights are selected to minimize the variance of the estimation error by accounting for the spatial correlation between the waterbody specific variables; weights are dependent on the distance between sampled locations (Isaaks and Srivastava, 1989). In this study, multiple correlated waterbody specific variables were used to estimate the spatial distribution of the dependent variable in the non-sampled locations (Isaaks and Srivastava, 1989). The primary variable subjected to co-kriging is the invasion status of zebra mussels or Eurasian watermilfoil. Therefore, the “sampled locations” were those confirmed to be infested, whereas “not sampled locations” were those that without infestation reports. The cross correlation between variables is used to improve the predictions because the predictions are derived from both primary and secondary variables (Isaaks and Srivastava, 1989). A complete description of the application of co-kriging is available elsewhere (Isaaks and Srivastava, 1989; Perez *et al.*, 2006).

Pearson correlation coefficient was calculated to determine the correlation between the six waterbody specific variables. Variables with a correlation coefficient  $\geq 0.1$  were selected to be included in the co-kriging models. Multiple co-kriging models were fit for both zebra mussels and Eurasian watermilfoil separately. Each model included the primary variable i.e. the status of the invasion and two correlated variables. All possible two-way combinations were fit. Considering the potential mutualism between zebra mussel and Eurasian watermilfoil suggested by multiple studies (MacIsaac *et al.*, 1996; Ricciardi, 2001), the variable pairing also included the use of invasion status of zebra mussels as a correlated variable used in co-kriging model to predict Eurasian milfoil and vice versa. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), a plot of model sensitivity (true positives) and 1-specificity (i.e., false positives) (Swets, 1988). AUC values lower than 0.7 are considered relatively inaccurate because the proportion of false and true positive results is not substantially different, whereas AUC values greater than 0.7 are generally considered appropriate (Swets, 1988). Models with AUC value greater than 0.7 were considered accurate in this study.

The variables contributing to the co-kriging model with highest AUC were chosen. Hence, final models consisted of the primary variable representing the invasion status of each



AIS and two other waterbody specific variables. AUC values were calculated for each of the co-kriging models by true validation, which was done by fitting models to the invasions by 2010 and validating using the invasions reported between 2011 and 2015. Results of the co-kriging analysis were the probability of finding an AIS invaded waterbody conditional to the presence of an invaded location in the proximity and the waterbody specific variables incorporated into the model. Small lag sizes (Ex. 0.04 km) and few lags (Ex.  $n=12$ ) were used in the computation of the co-kriging semivariogram. The use of small lag size and few lags was intended to reduce the exponential increase in the influence of an infested location to the nearby cells, i.e., to reduce the effect of high spatial autocorrelation (Isaaks and Srivastava, 1989). The choice of the parameter values for the co-kriging attributes such as the anisotropy factor and the angle were based on the spatial cluster analysis and directionality tests for the data (Kanankege *et al.*, 2018). The parameter values are summarized in the supplement document (Supplementary Table S1).

The performance of the final co-kriging models for zebra mussels and Eurasian watermilfoil was estimated based on the predictive powers of the candidate models. The predictive powers were measured estimating the sensitivity and specificity, and the AUC of the candidate models. In the context here, sensitivity and specificity reflect the ability of the model to predict invaded and not invaded waterbodies, respectively. Because the goal of the model was to predict potential infestations, high sensitivity, rather than high specificity, was targeted when optimizing the models. In addition to the true validation, the co-kriging models were cross validated using k fold cross validation ( $k=5$ ). Cross validation is a process in which a set of AIS infested locations were left out from the model fitting, and the fitted model output was used to estimate the probability of finding an AIS invasion at those left out locations (Fielding and Bell, 1997). Eighty percent of the cases were used for the model training, and testing was done using the 20% of the withheld cases for each validation. To maintain the consistency, the co-kriging parameters recognized during the true validation were used when fitting the models for the cross validation.

### ***Interpretation of the co-kriging outputs***

Predicted probabilities were extracted for each of the waterbodies from the probability output of the co-kriging models, for zebra mussels and Eurasian watermilfoil separately. The outputs were ranked into five 'risk rank' categories based on the quantiles of the output probability values. The risk ranks 1 through 5 were defined as: 5) very high, 4) high, 3)



intermediate, 2) low, and 1) negligible risk of AIS introduction. The co-kriging risk rank resulting with highest sensitivity and specificity was considered the threshold for each model. The calculated probabilities of AIS invasion using co-kriging represent current risk status. In the absence of effective eradication measures to remove AIS from invaded waterbodies, the waterbodies that are currently recognized to be at risk will remain in the same status while the intensity of the risk of invasion may increase when newly AIS invasions are reported (Fig. 2.5 and 2.6).

## 2.12 Results

The Pearson correlation coefficients for each variable pair are summarized in Table 2.4. The variable pair with the highest AUC value for the true validation of the zebra mussel model was the Eigenvector centrality of the watercraft movement network and the distance to the nearest major road (AUC= 0.78). Whereas, Eurasian watermilfoil was best predicted by the Eigenvector centrality of the watercraft movement network and the surface water connectivity to infested waterbodies (AUC= 0.76). The AUC values, sensitivity, and specificity at the threshold risk rank = 3 for the cross validations and true validation of co-kriging models are summarized in Table 2.5. The final model included the variables that were correlated with the invasion status and highly sampled.

**Table 2.4.** Pearson correlation coefficient for the six waterbody specific variables used in the study.

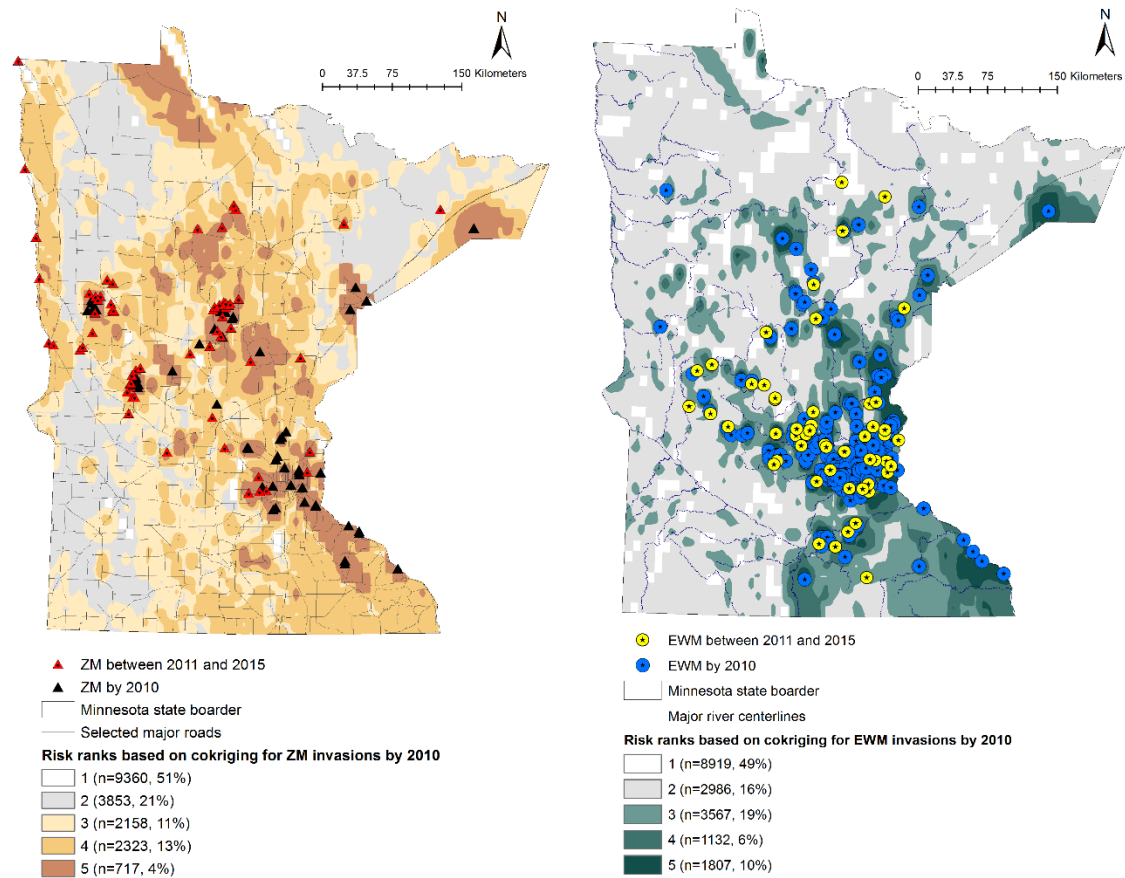
		<b>ZM* invasion status (primary variable)</b>	<b>EWM# invasion status (primary variable)</b>
<b>1</b>	ZM invasion status	1.00	0.10
<b>2</b>	EWM invasion status	0.10	1.00
<b>3</b>	Connectivity to another ZM invaded waterbody via a river or a stream	0.12	0.04
<b>4</b>	Connectivity to another EWM invaded waterbody via a river or a stream	0.09	0.10
<b>5</b>	Eigen vector centrality of the boater traffic network	0.28	0.34
<b>6</b>	Inverse of the Euclidean distance to the nearest major road	0.21	0.09

ZM\* = Zebra mussels and EWM# = Eurasian watermilfoil

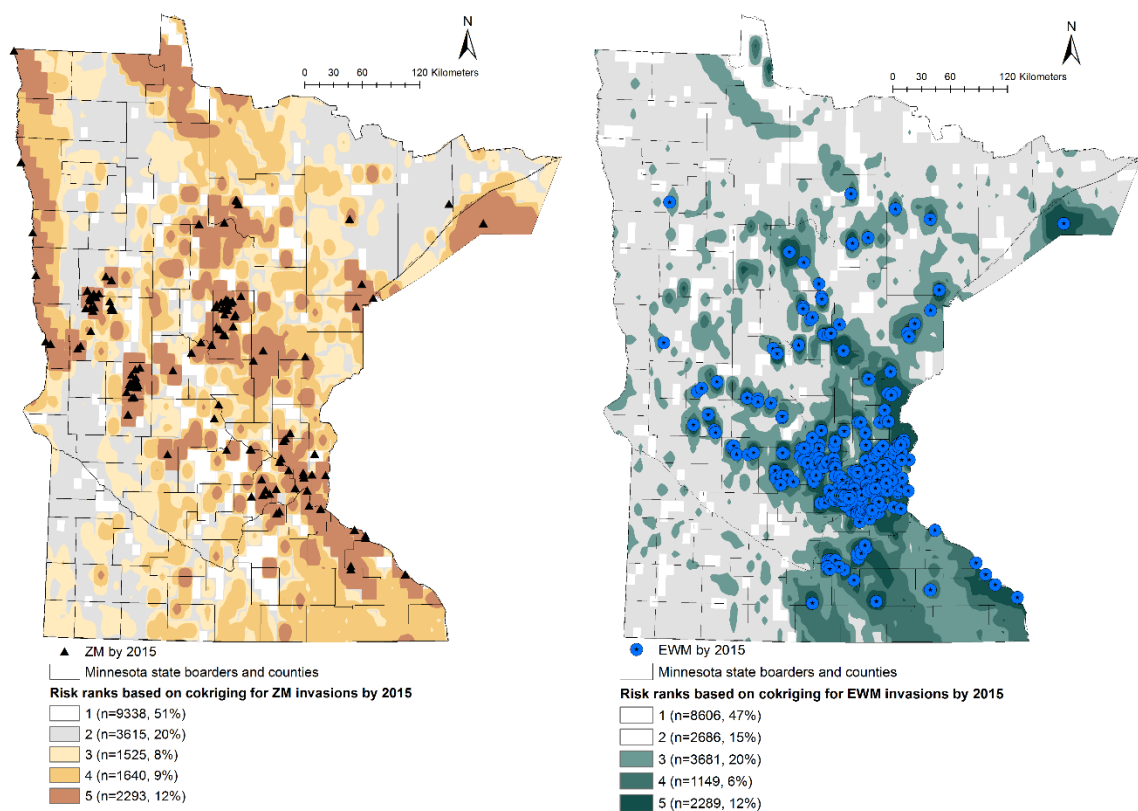
**Table 2.5.** Summary of co-kriging model validations for the probability of zebra mussel and Eurasian watermilfoil introductions in Minnesota. Cross validation was done using the k fold test (k=5). True validation was done by fitting models for invasions as of 2010 and validating using the invasions reported between 2011 and 2015. Area under the receiver operating characteristic curve (AUC), Sensitivity and Specificity at the threshold risk are summarized.

		Output maps for both zebra mussel and Eurasian watermilfoil co-kriging and the number		
		AUC	Sensitivity at risk rank 3	Specificity at risk rank 3
<b>Cross validation</b>	Zebra mussels	0.73	0.70	0.63
	Eurasian watermilfoil	0.79	0.82	0.74
<b>True validation</b>	Zebra mussels	0.78	0.78	0.72
	Eurasian watermilfoil	0.76	0.83	0.61

of waterbodies classified under each risk rank (Fig 2.5 and 2.6). Fig. 2.5 illustrates the risk maps for the models fitted for the invasions by 2010, whereas, Fig. 2.6 shows the risk based on the invasions by 2015. Therefore, by 2015, at the risk rank = 5, a total of 2,293 (12.45%) and 2,289 (12.43%) waterbodies were at very high risk of invasion by zebra mussels and Eurasian watermilfoil, respectively. Among the waterbodies at very high risk at risk rank 5 for both the AIS, 755 waterbodies were in common. Therefore, a total of 3,827 (20.78%) waterbodies were at high risk for either zebra mussel or Eurasian watermilfoil invasions.



**Fig. 2.5.** Co-kriging model outputs illustrating the probability of introduction of zebra mussels (ZM) and Eurasian watermilfoil (EWM) to Minnesota waterbodies, for the invasions as of 2010. The risk classes 1 through 5 indicates the intensity of the probability of introduction, where, class 5 represent a high probability of ZM or EWM introduction. The number of waterbodies under each category and as a percentage of the total waterbodies (n=18,411) is listed.



**Fig. 2.6.** Co-kriging model outputs illustrating the probability of introduction of zebra mussels (ZM) and Eurasian watermilfoil (EWM) to Minnesota waterbodies, for the invasion status of 2015. The risk classes 1 through 5 indicates the intensity of the probability of introduction, where, class 5 represent a high probability of ZM or EWM introduction. The number of waterbodies under each category and as a percentage of the total waterbodies (n=18,411) is listed.

### 2.13 Discussion

This study was aimed at predicting the risk of zebra mussels and Eurasian watermilfoil invasions in Minnesota using network analysis and co-kriging, a geostatistical modeling technique. Recognizing areas at high risk for invasion may facilitate early detection and efficient control through risk-based management. This study emphasized the use of co-kriging on observed data affected by underreporting and other reporting biases by augmenting the predictive power of one variable with the support of other correlated and highly sampled variables. In the absence of active surveillance, invasions are recorded based on public reporting and subsequent confirmation by the MNDNR. Therefore, presence of the AIS may be missed in some waterbodies due to insufficient coverage, resulting in underreporting. Results suggested that, by 2016, nearly 20% of the waterbodies in Minnesota were at high risk of invasions by either or both AIS. This included 2,293/18,411; 12.45% waterbodies at risk of zebra mussel invasions and 2,289/18,411; 12.43% waterbodies at risk of Eurasian watermilfoil invasions, whereas only 125/18,411 (0.67%) and 304/18,411 (1.65%) confirmed the invasions, respectively. Recognition of areas at high risk may act as an early warning system and help prioritization of water bodies for risk-based surveillance and management.

The key predictors of the best fitted co-kriging models, for both zebra mussels and Eurasian watermilfoil, were the distance to the nearest infested location and the boater traffic, i.e., Eigen vector centrality of the boater traffic network. This result emphasizes the proximity between waterbodies and human mediated dispersal as useful predictors of potential invasions (Roley and Newman, 2008; Karatayev *et al.*, 2015). The strong relationship between hitchhiking zebra mussel larvae along with the residual water, boat equipment and recreational gear is a known risk factor for invasions (Banha *et al.*, 2016). Affirmatively, the secondary variables in the final co-kriging model for zebra mussels were both indicators of human mediated dispersal of the AIS, the boater traffic and the distance to the nearest major road which represents the convenience for frequent accessibility. The final co-kriging model for Eurasian watermilfoil suggests that their distribution is attributable to the proximity between waterbodies as determined by the invasion status of Eurasian watermilfoil, the natural dispersal via connecting surface water such as rivers and the human mediated transportation (i.e. variables 2, 4, and 5). The predictive power of the boater traffic using the Eigen vector centrality measure is augmented with the use of the inverse distance to the nearest major road as a secondary predictor, which adjusted for the potential underreporting. The Pearson correlation between zebra mussel invasions and the inverse

of the distance to the nearest major road was 0.21 (Table 1.5), which was stronger than other variables. Distance to the nearest major road represents the convenience of frequent accessibility to the waterbody.

In the absence of active surveillance, AIS invasions are recorded based on public reporting and subsequent confirmation by the MNDNR (MNDNRAIS, 2016). Therefore, densely human populated areas are likely to be reported with invasions more frequently than less populated areas, where underreporting is possible (Aikio *et al.*, 2010; Kanankege *et al.*, 2018). Considering the commonalities between waterbodies with currently reported invasions and searching for waterbodies with similar characteristics using waterbody specific variables may be one of the solutions to correct for underreporting (Lund *et al.*, 2015). However, selecting waterbodies based on multiple criteria such as public water access, boater traffic, tourist activity etc. is challenging and through this study we provide a method which take the most correlated variables into account and produce risk maps and risk ranks for each waterbody, which may offer a better guidance to volunteers who search for potential invasions. This approach of risk-based and targeted surveillance would provide more opportunities to reduce the problem of underreporting.

An important strength of the present study is that the boater traffic was calculated at the waterbody level. This is more informative compared to the representation of boater movement by county centroids, such as the studies by Stewart-Koster *et al.*, 2015 and Buchan and Padilla, 2000 (Stewart Koster *et al.*, 2015; Buchan and Padilla, 2000). Representation of the boater traffic by county leads to either over estimation or underestimation of the importance of individual waterbodies (Stewart Koster *et al.*, 2015).

Areas at high risk for AIS infestations may be identified using a variety of modeling techniques. Species distribution modeling (Gallardo *et al.*, 2013), diffusion models (Buchan and Padilla., 1999), gravity models (Bossenbroek *et al.*, 2007), regression models (Buchan and Padilla., 2000), machine learning techniques (Tamayo and Olden, 2014), risk models (Leung *et al.*, 2012), and model combinations (Stewart-Koster *et al.*, 2015) are approaches commonly used for the estimation of AIS distribution risk. Some of the above-mentioned computationally complex modeling techniques are powerful when determining the risk of invasions; however, the complexity of these models can make the translation of the model output into practice a difficult task. Compared to above modeling techniques, co-kriging is a less complicated analysis. When translating the science to policy, the concept of using correlated and highly sampled variables to estimate unknown variables is rather simple and straightforward. Therefore, the use of co-kriging

as an introductory tool to assess the risk and introducing the method to the decision-makers perhaps is a step further into translating science into practice.

One limitation of our approach is that co-kriging interpolation assumes that the probability of AIS introduction is a continuous variable across geographical space (Isaaks and Srivastava, 1989). However, the probability of AIS introduction is waterbody specific and not a continuous variable. In this study, the assumption of continuous probability may be justified because Minnesota is a water rich state with over 19% of the state is consisting of lakes, ponds, rivers, and wetlands (MNDNR Hydrography, 2015). This assumption of continuous probability is also supported by the density and complexity of the overland boater traffic (Supplement Fig S2.1 and S2.2). Although, this simplification of continuous probability is held commonly in spatial modeling (Martinez *et al.*, 2008), the invasions only occur at the susceptible locations, i.e. the waterbodies. In co-kriging, probability is computed for cells and, here, we assumed the probability of infection to be zero for those cells in which no waterbody was found, whereas the probability of AIS introduction was computed for cells that was occupied, at least in part, by a waterbody. Presentation of co-kriging models in the format of isopleth maps with a continuous probability surface is common in the spatial modeling (Martinez *et al.*, 2008). As mentioned in the methods, magnitude and the duration of the infestation would have been ideal to be included in the analysis because it is a measure of the risk an infested waterbody pose on susceptible waterbodies (Roley and Newman, 2008). However, magnitude of invasions was not readily available because the collection of magnitude of invasions is a costly and labor-intensive process (Claudi and Mackie, 1993; Mellina and Rasmussen 1994) and the distribution of AIS within waterbodies is patchy based on the substrate compositions (Mellina and Rasmussen 1994; Downing and Anderson, 1985). Similarly, the assignment of surface water connectivity both upstream as well as downstream, without limiting the distances, may lead to potential overestimation of the risk of invasion. However, assignment of distance limits of upstream and downstream transmission was subjective as described by multiple studies (Bobeldyk *et al.*, 2005; Spencer and Carruthers, 2012; Osawa *et al.*, 2013). Another limitation is the lack of AIS distribution data in the states adjacent to Minnesota, which is important for effective cross-boundary control and preventive measures. For example, waterbodies in east central Minnesota are affected by both zebra mussels and Eurasian watermilfoil. However, the study described by Stewart-Koster *et al.* (2015) indicated low risk of the zebra mussel and Eurasian watermilfoil invasion across the border in northeastern Wisconsin (Stewart-Koster *et al.*, 2015). Our study does not account for zebra mussels and Eurasian watermilfoil invasions in the adjacent states



either, which indicates the risk of invasion may have been underestimated. Being confined within the political boundaries often results in reducing the model accuracies (Barnes *et al.*, 2014). The geographical area for the analysis was not expanded to the Midwest or great lakes because some of the required data, such as boater movement, was not available from all the locations.

As seen in Fig. 2.6, a total of 5,458 (29.64%) of the waterbodies were recognized to be equal or above the threshold risk rank 3 for zebra mussel invasions. Similarly, 7,119 (38.66%) of the waterbodies were predicted to be above the risk rank 3 for Eurasian watermilfoil invasions. From a management stand point, these numbers of waterbodies are still too high to plan a cost-effective risk-based surveillance or develop targeted management plans. Therefore, risk-based management using limited resources requires prioritizing the waterbodies at high risk for screening (Vander Zanden and Olden, 2008; Mandrak and Cudmore, 2015). This inherent difficulty of recommending sample sizes to be collected from risk regions is also discussed by another study where co-kriging was used to conduct a post hoc comparison of the association between highly pathogenic avian influenza (H5N1) incidences and intensity of surveillance activities of sampling wild birds by administrative region (Martinez *et al.*, 2008). Resource availability, degree of risk awareness, and participation in reporting by the region were recognized as key factors defining the extent of surveillance efforts (Martinez *et al.*, 2008). We suggest focusing on the waterbodies of biological and recreational importance. This can be a values-based judgment and should include a variety of stakeholders and agreed upon criteria. Prioritization of the waterbodies could also be done by conducting a risk-based survey by subdividing the counties into smaller polygons or using township areas. One such approach is the hexagonal tiling method, which is commonly used in ecological studies (Birch *et al.*, 2007). The risk rank generated from this study may also be useful to improve the MNDNR's Watercraft Inspection Program by recruiting watercraft inspectors at areas recognized to be at high risk for invasions and not currently inspected.

## 2.14 Conclusions

Risk-based management is not a novel concept (CRARM, 1997; Mandrak and Cudmore, 2015). However, the attempt to incorporate spatial models in invasion risk assessment to inform the decision and policy-making process may improve the efficiency and effectiveness of the AIS control programs, through targeted and risk-based sampling schemes (Lodge *et al.*, 2006; Vander Zanden and Olden, 2008). As demonstrated here, co-kriging enables predicting values for

locations without complete data, using correlated and highly sampled variables, which can be used as a solution to the underreporting in ecological and epidemiological studies. This work seeks to encourage the use of scientifically supported quantitative procedures such as network analysis and co-kriging to solve the problem of imperfect detections, which subsequently improve the early detection of biological invasions.

## **2.15 Acknowledgements**

This study was funded in part by the Minnesota Discovery, Research, and Innovation Economy (MnDRIVE) program and Office of the Vice President for Research (OVPR) of the University of Minnesota. We thank the Aquatic Invasive Species Prevention program of the Minnesota Department of Natural Resources for providing the incident data and sharing their experience.

## **2.16 References**

1. Aikio S, Duncan RP, Hulme PE. 2010. Herbarium records identify the role of long-distance spread in the spatial distribution of alien plants in New Zealand. *J Biogeogr.* 37,1740-1751.
2. Banha F, Gimeno I, Lanao M, Touya V, Duran C, Peribanez MA, Anastacio PM. 2016. The role of waterfowl and fishing gear on zebra mussel larvae dispersal. *Biol. Invasions.* 18, 115-125.
3. Barnes MA, Jerde CL, Wittmann M., Chadderton WL, Ding JQ, Zhang JL, Purcell M, Budhathoki M, Lodge DM. 2014. Geographic selection bias of occurrence data influences transferability of invasive *Hydrilla verticillata* distribution models. *Ecol. Evol.* 4, 2584-2593.
4. Birch CPD, Oom SP, Beecham JA. 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling.* 206, 347-359.
5. Bobeldyk AM, Bossenbroek JM, Evans-White MA, Lodge DM, Lamberti GA. 2005. Secondary spread of zebra mussels (*Dreissena polymorpha*) in coupled lake-stream systems. *Ecosci.* 12, 339-346.

6. Bossenbroek JM, Johnson LE, Peters B, Lodge DM. 2007. Forecasting the expansion of zebra mussels in the United States. *Conserv. Biol.* 21, 800-810.
7. Buchan LAJ, Padilla DK. 1999. Estimating the probability of long-distance overland dispersal of invading aquatic species. *Ecol. Appl.* 9, 254-265.
8. Buchan LAJ, Padilla DK. 2000. Predicting the likelihood of Eurasian watermilfoil presence in lakes, a macrophyte monitoring tool. – *Ecol. Appl.* 10, 1442-1455.
9. Claudi R, Mackie, GL. 1993. Practical manual for zebra mussel monitoring and control. CRC Press Inc. Florida. ISBN 0-8771-985-9.
10. Connelly NA, O'Neill CR, Knuth BA, Brown TL. 2007. Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. *Environ. Management.* 40, 105-112.
11. CRARM: The Presidential/ Congressional Commission on Risk Assessment and Risk Management. 1997. Framework for Environmental Health Risk Management. Commission on Risk Assessment and Management, Washington DC. Vol. 1.
12. Downing JA, Anderson MR. 1985. Estimating the standing biomass of aquatic macrophytes. *Can. J. Fish. Aquat. Sci.* 42:1860-1869.
13. ESRI: ArcMap Version 10.3.1. 2016. Environmental Research Institute, Inc. Redlands, CA, USA.
14. Fielding AL, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.*, 24, 38-49
15. Fisheries Stream Survey Manual. 2007. Stream Survey Methods. Published by Minnesota Department of Natural Resources. Special publication No. 165. Version 2.1. May, 2007. Available from: [http://files.dnr.state.mn.us/publications/fisheries/special\\_reports/165.pdf](http://files.dnr.state.mn.us/publications/fisheries/special_reports/165.pdf).
16. Gallardo B. 2014. Europe's top 10 invasive species: relative importance of climatic, habitat and socio-economic factors. *Ethology Ecol. Evol.* 26(2-3), 130-151.

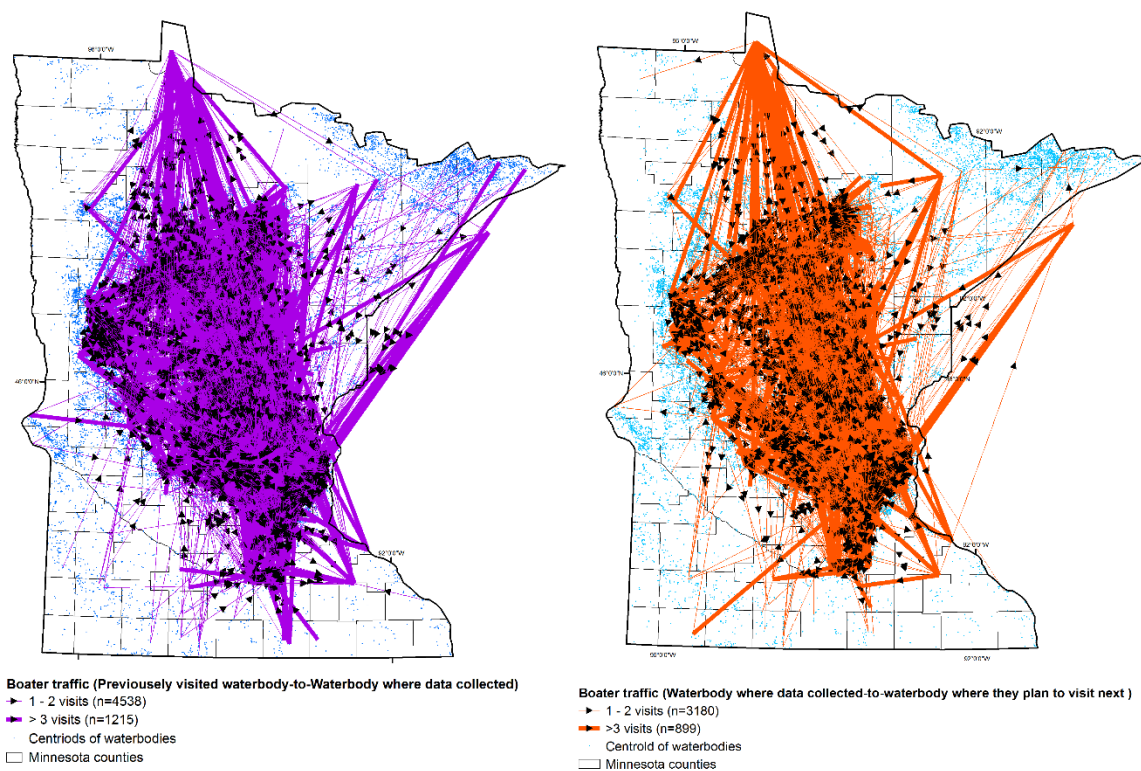
17. Gallardo B, Ermgassen P, Aldridge DC. 2013. Invasion ratcheting in the zebra mussel (*Dreissena polymorpha*) and the ability of native and invaded ranges to predict its global distribution. *J Biogeogr.* 40, 2274-2284.
18. Havel JE, Kovalenko KE, Thomaz SM, Amalfitano S, Kats LB. 2015. Aquatic invasive species: challenges for the future. *Hydrobiol.* 750,147-170.
19. Invasive Species Program MNDNR Annual Report: Invasive Species of Aquatic Plants and Wild Animals in Minnesota; Annual Report for 2015, Minnesota Department of Natural Resources, St. Paul, MN. Minnesota Department of Natural Resources Annual report of Invasive species of Minnesota of 2015. 2016. Available from: [http://files.dnr.state.mn.us/natural\\_resources/invasives/ais-annual-report.pdf](http://files.dnr.state.mn.us/natural_resources/invasives/ais-annual-report.pdf)
20. Isaaks EH, Srivastava RM. 1989. *Applied Geostatistics*. Oxford University Press. New York. ISBN-13: 978-0195050134.
21. Kanankege KST, Alkhamis MA, Perez AM, Phelps NBD. 2018. Zebra mussels and Eurasian watermilfoil reporting patterns in Minnesota. *J. Great Lakes Res.* doi: 10.1016/j.jglr.2018.02.008
22. Karatayev, AY, Burlakova LE, Mastitsky SE, Padilla DK. 2015. Predicting the spread of aquatic invaders: insight from 200 years of invasion by zebra mussels. *Ecol. Appl.* 25, 430-440.
23. Kelting DL, Laxson CL. 2010. Cost and Effectiveness of Hand Harvesting to Control the Eurasian Watermilfoil Population in Upper Saranac Lake, New York. *J. Aquat. Plant Manage.* 48,1-5.
24. Leung B, Roura-Pascual N, Bacher S, Heikkila J, Brotons L, Burgman MA, Dehnen-Schmutz K, Essl F, Hulme PE, Richardson DM, Sol D, Vila M. 2012. TEASIng apart alien species risk assessments: a framework for best practices. *Ecol. Letters.* 15, 1475-1493.
25. Lodge DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA, Carlton JT, McMichael A. 2006. Biological invasions: Recommendations for US policy and management. *Ecol. Appl.* 16, 2035-2054.

26. Lund K, Bloodsworth K, Wolbers T, Weling C, Gamble A. 2015. Guidance for conducting aquatic invasive species early detection and baseline monitoring in lakes. Invasive Species Program. Division of Ecological and Water Resources of the Minnesota Department of Natural Resources. Available from:  
[https://www.ifound.org/files/6714/4745/1209/ais\\_detection-baseline-monitoring.pdf](https://www.ifound.org/files/6714/4745/1209/ais_detection-baseline-monitoring.pdf)
27. MacIsaac HJ. 1996. Potential abiotic and biotic impacts of zebra mussels on the inland waters of North America. *Am. Zoologist* 36, 287-299
28. Mandrak NE, Cudmore B. 2015. Risk assessment: Cornerstone of an aquatic invasive species program. *Aquat. Ecosys. Heal. Manage.* 18, 312-320.
29. Martínez M, Perez AM, de la Torre A, Iglesias I, Muñoz MJ. 2008. Association between number of wild birds sampled for identification of H5N1 avian influenza virus and incidence of the disease in the European Union. *Transbound. Emerg. Dis.* 55, 393-403.
30. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. 2009. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound. Emerg. Dis.* 56, 109-120.
31. Mellina E, Rasmussen JB. 1994. Patterns in the distribution and abundance of zebra mussel (*Dreissena polymorpha*) in rivers and lakes in relation to substrate and other physicochemical factors. *Can. J. Fisheries Aquat. Sci.* 51, 024-1036.
32. Mills EL, Leach JH, Carlton JT, Secor CL. 1993. Exotic species in the Great-Lakes: A history of biotic crisis and anthropogenic introductions. *J Great Lakes Res.* 19,1-54.
33. Minnesota Statute 477A.19. 2016. Aquatic Invasive Prevention Aid. The office of the Reviser of the Statutes. Chapter 308. Article 1. Section 11. Available from:  
<https://www.revisor.mn.gov/statutes/?id=477A.19&format=pdf>.
34. MNDNR AIS: Minnesota Department of Natural Resources: Aquatic Invasive Species. 2016. Available from: <http://www.dnr.state.mn.us/invasives/ais/infested.html>
35. MNDNR WIP: Watercraft Inspection Program of the Minnesota Department of Natural Resources. 2014. Available from:

- [http://files.dnr.state.mn.us/natural\\_resources/invasives/mndnr\\_ais\\_watercraft\\_inspection\\_hanbook.pdf](http://files.dnr.state.mn.us/natural_resources/invasives/mndnr_ais_watercraft_inspection_hanbook.pdf).
36. MNGSC Hydrography: Minnesota GeoSpatial commons. MNDNR Hydrography data layer. 2015. [cited August 15, 2015]. Available from: <https://gisdata.mn.gov/dataset/water-dnr-hydrography>.
  37. MNGSC Roads: Minnesota GeoSpatial commons. Roads, Minnesota 2012 data layer. Published by Minnesota Department of Transportation (MnDOT). 2012. [cited: May 15, 2016]. Available from: <https://gisdata.mn.gov/dataset/trans-roads-mndot-tis>. Last updated: 02/25/2016.
  38. MNGSC Stream: Minnesota GeoSpatial commons. Stream Routes with Kittle Numbers and Mile Measures data layer. 2015. Content update: 11/01/2013. [cited: November 15, 2015]. Available from: <https://gisdata.mn.gov/dataset/water-measured-kittle-routes>.
  39. Oliver MA, Webster R, Lajaunie C, Muir KR, Parkes SE, Cameron AH, Stevens MCG, Mann JR. 1998. Binomial co-kriging for estimating and mapping the risk of childhood cancer. *Ima J. Math. Appl. Med. Biol.* 15, 279-297.
  40. Osawa T, Mitsuhashi H, Niwa H. 2013. Many alien invasive plants disperse against the direction of stream flow in riparian areas. *Ecol. Complex.* 15, 26-32.
  41. Perez AM, Thurmond MC, Carpenter TE. 2006. Spatial distribution of foot-and-mouth disease in Pakistan estimated using imperfect data. *Prev. Vet. Med.* 76, 280-289.
  42. Pysek P, Richardson DM. 2010. Invasive Species, Environmental Change and Management, and Health. Pages 25-55 in A. Gadgil and D. M. Liverman, editors. *Annual Review of Environment and Resources*, Vol 35. Annual Reviews, Palo Alto.
  43. Ricciardi A. 2001. Facilitative interactions among aquatic invaders: is an "invasional meltdown" occurring in the Great Lakes? *Can. Fish. Aquat. Sci.* 58, 2513-2525.
  44. Rogers DJ, Sedda L. 2012. Statistical models for spatially explicit biological data. *Parasitol.* 139: 1852-1869.

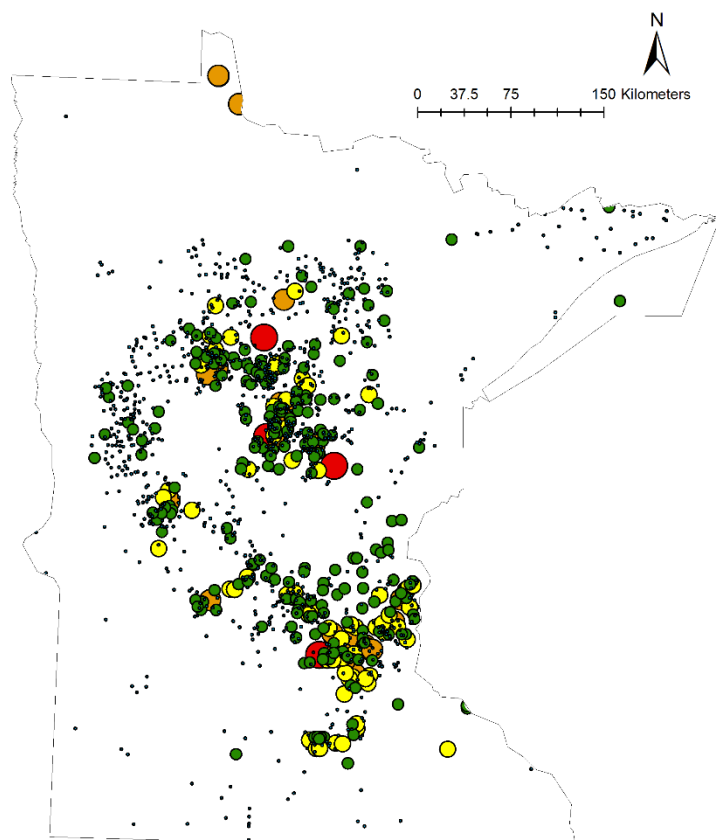
45. Roley SS, Newman RM. 2008. Predicting Eurasian watermilfoil invasions in Minnesota. *Lake Reserv. Manage.* 24, 361-369.
46. Seekamp E, McCreary A, Mayer J, Zack S, Charlebois P, Pasternak L. 2016. Exploring the efficacy of an aquatic invasive species prevention campaign among water recreationists. *Biol. Invasions.* 18, 1745-1758.
47. Simberloff D. 2009. The Role of Propagule Pressure in Biological Invasions. *Annual Review of Ecology Evolution and Systematics.* 40:81-102. Doi: 10.1146/annurev.ecolsys.110308.120304
48. Spencer DF, Carruthers RI. 2013. Predicting Eurasian watermilfoil's (*Myriophyllum spicatum*) distribution and its likely response to biological control in a spring-fed river. *J. Aquat. Plant Manage.* 51, 7-14.
49. Stewart-Koster B, Olden JD, Johnson PTJ. 2015. Integrating landscape connectivity and habitat suitability to guide offensive and defensive invasive species management. *J Appl. Ecol.* 52, 366-378.
50. Swets JA. 1988. Measuring the accuracy of diagnostic systems. *Science.* 240, 1285–1293.
51. Tamayo M, Olden JD. 2014. Forecasting the Vulnerability of Lakes to Aquatic Plant Invasions. *Invasive Plant Sci. Manage.* 7, 32-45.
52. Vander Zanden MJ, Olden JD. 2008. A management framework for preventing the secondary spread of aquatic invasive species. *Can. J. Fisheries Aquat. Sci.* 65, 1512-1522.
53. Vauclin M, Vieira SR, Vachaud G, Nielsen DR. 1983. The use of co-kriging with limited field soil observations. *Soil Sci. Society Am. J.* 47,175-184.

## 2.17 Supplement Tables and Figures



**Supplement Fig S2.1** The boater traffic between waterbodies based on the Watercraft Inspection Program conducted by Minnesota Department of Natural Resources. The data from year 2013 are illustrated. The panel A of figure represents the movement of boaters from previously visited waterbody-to-waterbody where the survey data were collected. The panel B represents the movement of boaters from waterbody where the survey data were collected-to-the waterbody where they plan to visit next.





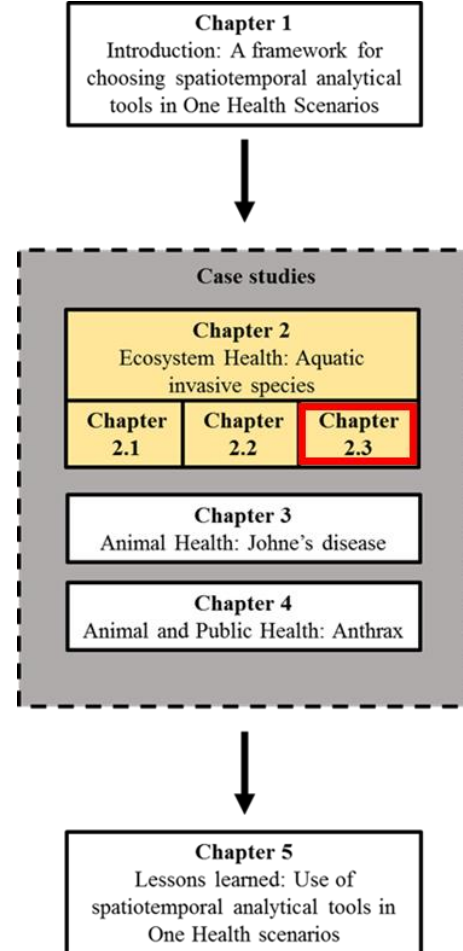
**Eigen Vector centrality based on the boater traffic network**

- 0.000004 - 0.049241
  - 0.049242 - 0.141691
  - 0.141692 - 0.285035
  - 0.285036 - 0.487783
  - 0.487784 - 1.000000
- Minnesota state boarder

**Supplement Fig 2.2** An illustration of the Eigen vector centrality for the waterbodies in the boater traffic network created using the surveys of Watercraft Inspection Program conducted by Minnesota Department of Natural Resources. The data from year 2013 are illustrated.

**Supplement Table S2.1.** Parameters used in the co-kriging models for Eurasian watermilfoil (EWM) and zebra mussels (ZM).

Parameter	ZM model*			EWM# model		
Lag size	0.025			0.04		
Number of lags	12			8		
Models	Exponential			Exponential		
Major range	0.15			0.15		
Search neighborhood parameters for each predictor variable	Primary variable: ZM invasion status	Boater traffic	Inverse of the Euclidean distance to nearest major road	Primary variable: EWM invasion status	Boater traffic	Connectivity to another EWM invaded waterbody via a river or a stream
Minimum number of neighbors to be included in the kriging process	5	200	200	50	50	200
Maximum number of neighbors to be included in the kriging process	2	2	2	2	2	2
Sector type	1	1	1	1	1	1
Angle	135°	135°	135°	120°	120°	120°
Anisotropy factor	2	1	1	1	1	1
Root mean Square value (RMS) of the model	0.057			0.1058		
RMSE standardized value of the model	3.86			0.836		
Nugget	0			Between 0 and 0.01		



### Chapter 2.3

*Developing a risk scoring system to estimate to recognize areas at high risk for Zebra Mussels and Eurasian Watermilfoil invasions in Minnesota: Combined interpretation of Maxent and co-kriging outputs*

This work has not been published/in-review elsewhere

## 2.18 Overview

Aquatic invasive species (AIS) disrupt the stability of ecosystems and are difficult to eradicate. Zebra mussels (*Dreissena polymorpha*) and Eurasian watermilfoil (*Myriophyllum spicatum*) are two aggressive AIS reported in Minnesota, U.S., since the late 1980s. By the end of 2015, 125 (0.67%) and 304 (1.65%) of the 18,411 waterbodies of Minnesota were infested with zebra mussels and Eurasian watermilfoil, respectively. Recognition of the areas at high risk of invasion through scientific evidence may facilitate early detection and the efficient allocation of limited resources to reduce the AIS spread. This study uses co-kriging and maximum entropy ecological niche (Maxent) models to predict the risk of AIS distribution and, ultimately, to inform the targeted surveillance activities. Co-kriging was used to calculate the probability of introduction of AIS, using waterbody-specific variables available from selected waterbodies. Whereas, Maxent was used to determine the ecological susceptibility of the waterbodies, using variables available for the entire state. Thus, the combined interpretation of Maxent and co-kriging results serves as a proxy for the probability of AIS infestation in waterbodies of Minnesota. Maxent results suggested that human population density was the most important predictor for the suitability of both the AIS, zebra mussels (41.7%) and Eurasian watermilfoil (58%), followed by annual mean temperature (19.1%) and precipitation during the wettest month (17.6%) for zebra mussels and Eurasian watermilfoil, respectively. The zebra mussel co-kriging model included distance to the nearest zebra mussel-infested location, boater traffic, and road access, whereas the Eurasian watermilfoil model included distance to the nearest Eurasian watermilfoil-infested location, boater traffic, and connectivity to infested waterbodies. Combined results from the Maxent and co-kriging outputs suggested that nearly 15% of the waterbodies in Minnesota were at high risk of invasion by zebra mussels or Eurasian watermilfoil. Results suggest that targeted sampling of the high-risk waterbodies may facilitate the early detection and prevention of AIS spread in Minnesota.

## 2.19 Introduction

Aquatic invasive species (AIS) have the potential to impair environment, economy, and public health (Pysek and Richardson, 2010; Havel et al., 2015). In an effort to control AIS, aggressive and costly programs have been implemented in Minnesota (MNDNR Annual Report, 2016). For example, since 2014, \$10 million per year has been allocated by the Minnesota legislature to provide resources for county based AIS prevention activities (MN Statute 477A.18, 2016). However, because the risk of AIS invasion had not been previously quantified, early detection of invasions has been challenging, therefore, the resources were distributed proportionally to the number of boat ramps and trailer parking spaces in each county (MNDNR Annual Report, 2016; MNDNR, 2016). Because of the high economic and conservation burden posed by the invasions, forecasting of the areas at high risk for invasions is considered as a research priority (Drake and Bossenbroek, 2009).

Areas at high risk for AIS infestations may be identified using a variety of spatial modeling techniques. Species distribution modeling (Quinn et al., 2014; Papes et al., 2016), diffusion models (Buchan and Padilla, 1999), gravity models (Bossenbroek et al., 2007), regression models (Buchan and Padilla, 2000; Drake and Bossenbroek, 2009; Stewart Koster et al., 2015), and empirical analyses and simulations (Latzka et al., 2016) are approaches commonly used for the estimation of AIS distribution risk. Majority of the above approaches were focused either on the prediction of habitat distribution or the probability of AIS introductions, separately, with few exceptions such as the work by Drake and Bossenbroek, (2009) and Stewart Koster et al. (2015) where they estimate distribution of the niches first and use statistical methods to recognize the niche boundaries. Successful invasions require both habitat suitability and the probability of the AIS introduction (i.e. translocation) (Havel et al., 2015). Therefore, an approach which combine the probability of introduction and the habitat susceptibility is preferable at predicting the risk of AIS invasions.

Zebra mussels (*Dreissena polymorpha*) and Eurasian watermilfoil (*Myriophyllum spicatum*) are AIS of major environmental and economic concern for Minnesota. Both AIS have been reported from Minnesota waters since 1989 and 1987, respectively (MNDNR Annual Report, 2016; <https://www.dnr.state.mn.us/invasives/ais/infested.html>). Therefore, these two AIS were chosen as indicator species for this study.

The objective of the study was to estimate the potential range expansion of zebra mussels and Eurasian watermilfoil in Minnesota using a combination of modeling techniques, namely:

Maximum entropy ecological niche modeling (Maxent), network analysis, and cokriging (a spatial interpolation technique). Species distribution models are used to identify statistically significant associations between AIS presence and environmental variables and to use those relationships to identify suitable environmental conditions in non-sampled locations (Elith et al., 2011). Because the two invasive species have been invading the Minnesota waters since over 20 years, the measurement of prologue pressure i.e. the ‘introduction effort’ need to be focused on both large scale across the region as well as on the individual waterbody. Therefore, in addition to the use of Maxent models which incorporate environmental variables available from the whole region, waterbody specific variables were used to prioritize the waterbodies at high risk of invasions using network analysis and cokriging. The advantage of using cokriging is that the technique enables to predict values for the locations without the data using other correlated and highly sampled variables, a way of addressing the availability of limited data (Isaaks and Sreevastava, 1989; Kanankege et al., 2018b).

Recognition of areas at high risk may act as an early warning system and help AIS prioritization of water bodies for a targeted and efficient allocation of limited resources to improve both defensive and offensive management strategies (Mandrak et al., 2015; Stewart Koster et al., 2015). Such risk targeted approaches certainly represent improvements over random selection of waterbodies for surveillance and risk-based management purposes (CRARM, 1997; Lodge et al., 2006; Vander Zanden et al., 2008).

## **2.20 Data and Methods**

### **2.20.1 Data**

A total of 18,411 point locations representing waterbodies of Minnesota were the population of waterbodies in this study. Waterbodies were mainly lakes and ponds (n=18,263) and were represented by the centroids of each waterbody. In addition to the lakes, several riverine locations (n=148) were included in the analysis. Riverine locations were identified at the rivers’ midpoint within each county. The locational data for the waterbodies were extracted from the GIS layer referred to as “MNDNR Hydrography,” which is available from the Minnesota GIS Commons (MNGSC Hydrology, 2015). AIS presence data were collected from the Minnesota Department of Natural Resources (MNDNR) database (MNDNRAIS, 2016). By the end of 2015,

there were 125/18,411 (0.67%) zebra mussels and 304/18,411 (1.65%) Eurasian watermilfoil infested waterbodies in Minnesota (MNDNRAIS, 2016; Kanankege et al., 2018b). The known presence of the AIS was used in the study regardless of the magnitude of infestation.

## 2.20.2 Methods

### *General approach*

Three modeling techniques were combined to estimate the risk of invasion by zebra mussels and Eurasian watermilfoil, namely: Maxent, network analysis, and cokriging. Maxent was used to predict the habitat distribution of each AIS separately, using the invasive species presence only data and the variables available for the entire study area.

Network analysis was used to generate a waterbody specific variable representing the importance of the waterbody in the network of boater traffic within Minnesota, using data from a survey conducted by the MNDNR at selected waterbodies (Kanankege et al., 2018b). Cokriging was used to estimate the probability of species introduction, using waterbody specific variables including the boater traffic (Kanankege et al., 2018b). The Maxent and cokriging outputs were interpreted in-parallel combination using a risk matrix to determine the number of waterbodies at high risk for invasions. All models were both cross validated by k fold test and true validated by fitting the models for the invasion status by 2010 and validating using the invasions occurred after 2010. The combined interpretation of the models using risk matrix was true validated using the risk ranks for Maxent and cokriging model outputs of 2010.

Finally, Maxent and cokriging models were fit for the invasion status by 2015, using the model parameters for year 2010. Based on the 2015 Maxent and cokriging models for each species, risk matrices were generated to calculate the number of waterbodies at high risk for the invasion as of 2015.

### *Maxent model to predict the habitat distribution*

Maxent is a species distribution modeling algorithm that makes use of presence only data (occurrences) and grid based environmental variables to estimate the potential range expansion (Elith et al., 2011; Merow et al., 2013). Maxent method has become popular for predicting the

spatial distribution of vectors, invasive species, and infectious diseases in ecological studies as well as in human and veterinary medicine (Papes et al., 2016; Alkhamis et al., 2016). A detailed description of the Maxent algorithm is available elsewhere (Phillips et al. 2006; Merow et al., 2013). Under the null model, species is assumed to be equally likely to occur anywhere in the landscape (Elith et al., 2011). The Maxent algorithm was used here to identify the distribution of reported AIS occurrences that maximized the entropy of predictions either by the most spread out distribution or the closest to uniform, under the null model. We fit two independent Maxent models for zebra mussels, and Eurasian watermilfoil. The Maxent program version 3.3.3k was implemented in Dismo, which is an R statistical package, was used (Phillips et al. 2006; Hijmans et al., 2016). The presence locations were assigned to the centroids of waterbodies rather than georeferencing the exact area where AIS were found. The reporting of the invasions is based on public reporting followed by confirmation by the MNDNR, therefore, presence locations are likely to be influenced by the human population density around the waterbodies, indicating that human population density may act as a confounder when introducing the AIS as well as when reporting the invasions (Kanankege et al., 2018a). Because of the logistic and financial constraints in conducting a surveillance, the available presence locations were used in the study.

Background data points, defined as the locations where the species could disperse, regardless of whether the waterbody has or has not been occupied by the species yet (Elith et al., 2011), were used as a null model to calculate the probability of species distribution. Because the AIS are not in an equilibrium with their environment, true absence points were not confirmed. Those locations considered as no AIS invasions were pseudo absent locations (i.e. the absence of AIS was not upheld by a survey or any other method). To keep the methods simple and compatible between the two AIS, instead of species specific tuning, the default logistic model for Maxent was used with the default convergence threshold, regulation, and count of iterations to ensure that the output probability estimates were values ranging between 0 and 1 (Merow et al., 2013).

Selected grid based predictor variables (n=23) which were used in the Maxent models included temperature and precipitation (n=19), geology (n=1), altitude (n=1), lake bathymetry (n=1), and the human population density (n=1). Although data on surface water temperature was not available, that variable has a linear relation with atmospheric temperature, which served as a suitable proxy (Read et al., 2014). Description of the variables used in the Maxent models are listed in the supplement document (Supplement Table S2.2). Global climate variables were downloaded from the WorldClim global climate database ([www.worldclim.org](http://www.worldclim.org)) (Hijmans et al.,



2005). The downloaded raster files were clipped by the geographical extent of the state of Minnesota. The WorldClim data are interpolated maps of the mean climatic data gathered from various sources between 1950 and 2000 (Hijmans et al., 2005). Soil type, mineral contents, and alkalinity have been recognized to influence the suitability of zebra mussels and Eurasian watermilfoil invasions (Bobeldyk et al., 2005; Roley and Newman, 2008; Quinn et al., 2014; Karatayev et al., 2015). Multiple experimental and environmental modeling studies indicate the influence of bedrock geology when determining the surface water chemistry (Yadav and Chakrapani, 2006; Olson and Hawkins, 2012), therefore, in the absence of water chemistry information from all the waterbodies, the bedrock geology data layer was used as a proxy. The bedrock geology map of Minnesota was downloaded from the Minnesota Geological Survey (Jirsa et al., 2011). A 1m resolution digital surface model that was generated from raw Light Detection and Ranging (LiDAR), a digital grid of terrain surface data, was used to represent the topography (DEM, 2012). The digital surface model grid was resampled using the bilinear method to have ~1 km<sup>2</sup> cell size. Lake depth has also been recognized as a risk factor for zebra mussels and Eurasian watermilfoil invasions, and thus, bathymetric data were incorporated into the analysis (Roley and Newman, 2008; Karatayev et al., 2015). Lake bathymetric data used here were a combination of the lake bathymetric Digital Elevation Model published by the MNDNR Fisheries Division, which was available from the Minnesota GIS Commons and the maximum lake depth data from the MNDNR Hydrography data layer (MNGSC Hydrology, 2015; MNGSC Bathymetric, 2015). Although human population density is not an environmental variable itself, it is used as a surrogate variable to account for the environmental changes that are human mediated. Human population density data were obtained from the LandScan Global Population Database, representing data from the census year 2011 (Landscan, 2012). The resolution and the projection of all environmental data layers were used at ~1 km<sup>2</sup> and the Geographic Coordinate System World Geodetic System 1984 (i.e. GCS\_WGS\_1984) Web Mercator projection, respectively.

Independent Maxent models were fit for zebra mussels, Eurasian watermilfoil, using all 23 variables, the variables with >10% relative contribution to the model prediction were selected (Table 2.6). A Spearman pairwise rank correlation was performed on all the standardized pairs of bioclimatic variables and relatively uncorrelated climate variables (Pearson  $r < 0.5$ ) were considered as the predictors contributing to each zebra mussels and Eurasian watermilfoil model (Dormann et al., 2013). Final models were selected using jackknife tests, a resampling method that excludes one variable at a time when running the model, to calculate the contribution of each environmental variable to the final model. The most parsimonious model, i.e. the model with

limited and most contributing predictors, which has the highest training gain, was selected as the best model (Baldwin, 2009).

Cross validation is a process in which certain AIS infested locations were left out from the analysis and the Maxent model was used to estimate the probability of finding an AIS infestation at those left out locations (Hijmans et al., 2016). The k fold-partitioning scheme of cross validation was used to evaluate the performance of the final Maxent models (Elith et al, 2011; Hijmans et al., 2016). In addition to the inbuilt cross validation of the Maxent package, we performed a manual 5 fold cross validation where each candidate Maxent model was tested five times (k=5) against 10,000 randomly generated background points (pseudo absences) (Elith et al, 2011). Eighty percent of the cases were used for the model training and testing was done using the 20% of the withheld cases at each validation.

Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), a plot of model sensitivity (true positives) and 1-specificity (i.e. false positives) (Swets, 1988). AUC values lower than 0.7 are considered relatively inaccurate because the proportion of false and true positive results is not substantially different, whereas AUC values greater than 0.7 are generally considered appropriate (Swets, 1988). Here, Maxent models with AUC greater than 0.7 were considered to provide an accurate prediction. All Maxent Models were tested for spatial sorting bias by calculating the calibrated AUC to assess the impact reporting bias from the areas where multiple locations were reported to have AIS invasions (Hijmans, 2012; Alkhamis et al., 2016). Spatial sorting bias is defined as “The difference between the geographic distance from testing presence to training presence sites and the geographic distance from testing absence (or testing background) to training presence sites” (Hijmans, 2012). If the calibrated AUC (cAUC) was close to 1 then there was no spatial sorting bias, whereas, if the value is close to 0, then it is considered that there is spatial sorting bias (Hijmans, 2012).

In addition to the cross validation, true validation of the Maxent models for each AIS was done by dividing the presence data at year 2010, fitting the Maxent for the 2010 presence data, and validating the predictions using the invasions reported between 2011 and 2015. There were 87/125 cases of zebra mussels and 213/304 cases of Eurasian watermilfoil cases by 2010. All the Maxent outputs were classified into five by the quantiles of predicted values and an AUC was generated. Models with AUC greater than 0.7 were considered to provide an accurate prediction.

### **Cokriging to estimate the probability of introduction**

Probability cokriging was used to estimate the probability of zebra mussel or Eurasian watermilfoil introduction into the waterbodies, conditional to the distance between locations and other waterbody specific variables. The full description of the cokriging process used in this analysis is described elsewhere (Kanankege et al., 2018b). Cokriging is a geostatistical interpolation technique that uses multiple correlated waterbody specific variables to estimate the spatial distribution of the dependent variable in the non-sampled locations, which in this study was the distribution of AIS (Isaaks and Sreevastava, 1989; Wackernagel 1994). “Sampled locations” were those known to be infested, whereas “non-sampled locations” were those that were not reported to be infested. Cokriging is a linear weighted averaging method in which weights are selected to minimize the variance of the estimation error by accounting for the spatial correlation between the waterbody specific variables, weights are depended on the distance between sampled locations (Isaaks and Sreevastava, 1989). A complete description of cokriging is available elsewhere (Isaaks and Sreevastava, 1989).

### ***Combined interpretation of Maxent and cokriging outputs using a risk matrix***

Predicted probabilities were extracted for each of the waterbodies from both MaxEnt and cokriging models, for zebra mussels and Eurasian watermilfoil, and were interpreted in series to indicate the overall risk of invasion at each waterbody. The probabilities resulted from both cokriging and Maxent were ranked into 5 classes based on the quantiles of the output probability values, they were considered as risk ranks (Fig 2.7 and 2.8). In the risk matrix, each cell represents the number of waterbodies categorized under each risk rank combination. The number of waterbodies with Maxent risk rank  $\geq 3$  were further narrowed down if the waterbodies resulted in a cokriging risk rank  $\geq 3$ . The combined risk was ranked into 3 categories, namely, 1) very high, 2) intermediate, and 3) negligible. Risk matrices were generated using the Maxent and cokriging models fitted for the invasions by 2010 as well as for the model fitted for the invasions by 2015.

Risk matrices are commonly used during Risk assessment process to define the level of risk by considering the probability of the event and the severity of the consequence (Dufour et al.,

2011). In the analysis, being recognized as a waterbody with high probability of species distribution by Maxent was considered as the “severity” and the probability of introduction recognized by cokriging output was considered as the “probability of the event”. Further details regarding the use of risk matrices are available elsewhere (Dufour et al., 2011). The combined interpretation was true validated by fitting Maxent and cokriging models for the invasions by 2010 and validating using the invasions reported after 2010.

## **2.21 Results**

### **2.21.1 Maxent and cokriging**

Maxent results suggested that human population density was the most important predictor for both zebra mussels (41.7%) and Eurasian watermilfoil (58%), followed by annual mean temperature (19.1%) and precipitation of the wettest month (17.6%) for zebra mussels and Eurasian watermilfoil, respectively. The cross validation AUC values for zebra mussels and Eurasian watermilfoil model fits were greater than 0.7, which suggests that the models were accurate. The cAUC values were 0.81 for zebra mussels and 0.71 for Eurasian watermilfoil; values were close to 1 (one), suggesting absence of spatial sorting bias (Table 2.6).

**Table 2.6.** Predictors used in the final Maxent models for zebra mussels (ZM<sup>#</sup>) and Eurasian watermilfoil (EWM<sup>‡</sup>).

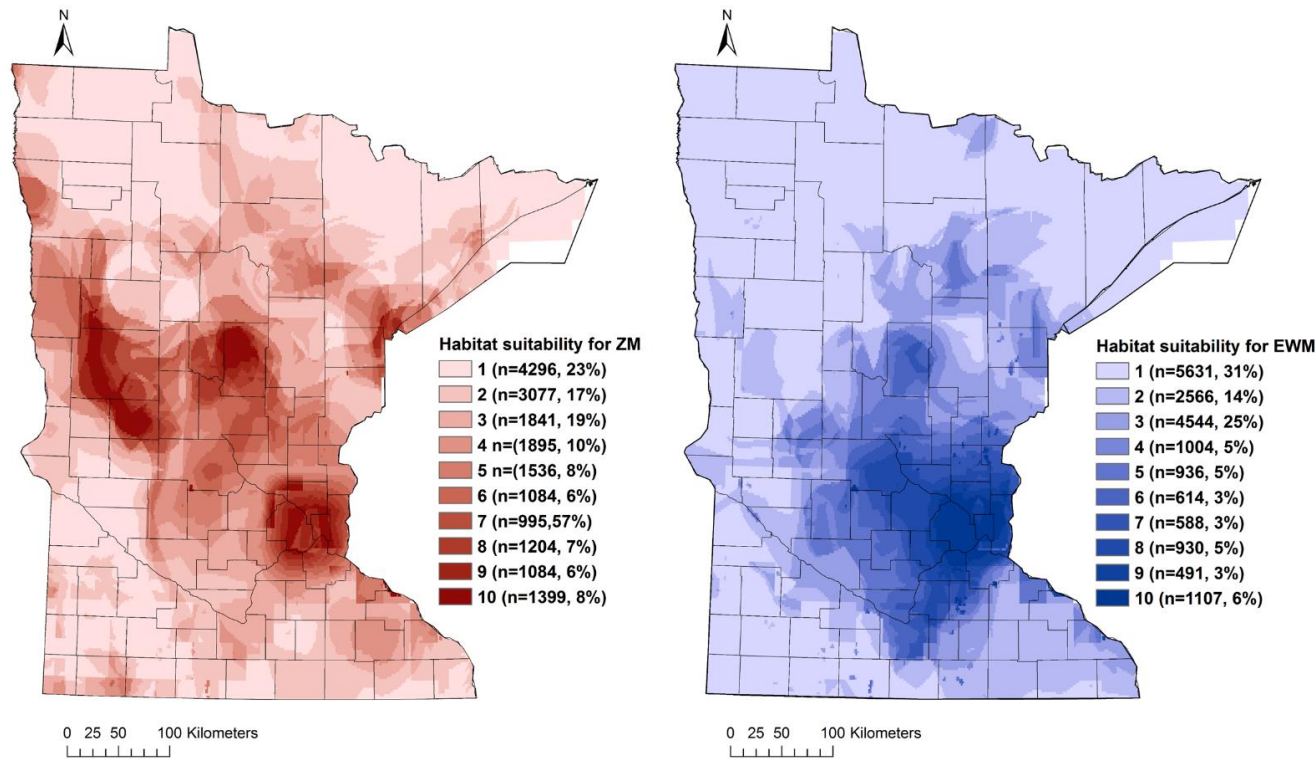
Variable	Percent contribution of the variable in the Maxent*	
	ZM <sup>#</sup>	EWM <sup>‡</sup>
Annual mean temperature	19.1%	
Isothermality (Mean diurnal range of temp / temp annual range)	2.7%	
Temperature seasonality (Standard deviation * 100)		3.3%
Mean temperature of warmest quarter		6.2%
Mean Temperature of coldest quarter		
Annual precipitation		2.5%
Precipitation of wettest month	16.7%	17.6%
Precipitation of driest month	6.8%	
Bedrock geology map of Minnesota	1.4%	0.2%
Minnesota digital surface model (altitude)	2%	7%
Lake depth	9.6%	5.2%
Human population density (per 2011 U.S. census)	41.7%	58%
Area under the receiver operating characteristic curve (AUC) value for the training data	0.92	0.90
Calibrated AUC (cAUC) values quantifying the magnitude of spatial sorting bias	0.81	0.71

Among the pairwise waterbody specific variables used in the cokriging models, the variable pair with highest AUC value (0.78) for the model for zebra mussels were the Eigenvector centrality of the watercraft movement network and the distance to the nearest major road. Whereas, for Eurasian watermilfoil, the best predictor variables were Eigenvector centrality of the watercraft movement network and the surface water connectivity to infested waterbodies (AUC=0.76) (Kanankege et al., 2018b) (Table 2.7).

Output maps for both Maxent and cokriging methods and the number of waterbodies classified under each category are seen in Fig 2.7 and 2.8. The AUC values, for the k fold cross validations and the AUC values, sensitivity, and specificity for the true validation of Maxent, cokriging, and combined interpretation of the models using risk matrices are summarized in Table 2.8. As of 2016, there were 1490 (8.09%) waterbodies at high risk of invasion by zebra mussels, respectively. Similarly, there were 2036 (11.06%) waterbodies at high risk of invasion by Eurasian watermilfoil as of 2016.

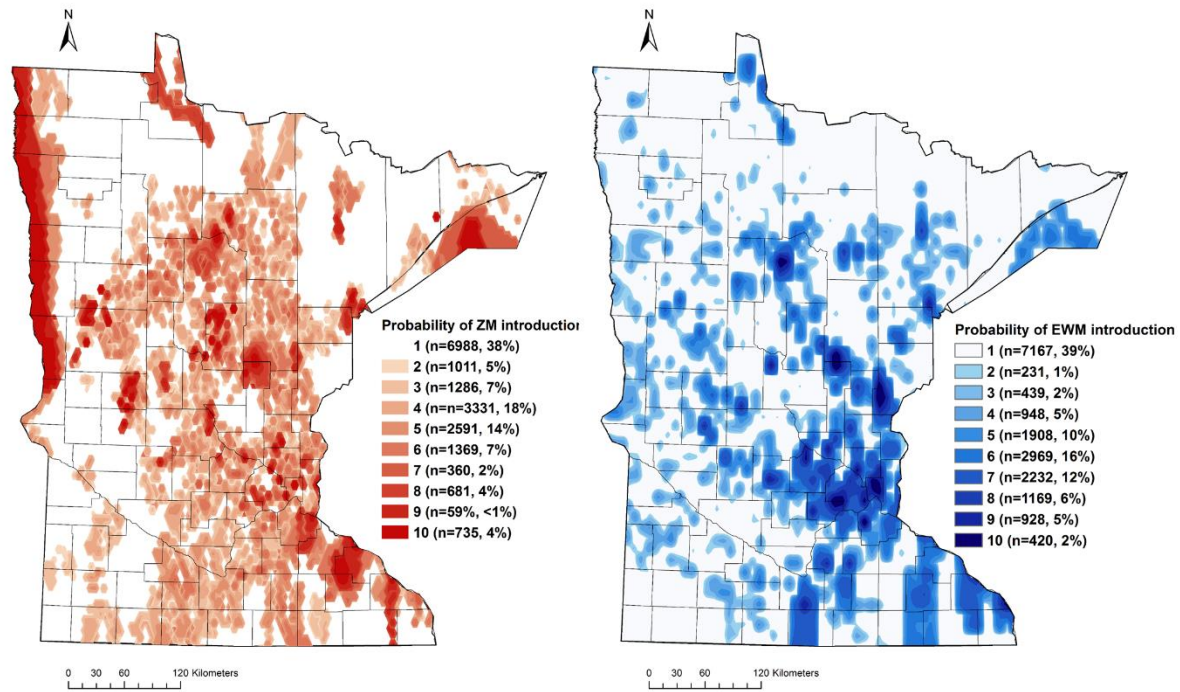
**Table 2.7.** Summary of model validations for both Maxent, cokriging and combined interpretation of the models using risk matrix. Cross validation was done using the K fold test. True validation was done by fitting models for invasions as of 2010 and validating using the invasions reported after 2010. Area under the receiver operating characteristic curve (AUC), Sensitivity (Se) and Specificity (Sp) at the risk rank  $\geq 3$  for Maxent and cokriging, and combined interpretation using the risk matrix are listed below. Combined interpretation was only subjected to true validation.

		Maxent			Cokriging ( <a href="#">Kanankege et al., 2018b</a> )			Combined interpretation using risk matrix		
		AUC	Se	Sp	AUC	Se	Sp	AUC	Se	Sp
Cross validation	Zebra mussels	0.94			0.72			N/A		
	Eurasian watermilfoil	0.93			0.79			N/A		
True validation	Zebra mussels	0.7	0.71	0.01	0.78	0.78	0.72	0.77	0.77	0.72
	Eurasian watermilfoil	0.7	0.79	0.31	0.76	0.83	0.61	0.78	0.77	0.67

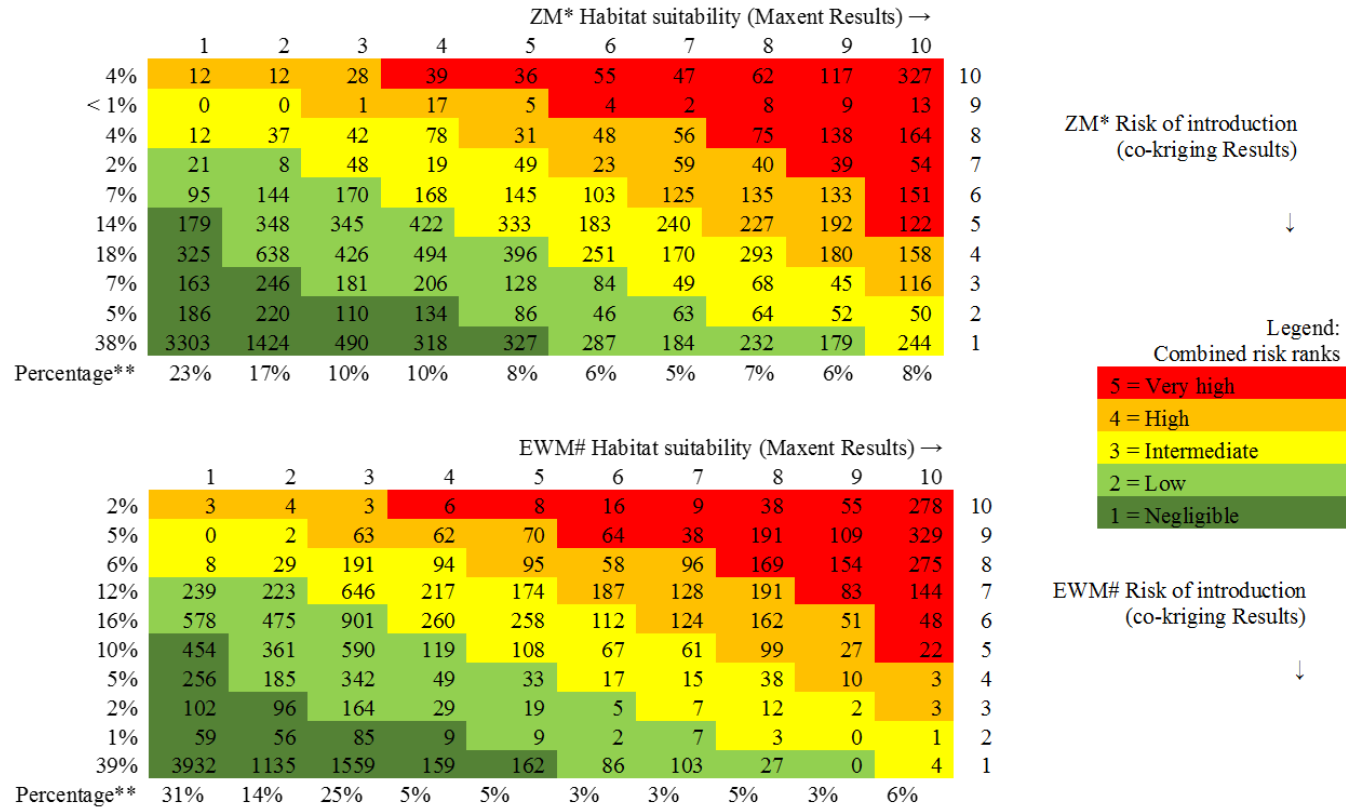


**Fig. 2.7.** MaxENT model outputs illustrating the habitat suitability of Minnesota for zebra mussels (ZM) and Eurasian watermilfoil (EWM). The predicted risk is for waterbodies only. The risk classes 1 through 10 indicates the intensity of the suitability of the locations, where, class 10 represent most suitable areas. The number of waterbodies under each category and as a percentage of the total waterbodies (n=18,411) is listed. The polygons represent counties of the state of Minnesota.





**Fig. 2.8.** Co-kriging model outputs illustrating the risk of introducing zebra mussels (ZM) and Eurasian watermilfoil (EWM) to Minnesota waterbodies. The risk classes 1 through 10 indicates the intensity of the probability of introduction, where, class 10 represent a high probability of ZM or EWM introduction. The number of waterbodies under each category and as a percentage of the total waterbodies (n=18,411) is listed. The polygons represent counties of the state of Minnesota.



**Table 2.8.** Risk matrices for (1) zebra mussels (ZM)\* and (2) Eurasian Watermilfoil (EWM)# invasion risk using the Maxent and co-kriging model outputs. Each model output was categorized into 10 classes where 1 represents the lowest risk, and 10 represent the highest risk. The number of waterbodies which has each of the risk class combination was illustrated in each cell of the matrices. Percentage\*\* are the marginal percentages i.e. the number of waterbodies in each category compared to the total number of waterbodies (n = 18,411). Combined risk of invasion is represented by the color scheme, where red and dark-green represent very high-risk and negligible risk of invasion, respectively.

## 2.22 Discussion

This study was aimed at describing and predicting the potential geographical range expansion of zebra mussels and Eurasian watermilfoil in Minnesota, in relation to the environmental preferences and probability of introduction, using species distribution (MaxENT) and geostatistical (co-kriging) modeling techniques, respectively. The outputs of the MaxENT and co-kriging models were interpreted in parallel to represent current (2016) status of invasion risk. According to the study, 1490 (8.09%) waterbodies are at risk for the invasion by zebra mussels or 2036 (11.06%) by Eurasian watermilfoil. Several waterbodies were at risk for the invasion by both the AIS, hence, the combined results suggested that nearly 15% of the waterbodies in Minnesota were at high risk of invasion by both AIS.

According to the MaxENT outputs, the human population density plays a significant role in determining the probability of both zebra mussels and Eurasian watermilfoil invasions. As described in the published literature, microscopic larval stages of zebra mussels and fragments of Eurasian watermilfoil are known to be transported along with human-mediated recreational activities (Roley and Newman, 2008; Banha et al., 2016).

The annual mean temperature (19.1%) and precipitation of the wettest month (16.7%) were the key climatic predictors in the MaxENT models for zebra mussels. When considering the lifecycle of zebra mussels, the optimum temperature between 10 – 15°C along with availability of food is suggested to support spawning and larval growth of zebra mussels (Chase et al., 1999). Per the data used here, the average annual temperature of Minnesota is ~10°C and ranges between -18°C and 33°C, stipulating appropriate temperatures for the survival of zebra mussels. The precipitation of the wettest month denotes potentially high water flows and, therefore, increases in catchment nutrient runoff which may have promote proliferation of zebra mussels (Taner MU, 2011). Past studies on zebra mussel invasions indicate that climate alone perhaps is a poor predictor of their distribution, while water chemistry including pH, calcium, magnesium, and chloride levels, and concentrations of total dissolved solids may also play a role defining the suitability of waterbodies for zebra mussel establishment (Karatayev et al., 2015). In the absence of water chemistry data for all the waterbodies in Minnesota, we incorporated the bedrock geology into the MaxENT models. Bedrock geology was incorporated because the soil type is considered as a determinant of water chemistry (Mellina and Rasmussen, 1994). However, the contribution of bedrock geology was 1.4% in the MaxENT model for zebra mussels, indicating that bedrock geology may not be a useful proxy to represent relevant water chemistry variables. It

would be more appropriate to collect field data regarding the water chemistry from the invaded waterbodies. Nevertheless, the counter argument that water chemistry may have been modified by the invasions need to be considered when incorporating the field data into predictive models (Karatayev et al., 2015).

The MaxENT model for Eurasian watermilfoil suggests that precipitation of the wettest month (17.6%) was the key climatic predictor contributing to determine the environment suitability for Eurasian watermilfoil. Eurasian watermilfoil is an aquatic plant which propagates by both sexual and vegetative means; runners and fragmentation are considered the primary methods of vegetative reproduction (Madsen et al., 2012). High precipitation results in increased volume and flow rates of the rivers, and this could be a way of explaining the relevance of precipitation of the wettest month in determining the environmental suitability for Eurasian watermilfoil. Moreover, the increase of nutrient runoff due to high precipitation has been suggested to support the proliferation of Eurasian watermilfoil (Bosch et al., 2009). Results demonstrate that the risk of Eurasian watermilfoil introduction and the establishment was less in Northeastern Minnesota. Previous studies suggest that waterbodies in Northeastern Minnesota tend to have comparatively deeper lakes, water with low alkalinity, and low human population densities, factors which tend to reduce the survival of Eurasian watermilfoil (Roley and Newman, 2008).

Species distribution models have certain limitations. The issue of pseudo-absences is one of them, where the locations with favorable environmental conditions were not invaded by the AIS due to biotic interactions, dispersal limitations, or effective preventive measures (Elith, 2011). In the context of this study, it is assumed that the waterbodies that are not reported with zebra mussels or Eurasian watermilfoil are free from those AIS, however, those are pseudo-absences, given that there is no active surveillance looking for AIS in waterbodies to confirm freedom from AIS.

The best fitted co-kriging models suggest that distance to the nearest infested location and boater movement between waterbodies play key roles in determining AIS dispersal. The zebra mussel co-kriging model further confirms that the distance to a major road is an important risk factor, suggesting the zebra mussel distribution is largely a human-mediated process and measures should be taken to mitigate hitchhiking of aquatic invasive species (Seekamp et al., 2016). Findings from the Eurasian watermilfoil co-kriging suggest that their distribution is attributable to both human-mediated transportation and natural dispersal via surface water.

In addition to the incorporation of robust geospatial techniques to assess the risk of invasion, there are several strengths of this study's approach as compared to a very similar study conducted

in Wisconsin, USA (Stewart-Koster et al., 2015). For example, the boater movement between waterbodies in this study was calculated for each waterbody, therefore, is more informative compared to the representation of boater movement by county centroids such as the study by previous studies (Buchan and Padilla, 2000). In addition to the boater traffic data incorporated into the network, the distance to the nearest major road was used as a proxy to represent the boater access making the human-mediated dispersal well represented. Because many (49.6%) of the Watercraft Inspection Program survey locations were infested with either zebra mussels or Eurasian watermilfoil, the boater traffic data may have been subjected to selection bias. Distance to the nearest major road was incorporated into the co-kriging model to adjust for that potential bias.

One of the limitations of the study here includes the lack of AIS distribution data in the states adjacent to Minnesota to implement effective cross-boundary control and preventive measures. For example, waterbodies in east-central Minnesota are affected by both zebra mussels and Eurasian watermilfoil. However, the study described by Stewart-Koster et al. 2015 indicated low risk of zebra mussel and Eurasian watermilfoil invasion across the border in the Northeastern Wisconsin (Stewart-Koster et al., 2015). Our study does not account for the zebra mussels and Eurasian watermilfoil invasions in the adjacent states either, which indicates the risk of invasion may have been underestimated. Being confined within the political boundaries often results in a shift of model accuracies for the species with fewer reported occurrences and non-identified geographic ranges (Barnes et al., 2014). The geographical area for the analysis was not expanded to the Midwest or great lakes because required data are not readily available from all the locations.

## 2.23 Conclusions

A combined approach of co-kriging and maximum entropy ecological niche (Maxent) model were used to predict the risk of AIS distribution and, ultimately, to inform the targeted surveillance activities. Co-kriging was used to calculate the probability of introduction of the two AIS, using waterbody-specific variables. Whereas, Maxent was used to determine the ecological susceptibility of the waterbodies. Thus, the combined interpretation of Maxent and co-kriging results serves as a proxy for the probability of AIS infestation in waterbodies of Minnesota. Combined results from the Maxent and co-kriging outputs suggested that nearly 15% of the waterbodies in Minnesota were at high risk of invasion by either zebra mussels or Eurasian watermilfoil.

## 2.24 Acknowledgments

This study was funded in part by the Minnesota Discovery, Research, and Innovation Economy (MnDRIVE) program and Office of the Vice President for Research (OVPR) of the University of Minnesota. We thank the Aquatic Invasive Species Prevention program of the Minnesota Department of Natural Resources for providing the incident data and sharing their experience.

## 2.25 References

1. Alkhamis MA, VanderWaal KL. 2016. Spatial and Temporal Epidemiology of Lumpy Skin Disease in the Middle East, 2012-2015. *Frontiers in Veterinary Science*; 3:19.
2. Baldwin RA. 2009. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy*. Dec;11:854-866.
3. Banha F, Gimeno I, Lanao M, Touya V, Duran C, Peribanez MA, Anastacio PM. 2016. The role of waterfowl and fishing gear on zebra mussel larvae dispersal. *Biol. Invasions*. 18, 115-125.
4. Barnes MA, Jerde CL, Wittmann ME, Chadderton WL, Ding JQ, Zhang JL, Purcell M, Budhathoki M, Lodge DM. 2014. Geographic selection bias of occurrence data influences transferability of invasive *Hydrilla verticillata* distribution models. *Ecol. Evol.* 4, 2584-2593.
5. Bobeldyk AM, Bossenbroek JM, Evans-White MA, Lodge, DM, Lamberti GA. 2005. Secondary spread of zebra mussels (*Dreissena polymorpha*) in coupled lake-stream systems. *Ecosci.* 12, 339-346.
6. Bosch I, Makarewicz JC, Bonk EA, Ruiz C, Valentino M. 2000. Responses of lake macrophyte beds dominated by Eurasian watermilfoil (*Myriophyllum spicatum*) to best management practices in agricultural sub-watersheds: Declines in biomass but not species dominance. *J Great Lakes Research*; 35: 99-108.
7. Bossenbroek JM, Johnson LE, Peters B, Lodge DM. 2007. Forecasting the expansion of zebra mussels in the United States. *Conservation Biology*. Jun;21:800-810.

8. Buchan LAJ, Padilla DK. 1999. Estimating the probability of long-distance overland dispersal of invading aquatic species. *Ecological Applications*. Feb;9:254-265.
9. Buchan LAJ, Padilla, DK. 2000. Predicting the likelihood of Eurasian watermilfoil presence in lakes, a macrophyte monitoring tool. – *Ecol. Appl.* 10, 1442-1455.
10. Chase ME, Bailey RC. 1999. The ecology of the zebra mussel (*Dreissena polymorpha*) in the lower Great Lakes of North America: I. Population dynamics and growth. *Journal of Great Lakes Research*; 25: 107-121.
11. CRARM: The Presidential/ Congressional Commission on Risk Assessment and Risk Management. 1997. Framework for Environmental Health Risk Management. Commission on Risk Assessment and Management, Washington DC. Vol. 1.
12. DEM: Minnesota LiDAR based Digital Elevation Model. 2012. 30-meter resolution. Published by Management Resources Management Information Systems Unit of Minnesota Department of Natural Resources. ID: elev\_dem03lbra4.
13. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecograph*; 36: 27-46.
14. Drake JM, Bossenbroek JM. 2009. Profiling ecosystem vulnerability to invasion by zebra mussels with support vector machines. *Theoretical Ecology*. Dec;2:189-198.
15. Dufour B, Plee L, Moutou F, Boisseleau D, Chartier C, Durand B, et al. 2011. A qualitative risk assessment methodology for scientific expert panels. *Revue Scientifique Et Technique-Office International Des Epizooties*; 30: 673-681.
16. Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*; 17: 43-57.
17. Havel JE, Kovalenko KE, Thomaz SM, Amalfitano S, Kats LB. 2015. Aquatic invasive species: challenges for the future. *Hydrobiol.* 750,147-170.

18. Hijmans RJ, Elith J. 2016. Species distribution modeling with R. 0.9-1. [cited: June 20, 2016]. Available from: <https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf>.
19. Hijmans RJ. 2012. Cross validation of species distribution models: removing spatial sorting bias and calibration with null model. *Ecology*; 93: 679-688.
20. Invasive Species Program MNDNR Annual Report: Invasive Species of Aquatic Plants and Wild Animals in Minnesota; Annual Report for 2015, Minnesota Department of Natural Resources, St. Paul, MN. Minnesota Department of Natural Resources Annual report of Invasive species of Minnesota of 2015. 2016. Available from: [http://files.dnr.state.mn.us/natural\\_resources/invasives/ais-annual-report.pdf](http://files.dnr.state.mn.us/natural_resources/invasives/ais-annual-report.pdf)
21. Isaaks EH, Srivastava RM. 1989. *Applied Geostatistics*. Oxford University Press. New York. ISBN-13: 978-0195050134.
22. Jirsa MA, Boerboom TJ, Chandler VW, Mossler JH, Runkel AC, Setterholm DR. 2011. S-21 Geologic Map of Minnesota-Bedrock Geology. Minnesota Geological Survey. Retrieved from the University of Minnesota Digital Conservancy, <http://purl.umn.edu/101466>.
23. Kanankege KST, Alkhamis MA, Perez AM, Phelps NBD. 2018a. Zebra mussels and Eurasian watermilfoil reporting patterns in Minnesota. *J. Great Lakes Res.* doi: 10.1016/j.jglr.2018.02.008
24. Kanankege KST, Alkhamis MA, Phelps NBD, Perez AM. 2018b. A probability co-kriging model to account for reporting bias and recognize areas at high risk for zebra mussels and Eurasian watermilfoil invasions in Minnesota. *Frontiers in Veterinary Science*. doi: 10.3389/fvets.2017.00231
25. Karatayev AY, Burlakova LE, Mastitsky SE, Padilla DK. 2015. Predicting the spread of aquatic invaders: insight from 200 years of invasion by zebra mussels. *Ecol. Appl.* 25, 430-440.
26. LandScan™. 2012. High Resolution Global Population Data. UT-Battelle, LLC. Oak Ridge National Laboratory. Contract No. DE-AC05-00OR22725. U.S. Department of Energy. <http://www.ornl.gov/landscan/>.



27. Latzka AW, Hansen GJA, Kornis M, Vander Zanden MJ. 2016. Spatial heterogeneity in invasive species impacts at the landscape scale. *Ecosphere*. Mar;7:15.
28. Lodge, DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA, Carlton JT, McMichael A. 2006. Biological invasions: Recommendations for US policy and management. *Ecol. Appl.* 16, 2035-2054.
29. Lund K, Bloodsworth K, Wolbers T, Weling C, Gamble A. 2015. Guidance for conducting aquatic invasive species early detection and baseline monitoring in lakes. Invasive Species Program. Division of Ecological and Water Resources of the Minnesota Department of Natural Resources. Available from:  
[https://www.ifound.org/files/6714/4745/1209/ais\\_detection-baseline-monitoring.pdf](https://www.ifound.org/files/6714/4745/1209/ais_detection-baseline-monitoring.pdf)
30. Madsen JD, Richardson RJ, Wersal RM. 2012. Managing Aquatic Vegetation. Small Impoundment Management in North America. In Neal JW, Willis DW, editors. American Fisheries Society, Bethesda. pp 275-305
31. Mandrak NE, Cudmore B. 2015. Risk assessment: Cornerstone of an aquatic invasive species program. *Aquatic Ecosystem Health and Management*; 18: 312-320.
32. Mandrak NE, Cudmore B. 2015. Risk assessment: Cornerstone of an aquatic invasive species program. *Aquat. Ecosys. Heal. Manage.* 18, 312-320.
33. Mellina E, Rasmussen JB. 1994. Patterns in the distribution and abundance of zebra mussel (*Dreissena polymorpha*) in rivers and lakes in relation to substrate and other physicochemical factors. *Can. J. Fisheries Aquat. Sci.* 51, 024-1036.
34. Merow C, Smith MJ, Silander Jr. JA. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*; 36: 001-012.
35. Minnesota Statute 477A.19. 2016. Aquatic Invasive Prevention Aid. The office of the Reviser of the Statutes. Chapter 308. Article 1. Section 11. Available from:  
<https://www.revisor.mn.gov/statutes/?id=477A.19&format=pdf>.

36. MNDNR AIS: Minnesota Department of Natural Resources: Aquatic Invasive Species. 2016. Available from: <http://www.dnr.state.mn.us/invasives/ais/infested.html>
37. MNDNR WIP: Watercraft Inspection Program of the Minnesota Department of Natural Resources. 2014. Available from: [http://files.dnr.state.mn.us/natural\\_resources/invasives/mndnr\\_ais\\_watercraft\\_inspection\\_hanbook.pdf](http://files.dnr.state.mn.us/natural_resources/invasives/mndnr_ais_watercraft_inspection_hanbook.pdf).
38. MNDR: Minnesota Department of Revenue. Aquatic Invasive Prevention Aid. 2016. [cited: June 10, 2016]. Available from: [http://www.revenue.state.mn.us/local\\_gov/prop\\_tax\\_admin/Pages/ais.aspx](http://www.revenue.state.mn.us/local_gov/prop_tax_admin/Pages/ais.aspx).
39. MNGSC Hydrography: Minnesota GeoSpatial commons. MNDNR Hydrography data layer. 2015. [cited August 15, 2015]. Available from: <https://gisdata.mn.gov/dataset/water-dnr-hydrography>.
40. MNGSC: Minnesota GeoSpatial commons. 2015. Lake Bathymetric Outlines, Contours, Vegetation, and DEM: Lake Bathymetric Digital Elevation Model (DEM) data layer. Published by MNDNR – Fisheries department. ID: water\_lake\_bathymetry. Last updated: April 24, 2015. [cited: May 15, 2016]. Available from: <https://gisdata.mn.gov/dataset/water-lake-bathymetry>.
41. Olson JR, Hawkins CP. 2012. Predicting natural base-flow stream water chemistry in the western United States. *Water Resources Research*. Feb;48.
42. Papes M, Havel JE, Vander Zanden MJ. 2016. Using maximum entropy to predict the potential distribution of an invasive freshwater snail. *Freshwater Biology*. Apr;61:457-471.
43. Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.
44. Pysek P, Richardson, DM. 2010. Invasive Species, Environmental Change and Management, and Health. Pages 25-55 in A. Gadgil and D. M. Liverman, editors. *Annual Review of Environment and Resources*, Vol 35. Annual Reviews, Palo Alto.

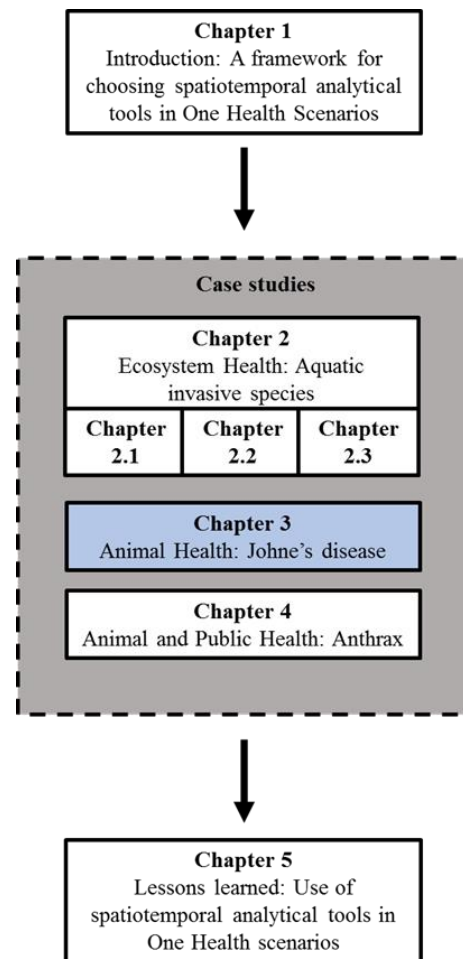
45. Quinn A, Gallardo B, Aldridge DC. 2014. Quantifying the ecological niche overlap between two interacting invasive species: the zebra mussel (*Dreissena polymorpha*) and the quagga mussel (*Dreissena rostriformis bugensis*). *Aquatic Conservation-Marine and Freshwater Ecosystems*. Jun;24:324-337.
46. Read JS, Winslow LA, Hansen GJA, Van den Hoek J, Hanson PC, Bruce LC, et al. 2014. Simulating 2368 temperate lakes reveals weak coherence in stratification phenology. *Ecological Modelling*; 291:142-150. doi: 10.1016/j.ecolmodel.2014.07.029
47. Roley SS, Newman, RM. 2008. Predicting Eurasian watermilfoil invasions in Minnesota. *Lake Reserv. Manage.* 24, 361-369.
48. Seekamp E, McCreary A, Mayer J, Zack S, Charlebois P, Pasternak L. 2016. Exploring the efficacy of an aquatic invasive species prevention campaign among water recreationists. *Biological Invasions*. 18: 1745-1758.
49. Stewart-Koster B, Olden JD, Johnson PTJ. 2015. Integrating landscape connectivity and habitat suitability to guide offensive and defensive invasive species management. *J Appl. Ecol.* 52, 366-378.
50. Swets JA. 1988. Measuring the accuracy of diagnostic systems. *Science*. 240, 1285–1293.
51. Taner MU, Carleton JN, Wellman M. 2011. Integrated model projections of climate change impacts on a North American lake. *Ecological Modelling*; 222: 3380-3393.
52. Vander Zanden MJ, Olden JD. 2008. A management framework for preventing the secondary spread of aquatic invasive species. *Can. J. Fisheries Aquat. Sci.* 65, 1512-1522.
53. Wackernagel H. 1994. Co-kriging versus kriging in regionalized multivariate data analysis. *Geoderma*; 62: 83-92.
54. Yadav SK, Chakrapani GJ. 2006. Dissolution kinetics of rock-water interactions and its implications. *Current Science*. Apr;90:932-937.

## 2.26 Supplement Tables and Figures

**Supplement Table S2.2.** Environmental variables used in the final MaxENT models.

Variable	Percent contribution of the variable in the MaxENT*		
	ZM <sup>#</sup>	EWM <sup>‡</sup>	Both <sup>†</sup>
Annual mean temperature	19.1%		0.8%
Isothermality (Mean diurnal range of temp / temp annual range)	2.7%		
Temperature seasonality (Standard deviation * 100)		3.3%	
Mean temperature of warmest quarter		6.2%	
Mean Temperature of coldest quarter			0.2%
Annual precipitation		2.5%	1%
Precipitation of wettest month	16.7%	17.6%	0.1%
Precipitation of driest month	6.8%		
Bedrock geology map of Minnesota	1.4%	0.2%	0.6%
Minnesota Digital Elevation Model - 30 Meter Resolution	2%	7%	
Lake depth	9.6%	5.2%	9.5%
Human population density (according to 2011 U.S. census)	41.7%	58%	87.6%
Area under the receiver operating characteristic curve (AUC) value for the training data	0.92	0.90	0.93
AUC for the testing data	0.94	0.93	0.87
Calibrated AUC (cAUC) values quantifying the magnitude of spatial sorting bias	0.81	0.71	0.86

ZM<sup>#</sup>=zebra mussels, EWM<sup>‡</sup>=Eurasian watermilfoil, and Both<sup>†</sup>=Inhabitation of both zebra mussels and Eurasian watermilfoil



### Chapter 3: Animal health

*The use of Getis Ord and Conditional autoregression as tools to study the spatial epidemiology of Johne's disease in Minnesota: A cross-sectional study based on a voluntary testing program*

**In review: BMC Veterinary Research (October 2018)**

Kanankege K, Machado G, Zhang L, Dokkebakken B, Schumann V, Wells S, Perez A, Alvarez J. Use of a voluntary testing program to study the spatial epidemiology of Johne's disease in Minnesota: A cross-sectional study.

### 3.1 Overview

One of the key steps in the management and control of chronic diseases, such as Johne's disease (JD), caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP), is the ability to track disease incidence over time. JD surveillance in U.S. dairy cattle is challenging due to lack of regulatory requirements, imperfect diagnostic tests, and associated expenses, including time and labor. An alternative approach is to use a voluntary testing programs to monitor JD. In this cross-sectional study, data from a voluntary JD testing program conducted by the Minnesota Dairy Herd Improvement Association were used to: a) explore whether such program provides representative information on the prevalence of JD in dairy herds in Minnesota, b) estimate JD distribution in Minnesota, and, c) identify herd and environmental factors associated with finding JD-positive cattle. Milk samples (n=70,809) collected from 54,652 unique cows from 600 Minnesota dairy herds between November 2014 and April 2017 were tested using a MAP antibody ELISA. Participant representativeness was assessed by comparing the number of JD-tested herds with the number of herds required to estimate the true disease prevalence per county based on official statistics from the National Agricultural Statistical Services. Multivariable logistic regression models, with and without spatial dependence between observations, were then used to investigate the association between herd status to JD (positive/negative), as indicated by milk ELISA results, and available covariates at the herd level. Within the study population, at least one test-positive cow was found in 414 of 600 (69%) herds. Regression results suggests that large herds that test frequently and are in loamy or silt soils are more likely to have at least one MAP test-positive cow. Residuals from a conditional autoregressive model indicated that, after adjusting for those variables, there was no spatial dependence in JD risk between neighboring dairies within 5 to 20Km. This applied study exemplifies the potential use of Dairy Herd Improvement Association's voluntary JD testing databases as a passive surveillance tools to monitor JD status and its association with epidemiological factors while identifying opportunities for improvement of the database.

### 3.2 Introduction

Johne's disease (JD) is a chronic, debilitating enteritis affecting ruminants caused by infection with *Mycobacterium avium* subsp. *paratuberculosis* (MAP) (Collins et al., 2003). Financial losses due to JD in the U.S. dairy industry have been estimated to range between \$200 to 250 million USD annually (Ott et al., 1999). Reduced milk production and quality due to reduced fat and protein content, increased premature mortality, weight loss, early culling, costs of testing and control, and reduced slaughter value are among the negative impacts of JD (Gonda et al., 07; Raizman et al., 07; Lombard, 2011; Garcia and Shalloo, 15). In addition, although not confirmed, the potential link between MAP and the development of Crohn's disease in humans further increases the hypothetical importance of JD (Waddell et al., 2016).

The management and control of a chronic disease such as JD in a proactive and organized way in the U.S. is challenging due to the lack of regulatory requirements for testing (USDA-APHIS, 2015), imperfect diagnostic tests (Nielsen and Toft, 2008), long-term survival of the pathogen outside the host (Whittington et al., 2005; Grewal et al., 2006), multiple routes of transmission, and the cost and labor necessary for long-term disease tracking (McKenna et al., 2006).

JD is widespread in the U.S, and herd prevalence has been estimated in 60.7% in Midwestern U.S. dairies (USDA 97) and 91.1% nationally (Lombard et al., 2013). However, because JD control is voluntary in the US, testing for JD is not mandatory (USDA-APHIS, 2017) which limits the availability of data and resources to monitor the disease. Studies in the U.S. State of Texas and Canada suggest that limited adoption and compliance with JD testing and control strategies in dairy farms is a result of a) the chronic nature of the disease progression, therefore, the absence of the "cues-to-action" (Benjamin et al., 2010), b) the farmers' perception of the limited cost-effectiveness of the herd control measures (Ritter et al., 2015), and c) not perceiving JD as a "hot topic" during communications with other farmers and veterinarians (Ritter et al., 2015).

Due to the lack of official disease monitoring, a common alternative for evaluating the epidemiological status of JD in a region is the use of data from voluntary testing programs, such as those collected by the Minnesota Dairy Herd Improvement Association (MNDHIA). Minnesota, a Midwestern state of the U.S., has nearly 460,000 dairy cattle and is among the top ten dairy state, 6<sup>th</sup> among the milk cow numbers and 8<sup>th</sup> in dairy herds per 2016 statistics (NASS 2018). A proportion of Minnesota dairy farms utilize the services of MNDHIA, a member of the

National Dairy Herd Improvement Association who service a testing and production recordkeeping system to U.S. dairy farms (<http://www.dhia.org/members.asp>). However, it is unknown if data collected by the MNDHIA is useful as a passive surveillance tool to monitor JD in Minnesota.

The objectives of the study here were to a) test whether the voluntary JD testing program conducted by the MNDHIA can provide representative information on the prevalence of JD in dairy herds in Minnesota, b) estimate the JD distribution in Minnesota using data from the MNDHIA voluntary JD testing program, and c) identify possible herd and environmental variables associated with increased risk of having JD test-positive cows, using the available data. We hypothesized that results from the voluntary JD program might be used to evaluate JD status in an area and inform management decisions made by the testing agencies, veterinarians, and dairy producers. In evaluating the use of the MNDHIA database as a passive surveillance tool, our overarching objective was to generate evidence that could influence management decisions by recognizing modifiable factors to reduce the JD risk at the individual, herd, and regional levels. Results can therefore help to design and implement effective surveillance programs for the U.S. dairy industry.

### **3.3 Data and Methods**

#### **3.3.1 Data**

Data from the voluntary JD testing program conducted by the MNDHIA were collected through a 2.5-year period, between November 1st, 2014 and April 30th, 2017. Although there were records from 723 JD testing herds, 123 herds were excluded from the study (Supplement Fig. S3.1). The reasons for exclusions were: herds located outside Minnesota; herds without location information, or herds that did not test the minimum number of cows required for the study, as explained below.

Milk samples (n=70,809) collected from 54,652 unique cows from 600 Minnesota dairy herds were included in this cross-sectional study. Samples were analyzed using the IDEXX MAP Enzyme-Linked Immunosorbent Assay (ELISA) (IDEXX Laboratories, Inc., Maine, USA) for detection of antibodies against *Mycobacterium avium* subsp. *paratuberculosis* (MAP) in milk.



Some herds (n=208, 35%) were tested only once during the 2.5-year study period. The median number of times a herd was tested was 2 (interquartile range between 1 and 11).

To identify relevant herd and environmental factors associated with JD risk at the herd level, the scientific databases ‘Web of Science’ and ‘PubMed’ were queried to find publications using the following search string: ‘Johne’s disease,’ AND ‘*Mycobacterium avium* subsp. *paratuberculosis*,’ AND ‘dairy cattle,’ AND ‘risk factors’. Reviews published in peer-reviewed journals in English language were selected, and reviews on human Crohn’s disease were excluded. A total of seven reviews published between 2001 and 2017 on JD (Kennedy and Benedictus, 2001; Tiwari et. al., 2006; McKenna et. al., 2006; Dore et. al., 2012; Elliott et. al., 2015; Garcia and Shalloo 2015; Rangel et al., 2015) were used to identify the most commonly recognized JD risk factors in North American dairy farms. Reviews were then examined for identification of relevant risk factors for North America, and primary articles cited there were used for identification of the variables to be considered in this study (Table 3.1). Herd features related to management practices including manure management, immediate culling of JD positive cows, management of the maternity pens and calves, and maintaining closed herd or purchasing animals from farms with improved management practices to control Johne’s disease (McKenna et. al., 2006) were excluded from analysis due to unavailability of relevant data in our dataset.

**Table 3.1.** Herd demographic factors and environmental factors associated with JD in North American dairy cattle, according to the published literature.

	<b>Covariates</b>	<b>Reference</b>
1	Herd size	Wells and Wagner, 2000, Scott et. al., 2007; Pillars et al., 2009
2	Testing frequency	USDA-APHIS-VS, 2010
3	Soil type (texture)	Ward and Perez, 2003
4	Soil pH	Kopecky, 1977; Johnson-Ifearegundun and Kaneene, 1999
5	Soil hydrologic characteristics i.e. run-off potential	Ward and Perez, 2003; Pickup et al., 2006
6	Agroecological zone (Montane/Boreal forest/Grassland/Parkland)	Scott et. al., 2007

7	Participating the USDA Voluntary Johne's control program i.e. VJDHSP (n=24)	<a href="#">USDA-APHIS, 2015; Collins et al., 2010</a>
8	Spatial dependence: Presence of neighboring dairy farms	<a href="#">Collins et al., 1994</a>

According to the published literature, herd size, testing frequency, and geographical region are associated with JD ([USDA-APHIS-VS, 2010; Bihrmann et al., 2012](#)), and information on those three variables was extracted from the MNDHIA database. Data on herd size were available in the form of snapshots at the beginning of 2015, 2016, and 2017. For the analysis, herd sizes were averaged across years. Farm addresses were verified and geocoded using ArcMap version 10.3.4 ([ESRI, 2016](#)). Because the spatial dependence of JD risk has been described for neighboring farms ([Collins et al, 1994](#)), the possible existence of a spatial pattern in the risk of JD was accounted for neighboring farms located within 5 through 20km in the analyses.

The voluntary participation in JD control programs has been also described as a factor associated with JD status in a farm ([Collins et al., 2010; Carter, 2012](#)). In the absence of data from a control program, information on whether farms included in this study were currently participating in the Voluntary Johne's Herd Status Program for Cattle (VJDSHP) program was retrieved from the Minnesota Board of Animal Health ([USDA-APHIS-VS, 2010; USDA-APHIS, 2015](#)). VJDSHP was introduced by USDA APHIS in 2002 as a gradual process which allows herds with low Johne's prevalence to be recognized as herds free from the disease or otherwise low prevalence ([USAHA, 1993; Rideout et al., 2003](#)). As of 2017, among the study population, 24/600 (4%) herds were part of the VDJSHP.

Layers of soil pH and soil type in Minnesota were obtained from the Natural Resource Conservation Services of the United States Department of Agriculture (USDA) ([Soil Survey Staff, 2017](#)). The hydrologic soils data were used to estimate the runoff potential of the soils, using the Hydrologic Soil map available from the Web Soil Survey ([Soil Survey Staff, 2018](#)). Because of the scarcity of accurate data, soil iron content was not considered in the study. Information on agroecological biome features such as grassland, shrubland, forest, and cropland were obtained from the Minnesota GIS Commons, which was based on remote sensing data creating the Land Cover Data Portal under the National Gap Analysis Project (GAP) ([MNGSC GAP, 2017](#)). The Minnesota GAP classification level 2 land-use/land-cover class code of the GAP data layer was used in the analysis (Supplement Table S3.1).

### 3.3.2 Methods

#### *Representativeness of MNDHIA Herds*

Minnesota counties were classified based on the number of NASS dairy herds present into three terciles (1 to 12, 13 to 42, and >43 dairy herds per county). Counties without records in the NASS 2012 statistics for milk cow herds were excluded for this calculation (n=1; Cook county). To evaluate whether the study herds were representative of all Minnesota dairy herds, the number of study herds per county (tested for JD) was compared to the appropriate sample size, i.e. number of dairy herds to be sampled from each county to estimate the true herd prevalence of JD using an imperfect test and adjusting for a finite population (calculated using the AusVet EpiTool Epidemiological calculator (URL: <http://epitools.ausvet.com.au>; [Humphry et. al., 2004](#); [Thrusfield 2005](#)). Total number of herds present in each county was extracted from the National Agricultural Statistical Services (NASS) 2012 Census of dairy herds ([NASS, 2012](#)). In addition, sample size calculations assumed an expected true herd prevalence of JD of 60% ([USDA, 1997](#)), a desired herd-level sensitivity and specificity of 70% ([VanLeeuwen et. al., 2001](#)), a precision for the estimate of +/-10% and a level of confidence of 80%. The number of herds included in the MNDHIA database was then compared to the sample size required to accurately estimate prevalence.

#### *Descriptive statistics and spatial pattern recognition*

The apparent JD prevalence, the spatial distribution of JD test-positive farms, and the presence of spatial autocorrelation in the risk of JD and in other covariates considered were visualized and, for the latter, estimated using Morans' *I* and Getis Ord  $G_i^*$  statistic ([Moran 1948](#); [Getis and Ord, 1992](#); [Ord and Getis, 1995](#)). Morans' *I* statistics measures the overall spatial autocorrelation of the herds based on both locations and value of the variable simultaneously ([Moran, 1948](#)). Whereas, areas where the local sum of values for a given variable significantly differed from the expected location sum were recognized by the Getis Ord  $G_i^*$  statistic ([Getis and Ord, 1992](#); [Ord and Getis, 1995](#)). The Getis Ord  $G_i^*$  statistics recognizes herds with high value of the variable next to herds with high values of the variable (high-high clusters), herds with low value of the variable next to herds with low values of the variable (low-low clusters), and non-matching pairs (high-low, or low-high values). Categorical variables, not suited for Getis Ord  $G_i^*$  analysis, were mapped for visualization.

### ***Classification of JD status***

The output variable (herd JD status) was dichotomized as follows: 1) herds that had sampled an appropriate number of cows to be considered ‘disease-free’ (see below) and had no positive results in any herd test during the study period (negative herds), and 2) herds with 1+ cows with ELISA positive results (positive herds). The appropriate sample size to certify disease freedom in a herd, i.e. minimum required sample size per herd, using an imperfect test and adjusting for a finite population was set for each herd using the AusVet EpiTool Epidemiological calculator (URL: <http://epitools.ausvet.com.au>; [Humphry et. al., 2004](#); [Thrusfield 2005](#)). Assumptions for the calculations were a) design prevalence, i.e. expected within-herd prevalence when infected, of 10% ([Raizman et. al., 2011](#)), and b) sensitivity of the diagnostic test of 52% (and the specificity of 98%) (IDEXX Laboratories Inc., Maine, USA.). Herds that tested fewer cows than the appropriate sample size and in which no positive cows were excluded from the analysis because their apparent JD disease-freedom state could not be reliably demonstrated.

### ***Regression Analysis***

The outcome variable used in all models was farm-level JD test-status (positive, negative). The herd size, which ranged between 1 and 1,929 cows, was categorized using Jenks’ natural breaks method ([Jenks, 1967](#)). The testing frequency per herd during the 2.5 years varied between once and 30 times, and was categorized into four classes (1, 2 to 5, 6 to 20, and >20). Soil pH values ranged between 5.6 and 7.5, and were categorized into three (<6.0, 6.0 to 7.0, and >7.0) groups. The soil type/texture was re-categorized into four classes based on the percentage of different types of particles as clay (>50% clay and ≤50% silt), sand (>50% sand and ≤50% clay), silt (>50% silt and ≤50% sand), and loam (equal proportions of sand, silt and clay, i.e. ≤50% sand, ≤50% clay, and ≤50% silt), following the soil texture triangle model described elsewhere ([Davis and Bennett, 1927](#); [Soil Science Division Staff, 2017](#)). Soil hydrologic characteristics were summarized into four categories: 1) Type A with low runoff potential, 2) Type B with moderately low runoff potential, 3) Type C with moderately high runoff potential, and 4) Type D with high runoff potential, when completely wet ([Soil Survey Staff, 2018](#)). Agroecological biome features included four categories, namely, crop/grassland, non-vegetated land, shrubland, and deciduous forest (Supplement Table S3.1). The VJDHSP participation status was incorporated into the model as a dichotomous variable (current participant vs. non-participant herds).

To avoid multicollinearity, a simple logistic regression was used to assess the marginal association between JD status and the covariates. The strength of the association between pairs of covariates was analyzed using Chi-square test followed by a Cramer's V test. Variables with p-value  $<0.2$  and that were not significantly associated among them (i.e., Chi-square p-values  $>0.05$  and Cramer's V  $>0.5$ ) were tested as candidate variables in the full multivariable logistic regression model. To prevent overfitting, the full model, including all possible 2-way interactions deemed biologically plausible, was subjected to backward stepwise regression based on the lowest Akaike Information Criterion (AIC) until the most parsimonious (final) model was fitted. The regression analysis was carried out using the R Statistical Software (Foundation for Statistical Computing, Vienna, Austria). For the descriptive analysis and logistic regression we used R packages 'Base' (R core team, 2017) and 'MASS' (Venables and Ripley, 2002).

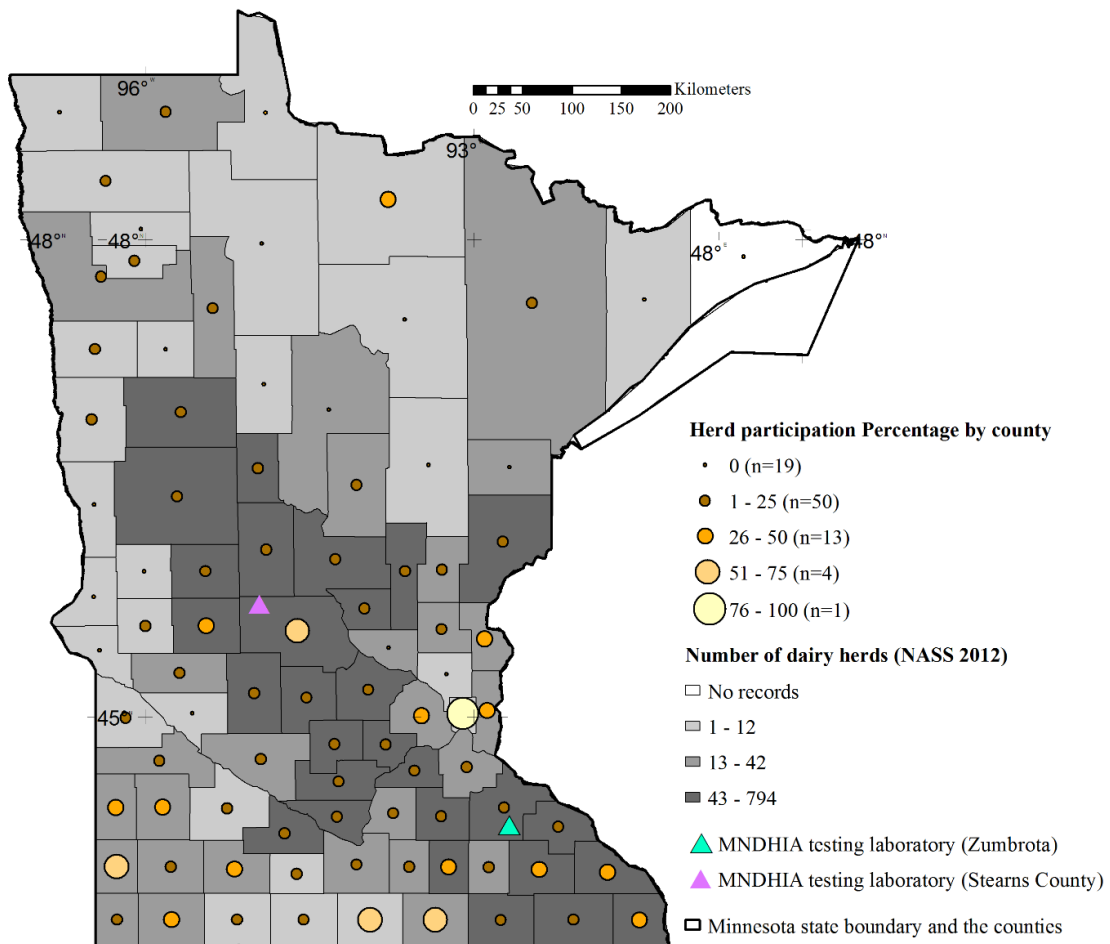
Evidence of spatial autocorrelation in the residuals of the final regression model was assessed using the global Morans  $I$  and local Getis Ord  $G_i^*$  tests. To account for potential spatial dependence in the outcome variable, a proper conditional autoregressive model (CAR) structure was included in the model (Besag, 1974) using the "spdep" R package (Bivand et al., 2013; Bivand and Piras 2015). Distances of 1km, 5km, 10Km, 15km, 20km and the minimum distance between herds that guaranteed all herds had at least one neighbor were tested alternatively to define the neighborhood matrix in the CAR (Banerjee et al., 2015). The isolated herds without a neighbor at each distance thresholds were assigned a lag value of zero at each model fit. The AIC values and significance of the spatial correlation parameter of the CAR model ( $\lambda$ ) from the regression models with and without the CAR model structure were compared to select the best model (Banerjee et al., 2015).

## 3.4 Results

### 3.4.1 Spatial representativeness

During the 2.5-year study period, 600/4,746 (13%) dairy herds in Minnesota tested at least once for JD at MNDHIA laboratories, representing 18.7 % (600/3,210) of the licensed dairy herds in Minnesota with permits to ship milk for human consumption (NASS, 2018). Figure 1 depicts the number of MNDHIA sampled dairy herds by county, the minimum required number of herds to estimate a true JD prevalence, and the number of dairy herds listed in the 2012 NASS

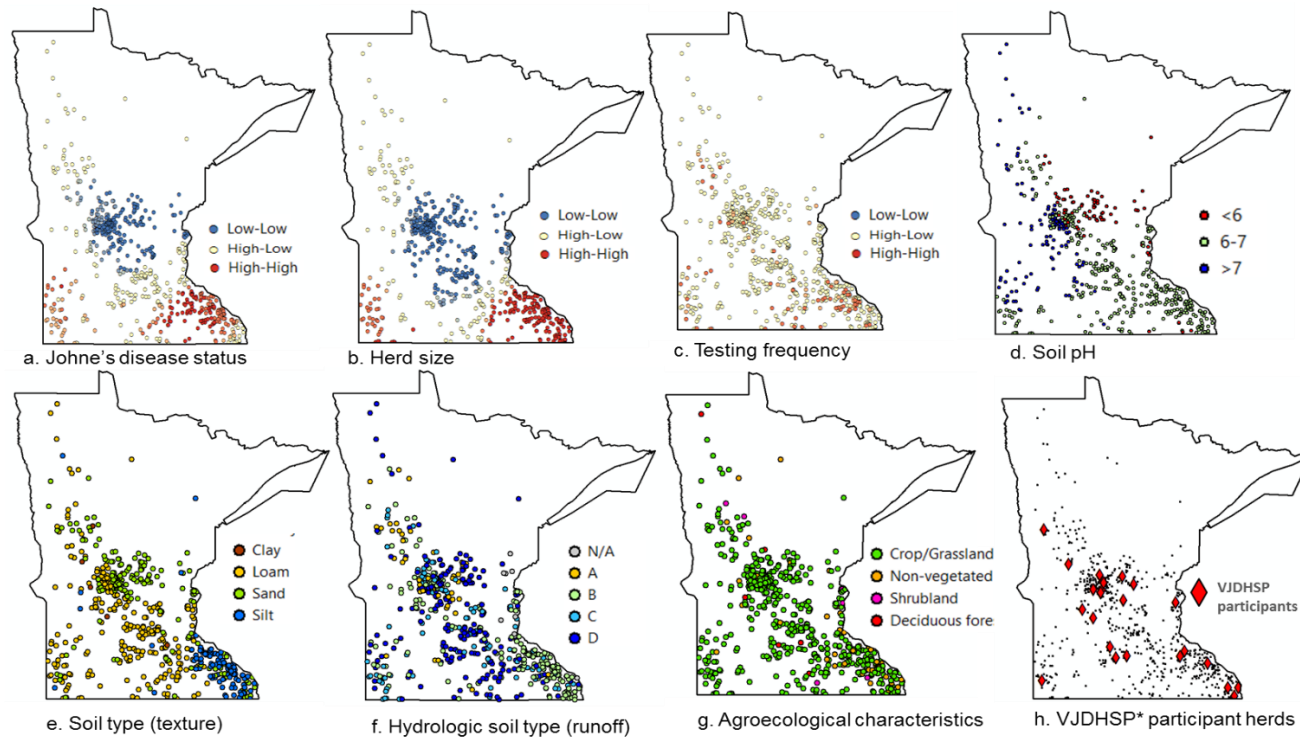
Census of Agriculture (by terciles). We observed that the distribution of the MNDHIA participants included in this study mirrors the pattern of the milk cow herds included in the USDA NASS 2012 report (Figure 3.1). Nevertheless, the minimum sample size required to estimate disease prevalence was not attained in any Minnesota county except Ramsey (where both the appropriate and observed sample sizes were 1).



**Fig. 3.1.** The participation of study herds as a percentage of the ideal sample size, by county, is summarised with the graduated symbols. The Minnesota Dairy Herd Improvement Association testing laboratories are illustrated with triangles. The background indicates the dairy cattle farm density by county, based on 2012 census of the National Agricultural Statistics Service (NASS, 2012).

### 3.4.2 Descriptive statistics and spatial pattern recognition

The apparent herd-level prevalence of MAP, based on having 1+ cows with a positive milk ELISA, was 69% (414/600). The MNDHIA herds in this study included both small (<100 cows, n= 332) and large ( $\geq$  100 cows, n=268) dairies. Figure 2 illustrates the Getis ord  $G_i^*$  local test results for JD status and the covariates in the geographical space. According to the Getis Ord  $G_i^*$  analysis, we concluded that a certain risk of JD infection was present in herds throughout Minnesota, although herds in the southeastern corner were more likely to be JD test-positive compared to the herds in the northcentral region (Fig 3.2: Panel a). The herd size of the study population showed a similar pattern where larger herds were located in the southeastern corner of the state, while smaller herds were in the northcentral region (Fig. 3.2: Panel b). We observed no spatial pattern in the testing frequency (Fig 3.2: Panel c). Spatial patterns were observed for many of the remaining covariates: soil pH, soil type (texture), soil hydrologic characteristics (runoff potential), and agroecological characteristics (Fig. 3.2). The soil pH in Western Minnesota contained predominantly alkaline soils. However, the soil type (texture), hydrologic soil types, and agroecological characteristics demonstrated intricate patterns of spatial distribution. According to the Cramer's V, none of the covariates were strongly associated with each other.



**Fig. 3.2.** Spatial patterns of Johne's disease status and the associated covariates. Points represent the location of study herds (n=600). Panels a, b, and c illustrates the results of the Getis Ord  $G_i^*$  local test where herds with high value of the variable next to herds with high values of the variable are represented in red (high-high clusters), herds with low value of the variable next to herds with low values of the variable represented in blue (low-low clusters), and no-matching pairs in yellow (high-low, or low-high values) (Getis and Ord, 1992; Ord and Getis, 1995). Panels e, f, g, and h simply depict the distribution of covariates.



### 3.4.3 Classification of JD status

Based on the appropriate number, i.e. minimum sample size per herd (Table 3.2) criterion, we observed 437/600 (72.83%) herds that had tested an appropriate number of cows during the study period. Of those, 186 herds (31%) had no test-positive cows.

**Table 3.2.** Minimum sample sizes to be collected required to estimate freedom from Johne's Disease (JD) at the herd level, using an imperfect test and adjusting for a finite population calculated using the AusVet EpiTool Epidemiological calculator (URL: <http://epitools.ausvet.com.au>).

Herd size (number of cows)	Sample size*
5-20	N/A (Recommendation: Same as herd size)
21-25	20
26-40	24
41-55	25
56-65	26
66-80	27
81-95	28
96-100	29
101-250	30
>251	31

\*Assumptions: a) design prevalence, i.e. expected within herd prevalence =10%, b) sensitivity of the diagnostic test of 52%, and c) expected number of test infected cows in a herd is a product of herd size and the design prevalence.

### 3.4.4 Regression results

Herd size, testing frequency, and soil type (texture) were retained in the final multivariable model (Table 3.3). The Moran's I and Getis Ord  $G_i^*$  statistics indicated that there was no spatial autocorrelation in the regression residuals ( $p$ -value > 0.05). Similarly, results of the fitted CAR model suggested that the spatial dependence for dairy herds located between 5 and 20km was not significant after adjusting for the covariates, and its AIC value was higher than that of the model without the spatial component (Table S2). The spatial correlation parameter of the CAR model ( $\lambda$ ) was not significantly different from zero for any of the distance thresholds considered (Table 3.3).

**Table 3.3.** Odds ratios, coefficients, and p-values of the association between epidemiological factors and herd-level Johne 's Disease status, based on ELISA assays performed on individual milk samples in 600 herds in Minnesota.

Covariates	N	Univariable analysis		Multivariable analysis			
		Odds ratio (CI <sub>95%</sub> )	p- value*	$\beta$	S.E.	Odds ratio (CI <sub>95%</sub> )	p-value*
Herd size			<0.001				<0.001
1 to $\leq$ 195	476	Ref	-	-	-	-	-
196 to $\leq$ 553	100	1.67(1.39-2.02)	<0.001	1.58	0.43	4.84(2.20-12.27)	<0.001
554 to 1929	24	13.74(2.86-246.87)	0.011	1.54	1.08	4.67(0.83-88.29)	0.153
Testing frequency			<0.001				<0.001
1	208	Ref	-	-	-	-	-
>1 to $\leq$ 5	208	3.04(2.03-4.57)	<0.001	1.03	0.21	2.81(1.86-4.29)	<0.001
6 to $\leq$ 20	80	13.15(6.16-32.64)	<0.001	2.48	0.43	11.85(5.44-29.82)	<0.001
21 to 30	104	64.30(19.64-396.39)	<0.001	3.77	0.73	43.34(13.00-269.33)	<0.001
Soil type (texture)			<0.001				0.017
Sand	189	Ref	-	-	-	-	-
Silt	115	2.87(1.69-5.01)	<0.001	0.63	0.32	1.88(1.01-3.54)	0.047
Loam	290	1.79(1.21-2.63)	0.003	0.63	0.22	1.88(1.21-2.93)	0.005
Clay	6	3.59(0.56-69.51)	0.25	1.62	1.15	5.04(0.68-102.54)	0.162
Soil pH			0.004				
<6.0	44	Ref	-				
6.0 to 7.0	467	2.15(1.14-4.03)	0.02				
>7.0	89	2.21(1.05-4.70)	0.037				
Hydrologic soil type			0.87				
A	57	Ref	-				
B	205	1.23(0.64-2.29)	0.527				
C	119	0.95(0.47-1.85)	0.874				
D	211	0.95(0.49-1.76)	0.872				
Participation in VJDHSP <sup>#</sup>			0.84				
Non-participants	576	Ref	-				
Participants	24	1.09(0.46-2.87)	0.84				
Agroecological zone			0.57				
Crop/Grassland	534	Ref	-				
Non-vegetated land	52	0.85(0.76-0.96)	0.76				
Shrubland	7	0.94(0.82-1.07)	0.50				
Deciduous Forest	7	1.08(0.95-1.22)	0.50				
Spatial dependence							
1km	0			$\lambda$ =N/A			
5km	344			$\lambda$ =0.00			0.962
10km	519			$\lambda$ =0.03			0.171
15km	565			$\lambda$ =0.02			0.256
20km	580			$\lambda$ =0.01			0.157
109.7km	600			$\lambda$ =-0.01			0.352

\*p-values from the Likelihood Ratio Test. Spatial correlation parameter ( $\lambda$ ).

<sup>#</sup> VJDHSP= USDA Voluntary Johne's control program.

### 3.5 Discussion

To our knowledge, this is the first study suggesting the use of an existing voluntary testing program to track JD in Minnesota. Study recognizes that the MNDHIA voluntary testing program may be a useful source to investigate the JD status of Minnesota dairies. The voluntary testing program participants in 2014-2017 represents 13% of the dairy herds across Minnesota and represents areas with both high and low density of dairy farms. Even though county-level JD herd prevalence could not be reliably estimated, because the required sample sizes were not achieved in most counties, we were able to establish the herd-level JD status (positive/negative) in most (600/723; 83%) of the herds participating in the program.

As per the epidemiological factor analysis, the most important epidemiological factors contributing to the JD status of a herd were herd size, testing frequency, and soil type, i.e. texture. We did not observe spatial dependence of the residuals of the regression model indicating that the observation of similar characteristics in JD status in the participant dairy herds were explained by the three covariates, namely, herd size, testing frequency, and soil type, i.e. texture. These results will be used to inform the potential use of the database as a surveillance tool and to suggest improvements in JD testing program conducted by MNDHIA.

The dairy herds in the southern Minnesota were more likely to be JD test-positive compared to herds in the north-central region (Fig. 3.2: Panel a). Interestingly, the spatial pattern of the herd size and the JD status of the study population showed a similar pattern (Fig. 3.2: Panel b). A comparable observation where participant herd sizes were unevenly distributed in space was also observed among participants in a Danish JD control program (Bihrmann. et al., 2012). Examination of the reasons why dairy farmers choose to test, or not, for JD status exceeded the scope of this study. It is envisaged that participation in voluntary testing and control programs varies due to multiple factors such as: a) farmers' belief in the importance of the JD (Ritter et. al., 2015), b) farmer's belief in control and prevention strategies including the investment of time and resources (Ritter et. al., 2016), and c) availability of the testing facilities and trained personnel to conduct testing at convenience (USDA-APHIS, 2010; Collins et al., 2010).

The herds located on silt, loam, or clay soils were more likely to be JD positive compared to the herds located on sandy soils. This observation contradicts early studies in the Midwest (Ward and Perez, 2003), which found that high silt content was associated with reduced detection of JD. However, studies by Dhand et al., (2009) and Salgado et al., (2011) described the increased likelihood of detecting JD on loamy and clay soils, respectively. This survival of MAP on loamy

soil was experimentally observed by Salgado et al. (2011), which concluded that MAP tends to migrate slower through loamy soils compared to sandy soils and thus loamy soils may have MAP contaminated upper soil layers and pasture. Furthermore, according to Salgado et al., (2011), in addition to soil type, the amount of rainfall and the soil pH also play an important role in the fate of MAP in the environment. Soil hydrologic characteristics and pH were not included in the final model, but it is worthwhile to acknowledge that soil texture, pH, and hydrologic characteristics are interconnected (Ditzler et al., 2017) and further analysis is needed to understand the association of soil features with MAP.

Limitations associated with this study include the restricted generalizability of the results because MNDHIA data represent only 13% of dairy herds in Minnesota. Additionally, the lack of information on herd characteristics other than herd size and testing frequency limited the value of our analysis. When data arise from imperfect surveillance systems, the interpretation of results must be done with caution because the covariates can be related either to the occurrence of the disease or to the efficiency of the data collection system (Vergne et al., 2014). However, evaluating an existing database in relation to the features such as benefit, coverage, data collection, representativeness, and sustainability are key steps when considering passive surveillance (Velasova et al., 2015).

Usefulness of the existing production and health data for veterinary surveillance, research, and management decision making is well acknowledged (Velasova et al., 2015). Although MNDHIA JD testing is not currently designed for multiple uses and does not offer to support a control program, we recognized opportunities to improve the MNDHIA database to be used as a passive surveillance tool. For example, this study has determined the number of herds and the number of animals to be sampled from each herd to ensure a reliable evaluation of its JD status (positive/negative). Additionally, improving the database enables collecting accurate farm location data, herd-level information, and biosecurity measures to facilitate a better assessment of JD status/risk. This study elucidates the evaluation of JD at herd and regional level using available data. The individual-level data analysis of the same dataset was presented elsewhere (Machado et al., 2018).

Establishing a novel JD surveillance program would be costly. A voluntary testing database could be a preferred solution for monitoring endemic chronic diseases like JD (Velasova et al., 2015). Moreover, the strength of using MNDHIA data as a source is that there are sixteen National Dairy Herd Improvement Association laboratories across the U.S. (<http://www.dhia.org/members.asp>) and their network already acts as a record keeping system for

dairy farming systems. Having a system to evaluate the JD status in a region would benefit the dairy industry in multiple ways: estimating herd-level JD status, recognizing the differences between participants and non-participants, understanding underlying risk factors and covariates in the neighborhoods, and eventually recognizing disease-free areas ([Garcia and Shalloo 2015](#)).

### **3.6 Conclusions**

In summary, results reported here suggest that a routinely generated database obtained from a voluntary testing program can be used as a passive surveillance tool to monitor the infection status and epidemiological determinants of JD. However, because the risk of introduction may always be present, successful prevention and control of JD depends on ongoing willingness to continue funding, surveillance, and research on the disease by both animal health authorities and the livestock industry ([Federal Register 65, 2000](#); [Kennedy and Benedictus, 2001](#); [Garcia and Shalloo 2015](#)).

### **3.7 Acknowledgements**

Mr. G. Palace and the members of Minnesota Dairy Herd Improvement Association for their support in data collection and Dr. J. Bell of the University of Minnesota for his inputs. This study was funded in part by the Minnesota Discovery, Research, and Innovation Economy (MnDRIVE) program and Office of the Vice President for Research (OVPR) of the University of Minnesota.

### **3.8 References**

1. Banerjee S, Carlin BP, Gelfand AE. 2015. Hierarchical modeling and analysis for spatial data. 2nd edition. CRC press, Boca Raton. FL. pp. 73-96.
2. Benjamin LA, Fosgate GT, Ward MP, Roussel AJ, Feagin RA, Schwartz AL. 2010. Attitudes towards biosecurity practices relevant to Johne's disease control on beef cattle farms. *Prev. Vet. Med.* 94(3-4): 222-230.

3. Besag J. 1974. Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion), *J. Royal Stat. Soc. Series B.* 36(2): 192–236.
4. Bihrmann K, Nielsen SS, Toft N, Ersboll AK. 2012. Spatial differences in occurrence of paratuberculosis in Danish dairy herds and in control programme participation. *Prev. Vet. Med.* 103(2-3): 112-119.
5. Bivand R, Piras G. 2015. Comparing Implementations of Estimation Methods for Spatial Econometrics. *J. Stat. Softw.* 63: 1-36.
6. Bivand RS, Hauke J, Kossowski T. 2013. Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geogr. Anal.* 45:150-179.
7. Carter MA. 2012. Prevalence and Prevention of Paratuberculosis in North America. *Jpn. J. Vet. Res.* 60: S9-S18.
8. Collins MT, Eggleston V, Manning EJB. 2010. Successful control of Johne's disease in nine dairy herds: Results of a six-year field trial. *J. Dairy Sci.* 93(4):1638-1643.
9. Collins MT, Sockett DC, Goodger WJ, Conrad TA, Thomas CB, Carr DJ. 1994. Herd prevalence and geographic distribution of, and risk factors for, bovine paratuberculosis in Wisconsin. *J. Am. Vet. Med. Assoc.* 204(4): 636-641.
10. Collins MT. 2003. Update on paratuberculosis: 1. Epidemiology of Johne's disease and the biology of *Mycobacterium paratuberculosis*. *Ir. Vet. J.* 56(11): 565-574.
11. Davis ROE, Bennett HH. 1927. Grouping of soils on the basis of mechanical analysis, Department Circular 419, U.S. Dept. of Agric., Washington, D. C.
12. Dhand NK, Eppleston J, Whittington RJ, Toribio J. 2009. Association of farm soil characteristics with ovine Johne's disease in Australia. *Prev. Vet. Med.* 89(1-2), 110-120.
13. Ditzler C, Scheffe K, Monger HC.(eds.). 2017. Soil Science Division Staff. Soil survey manual. USDA Handbook 18. Government Printing Office, Washington, D.C.

14. Dore E, Pare J, Cote G, Buczinski S, Labrecque O, Roy JP, Fecteau G. 2012. Risk Factors Associated with Transmission of *Mycobacterium avium* subsp *paratuberculosis* to Calves within Dairy Herd: A Systematic Review. *J. Vet. Intern. Med.* 26: 32-45.
15. Elliott GN, Hough RL, Avery LM, Maltin CA, Campbell CD. 2015. Environmental risk factors in the incidence of Johne's disease. *Crit. Rev. Microbiol.* 41(4): 488-507.
16. ESRI: Environmental Systems Research Institute. ArcMap Version 10.3.4. 2017. Environmental Research Institute, Inc. Redlands, CA, USA.
17. Federal register 65. 2000. Available online at:  
[https://www.aphis.usda.gov/animal\\_health/animal\\_diseases/johnes/downloads/docket98-037-2.pdf](https://www.aphis.usda.gov/animal_health/animal_diseases/johnes/downloads/docket98-037-2.pdf).
18. Garcia AB, Shalloo L. 2015. Invited review: The economic impact and control of paratuberculosis in cattle. *J. Dairy Sci.* 98(8): 5019-5039.
19. Getis A, Ord JK. 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr. Anal.* 24: 189-206.
20. Gonda MG, Chang YM, Shook GE, Collins MT, Kirkpatrick BW. 2007. Effect of *Mycobacterium paratuberculosis* infection on production, reproduction, and health traits in US Holsteins. *Prev. Vet. Med.* 80: 103-119.
21. Grewal SK, Rajeev S, Sreevatsan S, Michel FC. 2006. Persistence of *Mycobacterium avium* subsp *paratuberculosis* and other zoonotic pathogens during simulated composting, manure packing, and liquid storage of dairy manure. *Appl. Environ. Microbiol.* 72(1): 565-574.
22. Humphry RW, Cameron A, Gunn GJ. 2004. A practical approach to calculate sample size for herd prevalence surveys. *Prev. Vet. Med.* 65: 173-188.
23. Jenks GF. 1967. The data model concept in statistical mapping, *International Yearbook of Cartography.* 7: 186-190.
24. Johnson-Ifearulundu YJ, Kaneene JB. 1997. Relationship between soil type and *Mycobacterium paratuberculosis*. *J. Am. Vet. Med. Assoc.* 210: 1735-1740.

25. Kennedy DJ, Benedictus G. 2001. Control of *Mycobacterium avium* subsp paratuberculosis infection in agricultural species. *Rev. Sci. Tech. Office International Des Epizooties*. 20 151-179.
26. Kopecky KE. 1977. Distribution of paratuberculosis in Wisconsin, by soil regions. *J. Am. Vet. Med. Assoc.* 170(3): 320-324.
27. Lombard JE, Gardner IA, Jafarzadeh SR, Fossler CP, Harris B, Capsel RT, Wagner BA, Johnson WO. 2013. Herd-level prevalence of *Mycobacterium avium* subsp paratuberculosis infection in United States dairy herds in 2007. *Prev. Vet. Med.* 108: 234-238.
28. Lombard JE. 2011. Epidemiology and economics of paratuberculosis. *Vet. Clin. North Am. Food Anim. Pract.* 27: 525–535
29. Machado G, Kanankege K, Schumann V, Wells S, Perez A, Alvarez J. 2018. Identifying individual animal factors associated with *Mycobacterium avium* subsp paratuberculosis (MAP) milk ELISA positivity in dairy cattle in the Midwest region of the United States. *BMC Vet. Res.* 14: 28.
30. McKenna SLB, Keefe GP, Tiwari A, VanLeeuwen J, Barkema HW. 2006. Johne's disease in Canada Part II: Disease impacts, risk factors, and control programs for dairy producers. *Can. Vet. J.* 47: 1089-1099.
31. MNGSC GAP: Minnesota GeoSpatial commons. 2015. Minnesota Department of Natural Resources (DNR) - Division of Forestry. GAP Land Cover. Accessed 15 Aug 2017. Available from: <https://gisdata.mn.gov/dataset/biota-landcover-gap>
32. Moran PAP. 1948. The interpretation of statistical maps. *J. Royal Stat. Soc. Series B.* 37: 243–251.
33. NASS: National Agricultural Statistics Service. 2018. Milk Production. ISSN:1949-1557. Available from: <http://usda.mannlib.cornell.edu/usda/nass/MilkProd//2010s/2017/MilkProd-02-21-2017.pdf>.
34. NASS: National Agricultural Statistics Service. Census of Agriculture. 2012. United States Department of Agriculture, Washington DC: Government Printing Office. Available from:



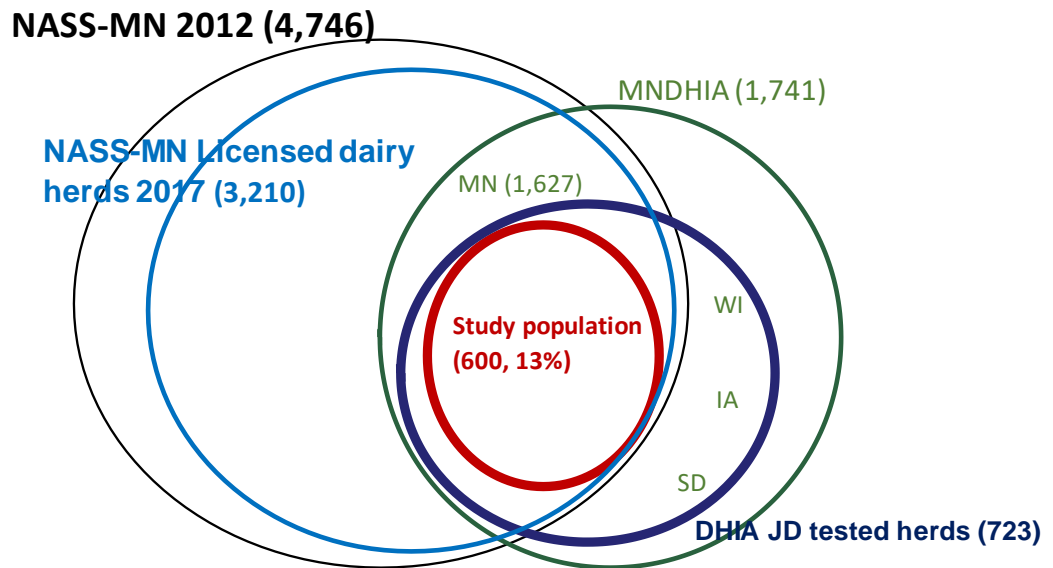
- [https://www.agcensus.usda.gov/Publications/2012/Full\\_Report/Volume\\_1,\\_Chapter\\_2\\_US\\_State\\_Level/st99\\_2\\_011\\_011.pdf](https://www.agcensus.usda.gov/Publications/2012/Full_Report/Volume_1,_Chapter_2_US_State_Level/st99_2_011_011.pdf). 2012; Table 11. p. 343
35. Nielsen SS, Toft N. 2008. Ante mortem diagnosis of paratuberculosis: A review of accuracies of ELISA, interferon-gamma assay and faecal culture techniques. *Vet. Microbiol.* 129: 217-235.
  36. Ord JK, Getis A. 1995. Local spatial autocorrelation statistics – Distributional issues and an application. *Geogr. Anal.* 1995; 27: 286-306.
  37. Ott SL, Wells SJ, Wagner BA. 1999. Herd-level economic losses associated with Johne's disease on US dairy operations. *Prev. Vet. Med.* 40: 179-192.
  38. Pickup RW, Rhodes G, Bull TJ, Arnott S, Sidi-Boumedine K, Hurley M, Hermon-Taylor J. 2006. *Mycobacterium avium* subsp paratuberculosis in lake catchments, in river water abstracted for domestic use, and in effluent from domestic sewage treatment works: diverse opportunities for environmental cycling and human exposure. *Appl. Environ. Microbiol.* 72: 4067–77.
  39. Pillars RB, Grooms DL, Woltanski JA, Blair E. 2009. Prevalence of Michigan dairy herds infected with *Mycobacterium avium* subspecies paratuberculosis as determined by environmental sampling. *Prev. Vet. Med.* 89: 191-196.
  40. R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
  41. Raizman EA, Fetrow J, Wells SJ, Godden SM, Oakes MJ, Vazquez G. 2007. The association between *Mycobacterium avium* subsp. paratuberculosis fecal shedding or clinical Johne's disease and lactation performance on two Minnesota, USA dairy farms. *Prev. Vet. Med.* 78:179–195.
  42. Raizman EA, Wells SJ, Munoz-Zanzi CA, Tavoranpanich S. 2011. Estimated within-herd prevalence (WHP) of *Mycobacterium avium* subsp paratuberculosis in a sample of Minnesota dairy herds using bacterial culture of pooled fecal samples. *Canadian Journal of Veterinary Research-Revue Canadienne De Recherche Veterinaire.* 75(2):112-116.

43. Rangel SJ, Pare J, Dore E, Arango JC, Cote G, Buczinski S, Labrecque O, Fairbrother JH, Roy JP, Wellemans V, Fecteau G. 2015. A systematic review of risk factors associated with the introduction of *Mycobacterium avium* spp. paratuberculosis (MAP) into dairy herds. *Can. Vet. J.* 56: 169-177.
44. Rideout BA, Brown ST, Davis WC, Gay JM, Giannella RA, Hines II ME, Hueston WD, Hutchinson LJ. et al. 2003. Committee on Diagnosis and Control of Johne's Disease Board on Agriculture and Natural Resources. Division on Earth and Life Studies. National Academy of Sciences. USA. ISBN: 0-309-56641-X, p.244, 6x9
45. Ritter C, Kwong GPS, Wolf R, Pickel C, Slomp M, Flaig J, Mason S, Adams CL, Kelton DF, Jansen J, De Buck J, Barkema HW. 2015. Factors associated with participation of Alberta dairy farmers in a voluntary, management-based Johne's disease control program. *J. Dairy Sci.* 98(11):7831-7845.
46. Salgado M, Collins MT, Salazar F, Kruze J, Bolske G, Soderlund R, Juste R, Sevilla IA, Biet F, Troncoso F, Alfaro M. 2011. Fate of *Mycobacterium avium* subsp paratuberculosis after Application of Contaminated Dairy Cattle Manure to Agricultural Soils. *Appl. Environ. Microbiol.* 77(6): 2122-2129.
47. Scott HM, Sorensen O, Wu JT, Chow EY, Manninen K. 2007. Seroprevalence of and agroecological risk factors for *Mycobacterium avium* subspecies paratuberculosis and *Neospora caninum* infection among adult beef cattle in cow-calf herds in Alberta, Canada. *Can. Vet. J.* 48(4): 397-406.
48. Soil Survey Staff. 2017. Gridded Soil Survey Geographic (gSSURGO) Database for State name. United States Department of Agriculture, Natural Resources Conservation Service. Available online at <http://datagateway.nrcs.usda.gov/>. Accessed 10 Feb 2018.
49. Soil Survey Staff. Hydrologic Soil Group. 2018. Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at the following link: <https://websoilsurvey.sc.egov.usda.gov/>. Accessed 10 Jan 2018.
50. Soil survey staff. USDA soil texture triangle classification. Based on: Davis ROE, Bennett HH. 1927. Grouping of soils on the basis of mechanical analysis, Dep. Circ. 419, U.S. Dep.

- of Agric., Washington, D. C. Available from:  
[https://www.nrcs.usda.gov/Internet/FSE\\_DOCUMENTS/stelprdb1044818.pdf](https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1044818.pdf). 2017.
51. Thrusfield M. 2007. *Veterinary Epidemiology*, 3rd Edition, Blackwell Science, Oxford, UK. pp. 233-234). ISBN: 978-1-405-15627-1
  52. Tiwari A, VanLeeuwen JA, McKenna SLB, Keefe GP, Barkema HW. 2006. Johne's disease in Canada - Part I: Clinical symptoms, pathophysiology, diagnosis, and prevalence in dairy herds. *Can. Vet. J.* 47(9): 874-882.
  53. USAHA: United States Animal Health Association. 1993. *Proceedings of the 97th Annual Meeting of the USAHA*.
  54. USDA: United States Department of Agriculture. 1997. Johne's disease on U.S. dairy operations. In *Johne's disease on U.S. Dairy Operations*. Washington DC: Animal and Plant Health Inspection Service.
  55. USDA-APHIS: United States Department of Agriculture - Animal and Plant Health Inspection Service. Disease information. 2015. Preventing Introduction of Johne's disease. Last modified: August 2015. Last cited: March 2018. Available from:  
[https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/animal-disease-information/cattle-disease-information/sa\\_johnes/ct\\_prevent-intro](https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/animal-disease-information/cattle-disease-information/sa_johnes/ct_prevent-intro).
  56. USDA-APHIS-VS: United States Department of Agriculture - Animal and Plant Health Inspection Service - Veterinary Services. 2010. Uniform Program Standards for the Voluntary Bovine Johne's Disease Control Program, APHIS. 91-45-016,  
<http://www.johnesdisease.org>. Accessed 10 Dec 2016. Available online from:  
[https://johnes.org/handouts/files/USDA\\_Program\\_Standards\\_Sept-2010.pdf](https://johnes.org/handouts/files/USDA_Program_Standards_Sept-2010.pdf)
  57. VanLeeuwen JA, Keefe GP, Tremblay R, Power C, Wichtel JJ. 2001. Seroprevalence of infection with *Mycobacterium avium* subspecies paratuberculosis, bovine leukemia virus, and bovine viral diarrhoea virus in Maritime Canada dairy cattle. *Can. Vet. J.* 42(3): 193-198.
  58. Velasova M, Drewe JA, Gibbons J, Green M, Guitian J. 2015. Evaluation of the usefulness at national level of the dairy cattle health and production recording systems in Great Britain. *Veterinary Record.* 177(12): 304-304

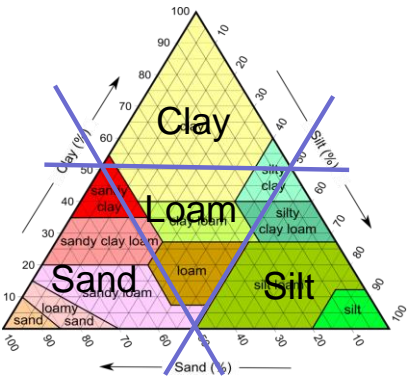
59. Venables WN, Ripley BD. 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN: 0-387-95457-0
60. Vergne T, Paul MC, Chaengprachak W, Durand B, Gilbert M, Dufour B, Roger F, Kasemsuwan S, Grosbois V. 2014. Zero-inflated models for identifying disease risk factors when case detection is imperfect: Application to highly pathogenic avian influenza H5N1 in Thailand. *Prev. Vet. Med.* 114: 28-36.
61. Waddell L, Rajic A, Stark K, McEwen SA. 2016. *Mycobacterium avium* sub spp. paratuberculosis detection in animals, food, water and other sources or vehicles of human exposure: A scoping review of the existing evidence. *Prev. Vet. Med.* 132: 32-48.
62. Ward MP, Perez AM. 2004. Association between soil type and paratuberculosis in cattle herds. *Am. J. Vet. Res.* 65: 10-14.
63. Wells SJ, Wagner BA. 2000. Herd-level risk factors for infection with *Mycobacterium paratuberculosis* in US dairies and association between familiarity of the herd manager with the disease or prior diagnosis of the disease in that herd and use of preventive measures. *J. Am. Vet. Med. Assoc.* 216: 1450-1457.
64. Whittington RJ, Marsh IB, Reddacliff LA. 2005. Survival of *Mycobacterium avium* subsp paratuberculosis in dam water and sediment. *Appl. Environ. Microbiol.* 71: 5304-5308.

### 3.9 Supplement Tables and Figures



**Supplement Fig. S3.1.** The study population (n=600; 13% of the 4,746 dairy herds in Minnesota ([NASS, 2012](#)) compared to the participants in the Minnesota Dairy Herd Improvement Association (MNDHIA; n=1,741). The 600 herds here 18.7 % of the licensed dairy herds in Minnesota with permits to ship milk ([NASS, 2018](#)). Between November 2014 and April 2017, there were 723 herds tested for JD at MNDHIA. Among those, 123 were excluded due to locations outside Minnesota (MN) including Iowa (IA), Wisconsin (WI), and South Dakota (SD); herds without location information; and, 3) inadequate sampling.

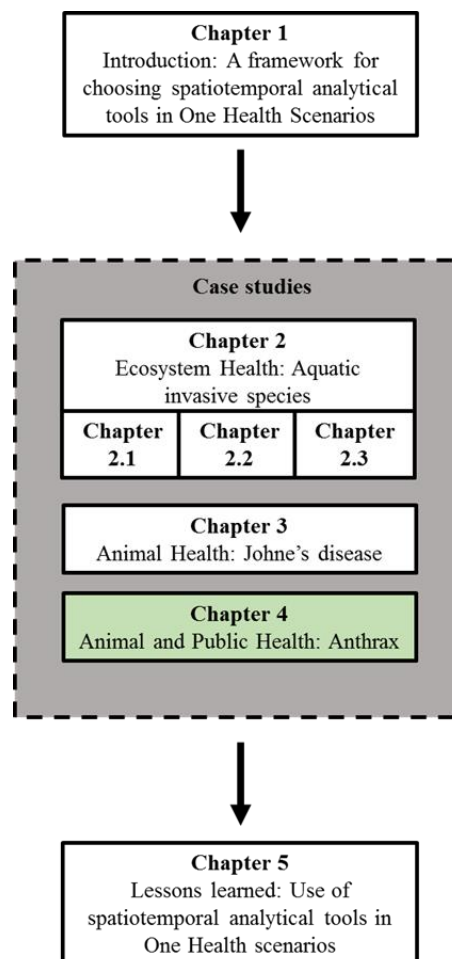
**Supplement Table S3.1.** Environmental factors assessed for association with Johne’s disease status in Minnesota dairy herds participating in the voluntary testing program conducted by Minnesota Dairy Herd Improvement Association.

Variable	Description and relevant references	Data source
1. Soil type (texture)	<p>Soil types re-categorized based on the percentage of the clay, sand, and silt content, as seen in the figure below i.e. USDA soil classification (Davis and Bennett, 1927; Soil Science Division Staff, 2017)</p> 	<p>Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States - 10 meter resolution.  <a href="http://datagateway.nrcs.usda.gov">http://datagateway.nrcs.usda.gov</a></p>
2. Soil pH	<p>Values ranged between 5.6 and 7.5. Missing data were represented with a zero</p>	<p>Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States – 10 meter resolution.  <a href="http://datagateway.nrcs.usda.gov">http://datagateway.nrcs.usda.gov</a></p>
3. Hydrologic soil group (run-off potential)	<p>Seven categories: A, B, C, D, A/D, B/D, and C/D. A/D, B/D, C/D, and D categories were combined to one category and labelled “D” for the analysis here.</p>	<p>Web soil Survey.  <a href="https://websoilsurvey.sc.egov.usda.gov/">https://websoilsurvey.sc.egov.usda.gov/</a></p>

4. Agroecological zone	The National Gap Analysis Project (GAP) vegetation information were categorized into four levels. The Level 2 categorization was used in the analysis.	Upper Midwest Gap Analysis Program (UMGAP)( <a href="http://www.umesc.usgs.gov/reports_publications/psrs/psr_1999_04.html">http://www.umesc.usgs.gov/reports_publications/psrs/psr_1999_04.html</a> ) of the U.S. Geological Survey. 30-meter resolution. ( <a href="http://gapanalysis.usgs.gov">gapanalysis.usgs.gov</a> )
------------------------	--	--

**Supplement Table S3.2.** Akaike information criterion (AIC) values for the model fits.

Model	AIC
Final regression model	591.47
Final regression model with iCAR at following distances:	
1Km	640.64
5Km	640.65
10Km	638.78
15Km	639.36
20Km	638.65
109.7Km	639.79



#### Chapter 4: Animal and Public health

*The use of spatiotemporal cluster analysis as a tool to recognize historic patterns of Anthrax in endemic areas: A comparison between Minnesota and Kazakhstan*

**In review: PLoS ONE (December 2018)**

Kanankege KST, Abdrakhmanov SK, Alvarez J, Glaser L, Bender JB, Mukhanbetkaliyev YY, Korennoy FI, Kadyrov AS, Abdrakhmanova AS, Perez AM. Comparison of spatiotemporal patterns of historic animal Anthrax outbreaks in Minnesota and Kazakhstan.



## 4.1 Overview

Disease spread in populations is a consequence of the interaction between host, pathogen, and environment, i.e. the epidemiological triad. Yet the influences of each triad component may vary dramatically for different settings. Comparison of environmental, demographic, socio-economic, and historical backgrounds may support tailoring site-specific control measures. Because of the long-term survival of *Bacillus anthracis*, Anthrax is a suitable example for studying the influence of triad components in different endemic settings. We compared the spatiotemporal patterns of historic animal Anthrax records in two endemic areas, located at northern latitudes in the western and eastern hemispheres. Our goal was to compare the spatiotemporal patterns in Anthrax progression, intensity, direction, and recurrence (disease hot spots), in relation to epidemiological factors and potential trigger events. Reported animal cases in Minnesota, USA (n=289 cases between 1912 and 2014) and Kazakhstan (n=3,997 cases between 1933 and 2014) were analyzed using the spatiotemporal directionality test and the spatial scan statistic. Over the last century Anthrax occurrence in Minnesota was sporadic whereas Kazakhstan experienced a long-term epidemic. Nevertheless, the seasonality was comparable between sites, with a peak in August. Declining number of cases at both sites was attributed to vaccination and control measures. The spatiotemporal directionality test detected a relative northeastern directionality in disease spread for long-term trends in Minnesota, whereas a southwestern directionality was observed in Kazakhstan. In terms of recurrence, the maximum timespans between cases at the same location were 55 and 60 years for Minnesota and Kazakhstan, respectively. Disease hotspots were recognized in both settings, with spatially overlapping clusters years apart. Distribution of the spatiotemporal cluster radii between study sites supported suggestion of site-specific control zones. Spatiotemporal patterns of Anthrax occurrence in both endemic regions were attributed to multiple potential trigger events including major river floods, changes in land use, agriculture, and susceptible livestock populations. Results here help to understand the long-term epidemiological dynamics of Anthrax while providing suggestions to the design and implementation of prevention and control programs, in endemic settings.

## 4.2 Introduction

Anthrax is a zoonotic disease affecting livestock, wildlife, and humans which caused by the spore-forming bacterium *Bacillus anthracis* (Sternbach, 2003). Anthrax spores survive for extended periods of time in the environment and rapidly return to vegetative stages once

conditions are favorable (Driks, 2009). Therefore, recurrence of Anthrax is observed in endemic areas years apart (Driks, 2009; Elvander et al., 2017), which makes the recognition of risk areas and control zones a challenging task. Epidemiological determinants of Anthrax include soil pH, calcium, and organic matter content (Hugh-Jones et al., 2009), precipitation, temperature, wind, and vegetation biomass (Turnbull et al., 1988; Blackburn et al., 2007; Abdrakhmanov et al., 2017a). Additionally, river drainages/flood plains, wild herbivore population, human population densities, and anthropogenic activities, such as limestone mining or road building, are recognized to play a role in the occurrence of Anthrax outbreaks (Sternbach, 2003; Ottorino et al., 2008).

Recurrence of Anthrax in endemic areas is often analyzed and modeled in relation to the environmental characteristics of the setting, to identify high risk areas (Blackburn et al., 2007; Abdrakhmanov et al., 2017a), despite differences in the demographic, socio-economic, and historical backgrounds. However, a recent study which attempted to predict habitat suitability for *Bacillus anthracis* in Kazakhstan using an ecological niche model (ENM) fitted to USA data (and vice versa), failed to accurately predict the pathogen presence in novel landscapes (Mullins et al., 2013). Their observed differences in ENM predictions were attributed to the genetic-ecological divergence of the pathogen at different locations, suggesting that site-specific features should be considered when assessing the disease epidemiology and, ultimately, planning for prevention and control activities. Additionally, one may argue that it is equally important to understand the influence of spatiotemporal epidemiological factors and trigger events that led to the Anthrax occurrences, in addition to the recognition of ecological suitability. The inability to accurately predict the pathogen presence in novel landscape as seen in Mullins et al. (2013) study potentially is a result of the differences in the demographic, socio-economic, and agricultural-livestock production backgrounds play a major role in determining the spread of pathogens in different settings.

As suggested by the epidemiological triad concept, the interaction between host, pathogen, and environment results in disease patterns of populations. Yet the influences of each triad component may dramatically vary for different spatiotemporal conditions. The spatiotemporal characterization of historic Anthrax occurrence would facilitate identifying the patterns of climatic, anthropogenic, socioeconomic, agricultural, and environmental changes that may have led to the relevant disease outbreaks. Understanding those epidemiological factors and trigger events which may have influenced the spread of Anthrax, especially at the hotspots in which Anthrax recurs, may support suggesting tailored and site-specific risk zones to improve planning of control measures.

The site-specific spatiotemporal patterns of Anthrax spread could be better demonstrated in a generalizable manner if more than one endemic site was assessed. Here, two distinctly different Anthrax endemic sites, located at similar latitudes in the eastern and western hemispheres were selected for this study, namely, the U.S. state of Minnesota (latitudes 44° through 48°, North) and the Republic of Kazakhstan (latitudes 42° through 56°, North). Moreover, the socio-economic and demographic conditions of the two settings have been substantially different, historically. Both Minnesota and Kazakhstan have recorded Anthrax cases since early 1900's ([Abdrakhmanov et al., 2017a](#); [Bender et al., 2006](#)). Although ENM approaches were used to compare between the suitability of the two sites ([Blackburn et al., 2007](#); [Abdrakhmanov et al., 2017a](#); [Mullins et al., 2013](#)), the historic spatiotemporal patterns of both sites have not been analyzed retrospectively to identify potential disease hotspots and the trigger events.

The objectives of this study were to understand and compare the progression, intensity, direction, and recurrence of animal Anthrax cases in relation to epidemiological factors and trigger events in Minnesota and Kazakhstan over an extended period of time. The analysis was intended to generate hypotheses regarding the association of the outbreaks with site specific trigger events, to ultimately suggest control zone boundaries based on the reported data.

## 4.3 Data and Methods

### 4.3.1 Data

In Minnesota, 289 animal Anthrax cases reported between 1912 and 2014 to the Minnesota Board of Animal Health (MNBAH) were used in the analysis. For events occurred between 1920 and 1999 the geographical coordinates were obtained using historic aerial images whereas, for those cases that occurred after 2000, coordinates were recorded during site visits ([Bender et al., 2006](#)). In addition to location details, the database contained the number and species affected and date in which the disease was first reported. Both laboratory confirmed (n=233, 80.6%) and suspected cases were included in the analysis.

For the Republic of Kazakhstan, laboratory confirmed cases reported by the Cadastral register of stationary unfavorable foci on Anthrax between 1933 and 2014 (n=3,997) were analyzed. Kazakhstan's database contained coordinates of the location, year, number, animal species affected, and for 2,027 (53.6%) cases, the date of report.

### 4.3.2 Methods

A descriptive analysis was performed to characterize the animal species affected and the locations with recurrent Anthrax cases. Epidemic curves were used to visualize the annual progression of Anthrax cases. Seasonality of animal Anthrax occurrence was explored by plotting the monthly sum of cases in each location. Because precise date of report was only available for 53.6% of the records from Kazakhstan, the remaining 1,970 records in which only the year was provided were eliminated of the temporal analysis. Additionally, the G-rates of animal Anthrax, i.e. the number of cases over a land area unit in one year (Chen and Wang, 2017), were calculated to compare the heterogeneity of disease occurrence of both sites.

Spatiotemporal directionality of Anthrax spread was analyzed using the spatiotemporal directionality test implemented in the ClusterSeer v.2.05. software (User manual, 2012: <https://www.biomedware.com/>). The spatiotemporal directionality test was used to detect the space-time interaction through a calculation of the mean direction of Anthrax spread. In this analysis, the directionality is measured by means of the angle, i.e. the rotation from the horizontal axis (with east corresponding to 0°), and the angular concentration (AC). The AC quantifies the magnitude of the variance in the angles between vectors connecting pairs of cases (value between 0 and 1), where the closer the values get to one, the higher the consistency in the direction of disease spread. A relative matrix was computed for the analysis, which uses vectors connecting each case to all the following cases to measure the directionality. The significance of the mean direction was evaluated through a randomization procedure, which retains constant geographic locations and randomly assigns the time between pairs of cases. The p-value was determined by comparing the AC from the original (non-randomized) data to the null distribution generated through the randomization process in 999 iterations (User manual, 2012: <https://www.biomedware.com/>).

The permutation model of the scan statistic was used to quantify the spatiotemporal pattern of disease occurrence in both settings, with the intention to identify areas with recurrent outbreaks (disease hot spots) (Kulldorff et al., 2005). The spatiotemporal permutation model creates cylinders of candidate clusters of disease throughout the study area and time period. The base and height of the cylinder, representing space and time respectively, varies up to a maximum value of 50% of the study population (Kulldorff et al., 2005), which determines the possible maximum size of the cluster. The null hypothesis of the permutation model of the scan statistic is based on

an even distribution of cases in time and space regardless of the distribution of controls or the population at risk. The ratio between observed and expected number of cases within each candidate cylinder is computed and the significance of the cluster is tested using a Monte Carlo simulation process in which time labels are randomly assigned to each location 999 times (Kulldorff et al., 2005). Candidate clusters with a  $P \leq 0.05$  were assumed to represent space-time clusters at high risk for occurrence of the disease. The test was implemented using the SaTScan software (<http://www.satscan.org/> version 9.6). The spatiotemporal permutation model of the spatial scan statistic was set to a maximum of 10% and 5% for the spatial and time windows, respectively. The sensitivity of space and time parameters was assessed by repeating the analysis with maximum window sizes of 50%, 30%, 20%, 10%, and 5%. The resulting clusters characteristics were compared in relevance to the location, radii, and the stakeholder's/experts' knowledge of outbreaks. The stakeholders/experts were MNBAH and the research experts of Seifullin Kazakh Agrotechnical University, Kazakhstan.

#### **4.3.3 Investigation of potential trigger events related to historic outbreaks**

To describe and hypothesize on the underlying trigger events related to the results of the detected spatiotemporal clusters, historic records related to livestock distribution, agriculture and crop production, land use, and environmental events such as major floods and droughts, were investigated.

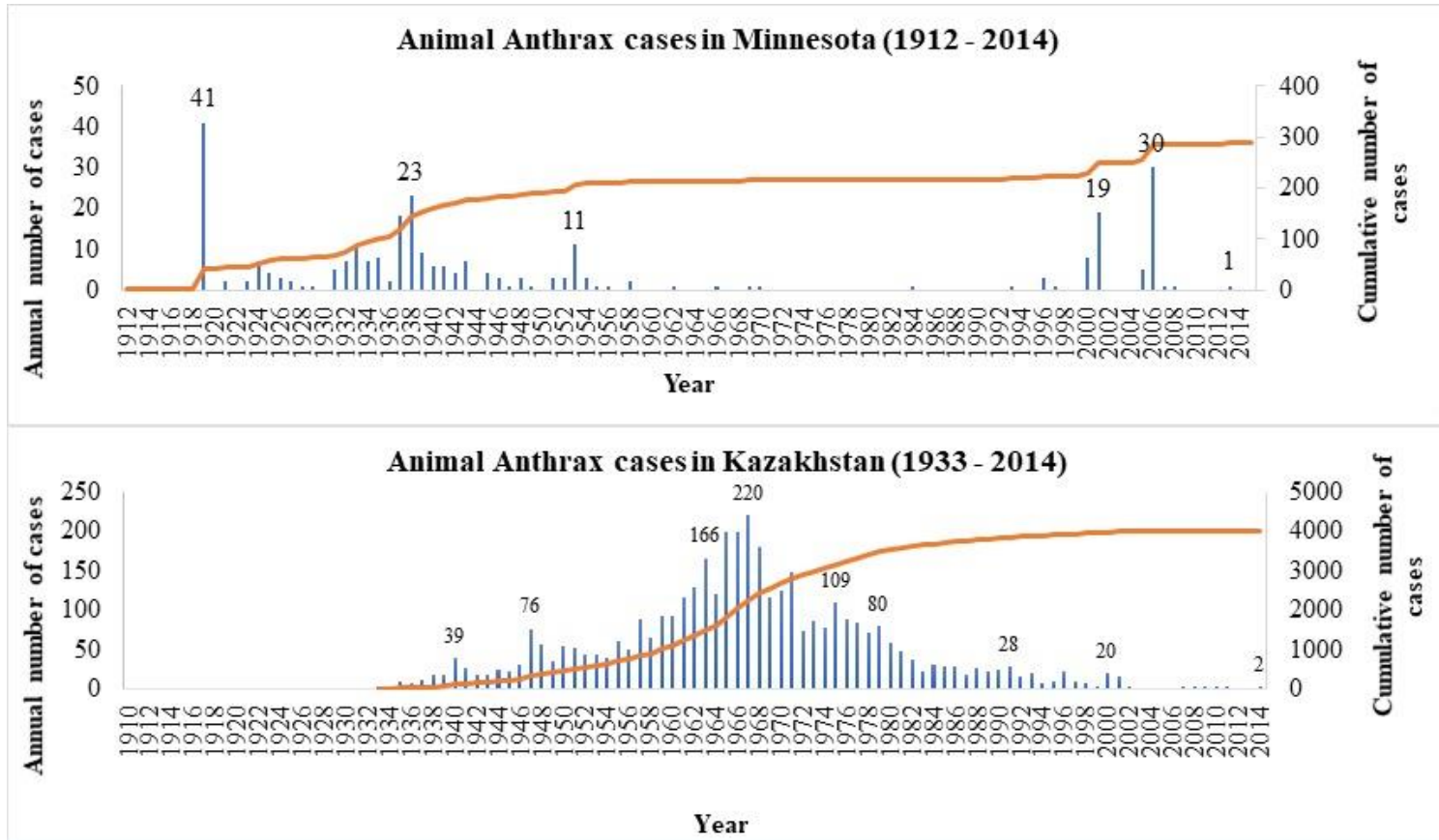
#### 4.4 Results

Minnesota experienced a period of multiple outbreaks prior to 1958, with three peaks in 1919, 1938, and 1953, respectively (Fig. 4.1). After 1958, the number of animal Anthrax cases decreased until 2000-2005, a period in which another major outbreak was observed (Fig. 4.1). The epidemic curve of Kazakhstan was consistent with a prominent and lengthy epidemic of animal anthrax that began in the 1940's and decreased during the 2000 and 2005 period, with a peak of 220 cases reported in 1968 (Fig. 4.1). Most of the cases in both Minnesota and Kazakhstan were reported in summer months, with a peak in August (Minnesota with  $n=90/289$ ; 31.1% and Kazakhstan:  $n=499/3,229$ ; 24.6% respectively) (Fig. 4.2).

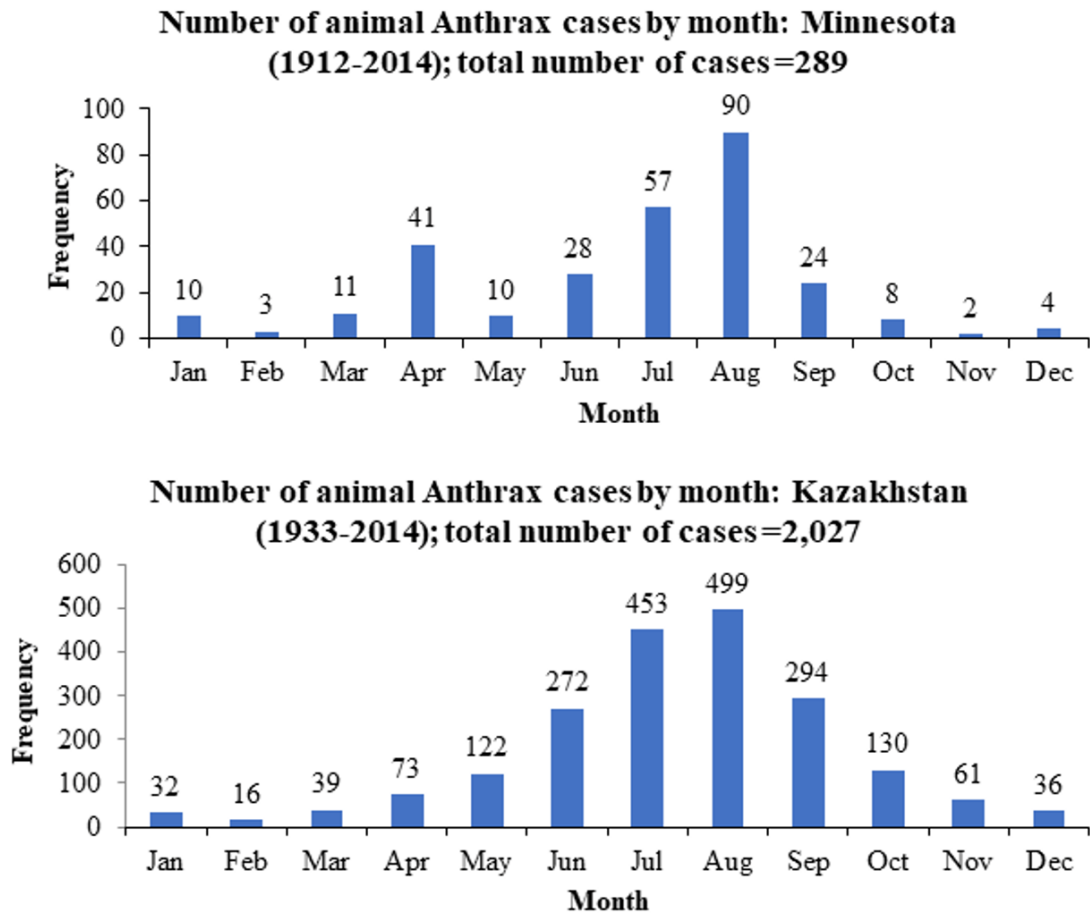
Anthrax cases were not homogeneously spread across the geographical space in Minnesota or Kazakhstan before and after the most recent/peak outbreak (Fig. 4.3). In Minnesota, a total of 222 unique locations with Anthrax cases were reported between 1912 and 2014. Most of the locations affected before the year 2000 were located in the southwestern part of Minnesota, whereas most of the cases reported after 2000 were located in the north-western corner of the State (Fig. 4.3). A total of 209 cases occurred in cattle, 119 in horses, 100 in swine, 55 in canids, and 9 in small ruminants. Some ( $n=28$ ) locations/farms reported recurrence of the disease  $\geq 2$  years apart, with a mean, median, and maximum time-spans of 10.9, 7, and 55 years (1941 and 1996) between recurrent cases in the same location.

In Kazakhstan, a total of 1,797 unique locations with animal Anthrax cases were reported between 1933 and 2014. The number of cases by species in Kazakhstan were 2,397 cattle, 1,210 small ruminants, 225 equines, 152 swine, 6 camels, 5 canids (dogs and foxes), 1 mink, and 1 borrowing animal case. Among the 1,797 locations/farms, there were 970 reported recurrences  $\geq 2$  years apart, with a mean, median, and maximum time-span of 7.47, 5, and 60 years (1947 and 2007) between recurrent cases in the same location, respectively.

According to the G-rate indicator of heterogeneity, compared to Minnesota (255 total cases in a 225,180 km<sup>2</sup> area,  $G\text{-rate}=0.012\pm 0.02$ ), the overall number of cases reported from Kazakhstan (3,997 total cases in a  $2.725\times 10^6$  km<sup>2</sup> area,  $G\text{-rate } 0.0178\pm 0.019$ ) was relatively high (unit of G-rate: number of cases per 1000 km<sup>2</sup> within a year).

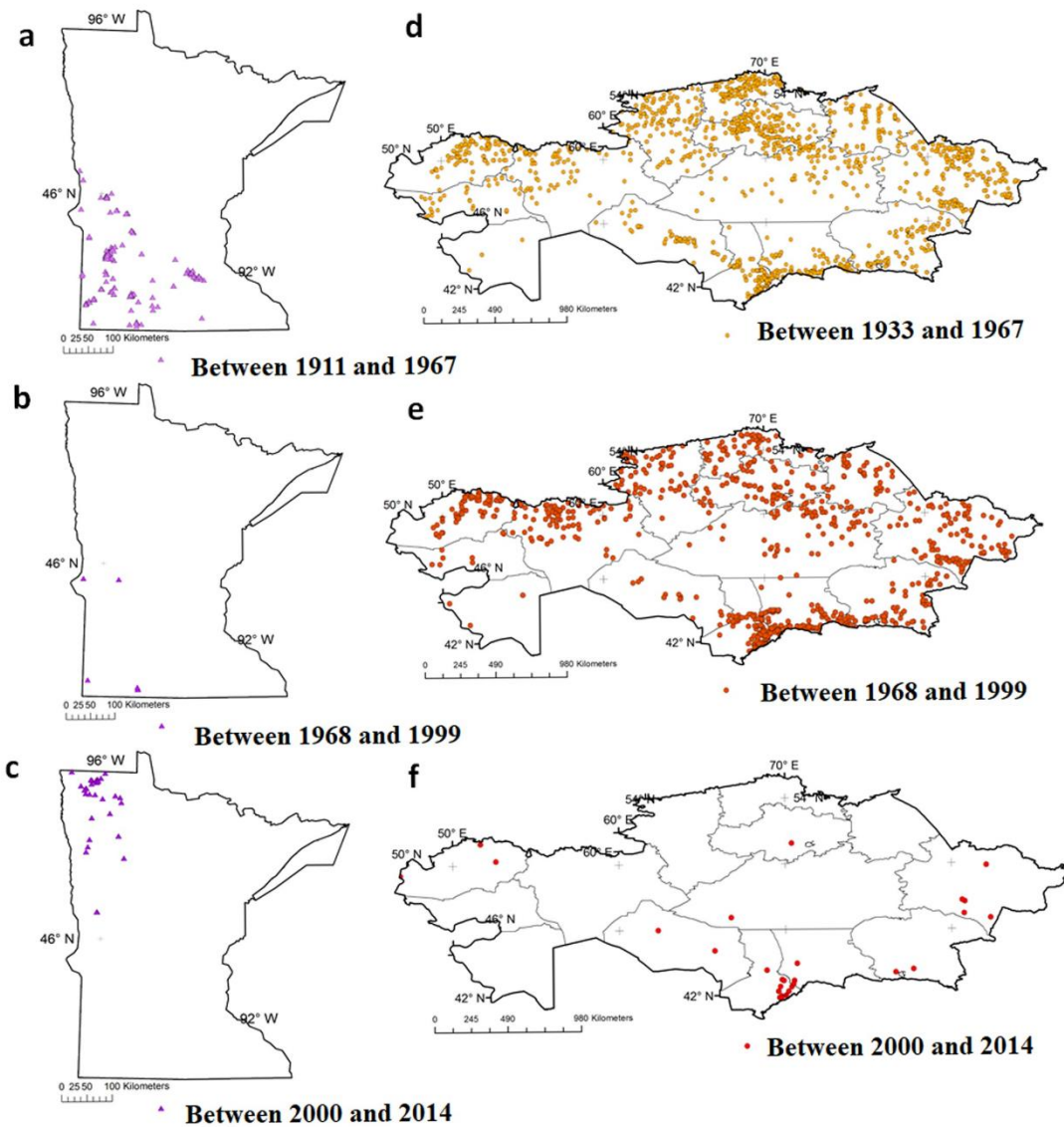


**Fig. 4.1. Epidemic curves illustrating the number of animal Anthrax outbreaks.** Outbreaks were reported between 1912 and 2014 in Minnesota and between 1933 and 2014 in Kazakhstan (Bars indicate annual number of reported cases, left axis, and lines indicate cumulative number of cases, right axis).



**Fig. 4.2. Number of animal Anthrax cases by month in Minnesota (1912-2014) and Kazakhstan (1933-2014).** The number of cases included in the monthly analysis (n) were 289 and 2,027 for Minnesota and Kazakhstan (53.6% of the total records of Kazakhstan) respectively.





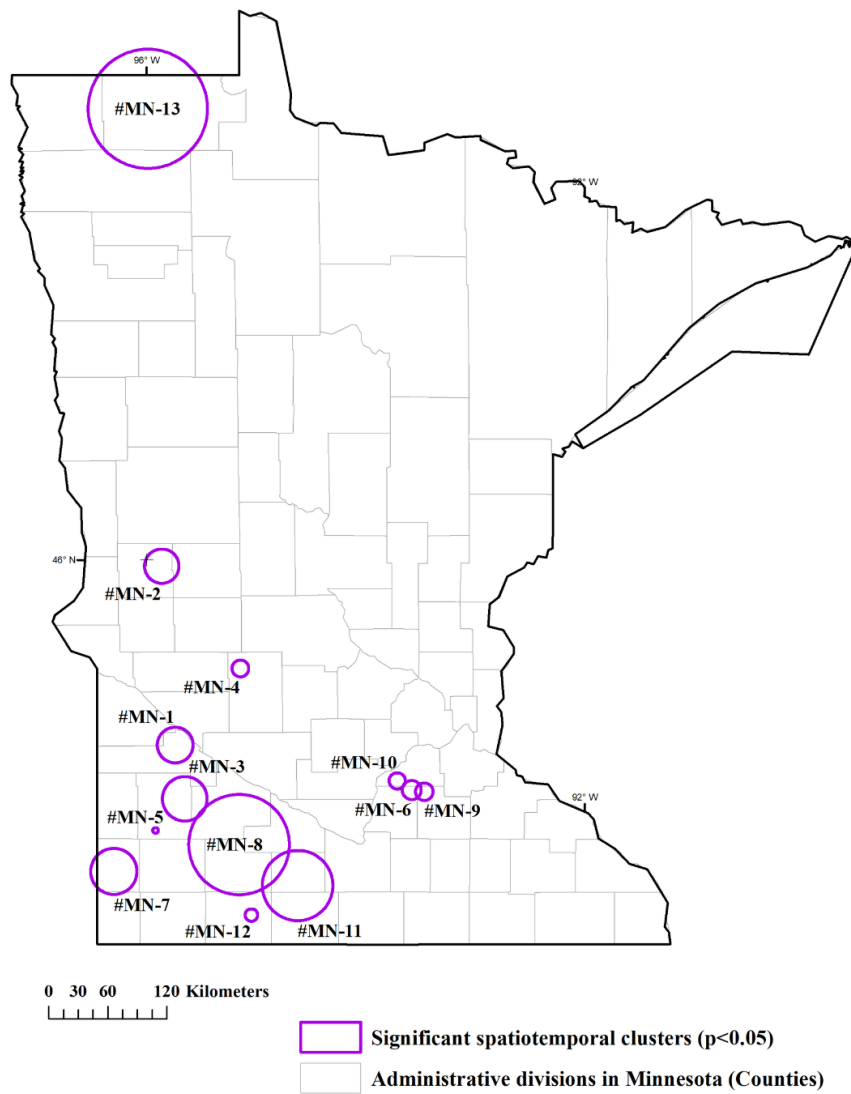
**Fig. 4.3.** Maps illustrating the distribution of animal Anthrax at certain time intervals.

Panels a, b, and c represents cases in Minnesota between 1912 and 2014. Panels d, e, and, f represents cases in Kazakhstan between 1933 and 2014. According to the spatiotemporal directionality test performed using the relative time matrix, a northeastern directionality ( $42^\circ$ ) was found in Minnesota and a southwestern directionality for Kazakhstan ( $260^\circ$ ) with low ( $\sim 0.1$ ) though significant ( $p < 0.05$ ) AC values in both locations.

The spatiotemporal permutation model of the scan statistic detected 13 significant ( $p < 0.05$ ) clusters in Minnesota with spatial and temporal sizes spanning between 2 and 80 km and 1 to 2 years and observed-to-expected ratios (O/E) ranging between 4 and 63 (Fig. 4.4 and Table 4.1). There was a location where Anthrax was first reported in 1941 and then again in 1996, i.e. 55-years apart. When the recurrence occurred in 1996 this area resulted in an outbreak, as indicated by cluster #MN-12. In Kazakhstan 17 significant ( $p < 0.05$ ) spatiotemporal clusters with radii ranging between 18 and 309 km and temporal lengths between 1 to 3 years, and O/E between 3 and 57 were obtained ( $p < 0.05$ ) (Fig. 4.5 and Table 4.1). The location where Anthrax was reported 60 years apart, i.e. first in 1947 and again in 2007, was located within the area captured by cluster #KZ-4, towards the west of Astana, Kazakhstan. When the spatiotemporal cluster radii were compared between the sites (Table 4.1), relatively smaller spatiotemporal clusters ranging between 2.07 to 39.9 km (median=12.1 km and average= 14.8 km) were found for Minnesota, whereas, Kazakhstan's clusters ranged between 18.4 to 309.2 km (median = 66.6 km and average=98.0 km) (Supplementary Fig. S4.1).

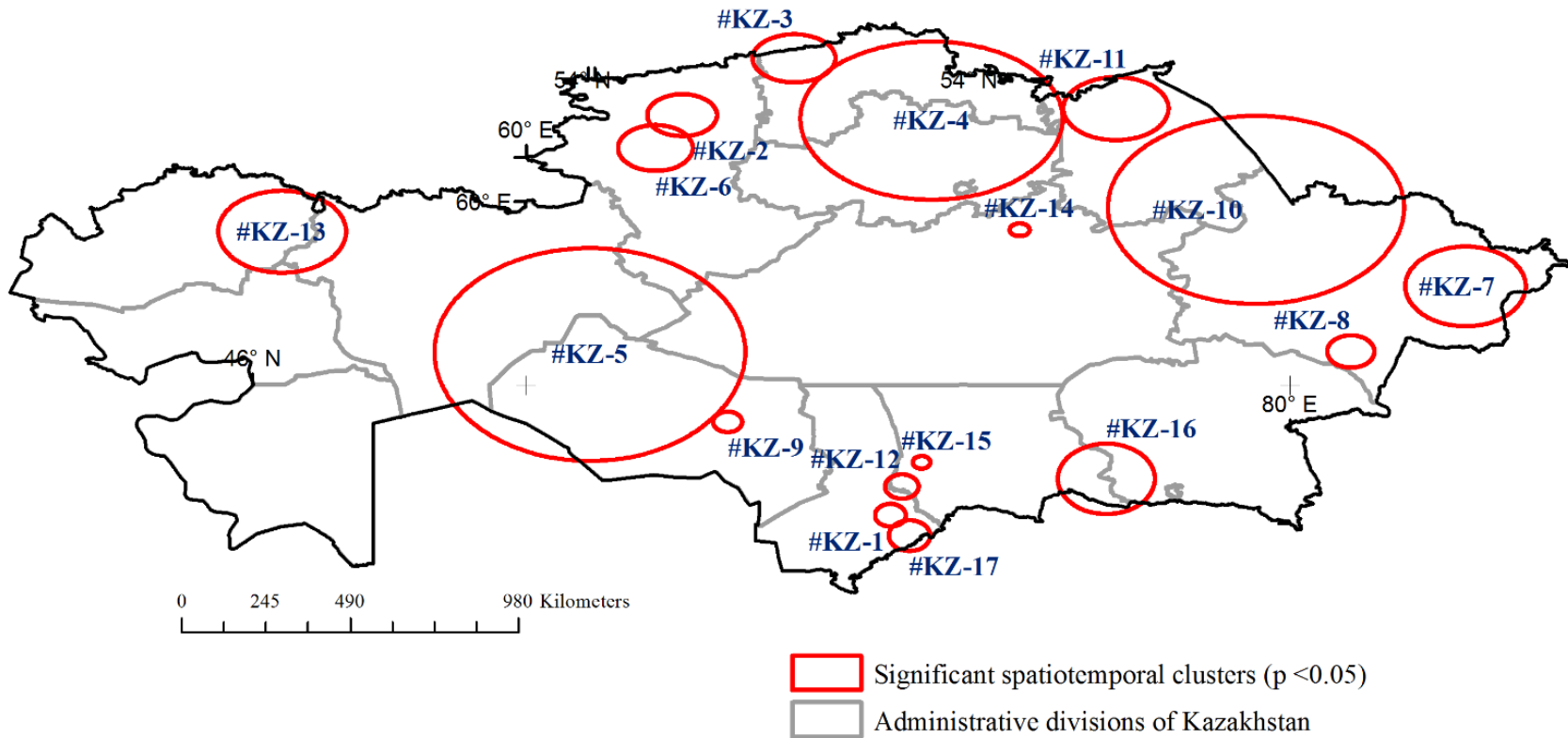
The sensitivity analysis (Supplementary Fig S4.2) revealed no changes in the approximate location of the clusters. With space-time window sizes 10% and 5%, the location and the range of cluster radii were comparable. However, when used 5% space-time window, three separate clusters were detected in northwestern Minnesota: one cluster in 2001 and two in 2006. Whereas, in Kazakhstan, with the 5% space-time window parameters, two clusters with  $< 1$  km were detected in the southeastern region of the country in year 2000.

The summary of characteristics of the spatiotemporal patterns recognized through above analyses, representing the disease progression (epidemic curves), intensity (G-statistics), direction (spatiotemporal directionality test), and recurrence i.e. disease hot spots (spatiotemporal cluster analysis) were compared between the sites (Table 4.2). The epidemiological factors and relevant trigger events associated with spatiotemporal clusters are summarized in Supplementary Table 4.1 and 4.2.



**Fig. 4.4. Spatiotemporal clustering of animal Anthrax cases in Minnesota (time: 1912-2014).**

Clusters were detected using the space-time permutation model of the spatial scan statistic with the spatial-window set to 10% and 5%-time window. Clusters are numbered according to the order of the start date of the outbreak.



**Fig. 4.5. Spatiotemporal clustering of animal Anthrax cases in Kazakhstan (time: 1933-2014).** Clusters were detected using the space-time permutation model of the spatial scan statistic with the spatial- and time-windows set to 10% and 5%, respectively. Clusters are numbered according to the order of the start date of the outbreak.

**Table 4.1.** Summary of spatiotemporal clusters of Anthrax detected in Minnesota (between 1912 and 2014) and Kazakhstan (between 1933 and 2014), using the spatiotemporal permutation model of the scan statistics. The spatial and temporal windows of the scan statistic were set to 10% and 5% respectively.

Cluster ID	Radius (km)	Number of case locations within the cluster	Start (Year)	End (Year)	Observed-to-expected ratio (O/E)
<b>Minnesota</b>					
#MN-1	12.88	22	1919	1919	5.72
#MN-2	12.14	8	1923	1924	28.44
#MN-3	16.16	5	1925	1926	29.39
#MN-4	5.99	8	1931	1932	19.20
#MN-5	2.07	5	1932	1933	16.00
#MN-6	6.91	6	1933	1933	12.08
#MN-7	16.89	13	1937	1938	5.07
#MN-8	36.61	13	1938	1939	6.16
#MN-9	6.36	6	1945	1946	27.43
#MN-10	5.88	3	1948	1948	72.00
#MN-11	25.80	4	1953	1953	26.18
#MN-12	4.59	4	1996	1997	48.00
#MN-13	39.87	14	2001	2001	10.61
<b>Kazakhstan</b>					
#KZ-1	32.94	18	1934	1937	10.77
#KZ-2	61.04	15	1935	1938	34.48
#KZ-3	69.90	24	1940	1940	57.41
#KZ-4	229.99	270	1944	1947	5.38
#KZ-5	309.19	38	1945	1947	7.70
#KZ-6	66.65	15	1949	1952	9.50
#KZ-7	115.82	43	1951	1954	4.59
#KZ-8	46.97	23	1954	1957	4.89
#KZ-9	29.86	6	1956	1958	10.65
#KZ-10	273.36	157	1956	1959	2.70
#KZ-11	91.66	36	1968	1969	5.61
#KZ-12	35.48	10	1971	1971	13.09
#KZ-13	119.16	42	1981	1983	8.65
#KZ-14	18.67	4	1982	1983	41.98
#KZ-15	18.40	7	1985	1987	35.13
#KZ-16	102.38	51	1989	1992	5.25
#KZ-17	44.72	36	2000	2001	7.81

**Table 4.2.** Summary of the similarities and differences of spatiotemporal patterns in Minnesota and Kazakhstan.

	<b>Minnesota, USA</b>	<b>Kazakhstan</b>
<b>Similarities</b>		
Endemicity	√	√
Seasonality with a peak in August	√	√
The time-span of anthrax recurrence at a given location	55 years apart	60 years apart
Spatially overlapping spatiotemporal clusters indicating recurrence of outbreaks in the same region (“Hot spots”)	√	√
River floods over the previous burial sites	√	√
Decreasing trend of Anthrax cases due to control measures	√	√
<b>Differences</b>		
Distribution of cases over time	Sporadic	Long-term epidemic
Relative time spatiotemporal directionality of Anthrax spread	Northwestern directionality (42°)	Southwestern directionality (260°)
Changes in livestock production and distribution	Establishment of livestock production in south central Minnesota: Beef and dairy cattle	Increase of livestock production since 1934
Livestock movement patterns	State government regulated	Seasonal movements with nomadic migrations (Historically)
Open-range summer livestock grazing	Reduced after mid-20 <sup>th</sup> century	Continued practice
Vaccination	As needed. Preventive vaccination is recommended in the areas where Anthrax was reported previously	Inadequate or absence of vaccination before 1950’s. Mass vaccination starting late 1960’s.
Drastic changes in land use	N/A	Virgin lands campaign
Radii of spatiotemporal clusters	Ranging between 2.07 to 39.9 km with a median of 12.1 km	Ranging between 18.4 to 309.2 km with a median of 66.6 km

## 4.5 Discussion

The study compared historic animal Anthrax occurrence in two endemic areas at similar northern latitudes in the western and eastern hemispheres, where socio-economic and demographic conditions were drastically different, with the objective of comparing spatiotemporal patterns of disease progression, intensity, direction, and recurrence (disease hot spots) in relation to potential trigger events. We expect that the analysis here would be useful when suggesting control zone boundaries for each site, based on the reported data.

The progression and intensity of Anthrax was site-specific with few comparable characteristics between the assessed settings; sporadic Anthrax outbreaks were observed in Minnesota, whereas a long-term epidemic was reported in Kazakhstan. This long-term epidemic in Kazakhstan may be attributable to the absence and inadequacies of mass animal vaccination during early years ([Abdrakhmanov et al., 2017a](#); [Abdrakhmanov et al., 2017b](#)). According to G-rate indicator values of the heterogeneity, in Kazakhstan it was 1.5 times higher than in Minnesota (i.e. the difference of G-rates is  $0.0178 - 0.012 = 0.0066$ ). This difference in heterogeneity further indicates the requirement of mass vaccination and control strategies in Kazakhstan compared to Minnesota. In alignment with the literature from the northern hemisphere ([Hugh-Jones et al., 2002](#)), the majority of the outbreaks were reported during the summer months in the regions with peaks in August (Fig. 4.2), suggesting an association with the practice of livestock grazing during summer (approx. May through end of September). The summer grazing practice, especially before the mid-20<sup>th</sup> century, coincides with the potential exposure to pasture/soil-contaminated with pathogens ([Sorge et al., 2015](#); [Morgan et al., 2006](#)).

The relative directionality of long-term trends in Anthrax cases in Minnesota suggested a northeastern directionality of Anthrax compared to the early cases of the disease, which were reported in the south-west of Minnesota. There was a sudden change in the geographical area in which Anthrax cases were reported in the State after 1999 (Fig. 4.1). The early Anthrax cases were reported from southwestern Minnesota, in the Minnesota River valley, whereas the outbreaks reported between 2000 and 2014 were reported from the northwestern corner of the state, in the Red River valley. This northwestern trend may involve multiple trigger events during the period immediately before the year 2000 including the Red River flooding of 1997 ([Kunkel et al., 2003](#)), Anthrax outbreaks of North Dakota in 1998 ([Hughes-Jones, 1999](#)) and the potential wildlife behavioral and migration patterns in the Northern Great Plains including Bison and other habitat sharing ungulates (Elk and White tailed deer) ([NPS Bison Report, 2014](#); [Morris et al.,](#)

2016; Kenefic et al., 2009). Whereas, in Kazakhstan, the relative time measure of the spatiotemporal directionality test identified a southwestern directionality indicating higher potential of Anthrax in the southwestern area in the last decades compared to the past. However, Anthrax occurrence in Kazakhstan was observed throughout the country since early years which is potentially attributable to seasonal movements of livestock with nomadic migrations (Robinson et al., 2003).

The recurrence patterns of Anthrax were comparable between the sites. The maximum time-span of anthrax recurrence at the same location in Minnesota and Kazakhstan was 55 and 60 years respectively. *B. anthracis* spores have a high environmental resistance under certain environmental conditions such as depth and organic content of contaminated soil, and a half-life of nearly 100 years has been reported (Halvorson, 1997). The active lifecycle of *B. anthracis* in soil involving bacteriophage-mediated ecological adaptations also suggested to play a role (Schuch and Fischetti, 2009). Occurrence of cases in the same locations could be therefore due to disturbance of buried carcasses containing viable spores but also could be a result of recurrent recontamination of the soil due to unappreciated death of anthrax-infected animals (Hugh-Jones et al., 2002; Bellan et al., 2013).

Among the 13 spatiotemporal clusters detected in Minnesota, six were spatially overlapping, indicating potential hot spot areas: Clusters #MN-6, 9, and 10 and Clusters #MN-3, 8, and 11. The major trigger events/factors that are likely contributing to Anthrax outbreaks in Minnesota include: 1) favorable soils in western and northwestern Minnesota (Blackburn et al., 2007), 2) historic river flooding events in the Minnesota and Red rivers that contribute to spread of pathogens and spores to extensive areas (Kunkel et al., 2003), and 3) changes in distribution of susceptible livestock populations and management including extensive beef cattle production during early 1900's in the southern prairies of Minnesota and summer grazing (Sorge et al., 2015; Crickman, 1934) (Supplement Table S4.2). The declining trend of recent cases in Minnesota may be attributable to direct and indirect preventive measures including: 1) flood control programs implemented targeting Minnesota and Red river areas (Hearne and Kritsky, 2010; Anonymous 1980); 2) prompt vaccination (Ward, 1914); and reduced summer grazing and increased feedlot management practice during late 2000's (Sorge et al., 2015).

There were 17 spatiotemporal clusters in Kazakhstan, among which eight were overlapping indicating potential hot spot areas: #KZ-1, 12, 15, and 17; #KZ-2 and 6; and #KZ-5 and 9. The major trigger events/factors that are likely contributors to Anthrax outbreaks in Kazakhstan include: 1) favorable soils (Abdrakhmanov et al., 2017a); 2) insufficient vaccination



coverage or absence of vaccination (Abdrakhmanov et al., 2017a); 3) livestock growth since 1934 (National Committee of Statistics (In Russian): <http://istmat.info/node/21348>) including the influx of inhabitants along with their livestock from other USSR regions (Adamovich and Nikonov, 1970; McCauley, 1976; Rowe, 2011); 4) seasonal movements of livestock with nomadic migrations, historically (Robinson and Milner-Gulland, 2003); 5) land use, especially the Virgin Lands Campaign which lead to disturbance of previously unused landscapes where some of the previous Anthrax burial sites were located, which may have resulted in bringing the anthrax spores onto the surface and its intensive wind mediated spread over long distances (Turnbull et al., 1998; Robinson and Milner-Gulland, 2003; McCauley, 1976; Rowe, 2011); 6) river flooding, especially the Syrdarya river flooding which might be related to disturbances at old carcass burial sites (Abdrakhmanov et al., 2017b; Arkhipkin et al., 2010); 7) range-land based livestock production (Abdrakhmanov et al., 2017b; Bourn, 2004) where susceptible populations may come in contact with the contaminated soils/pasture; and 8) intense agriculture in the southern Kazakhstan where high human population density, intensive agriculture, together with livestock population collectively may have resulted in frequent disturbance of anthrax burials and animal contacts with the pathogen (Adamovich and Nikonov, 1970) (Supplement Table S4.3). The declining trend of the recent cases in Kazakhstan may be attributable to direct and indirect preventive measures including: mass vaccination and mandatory disposal by burning mandatory disposal by burning carcasses of animals that had died from anthrax (Lukhnova et al., 2004); 2) development and implementation of a whole complex of preventive, antiepidemic and antiepidemic measures, continuous epizootic monitoring and recording of anthrax burial sites (Abdrakhmanov et al., 2017b).

The sensitivity analysis indicated that the location and range of the radii remained compatible at both 10% and 5% space-time window parameter settings (Supplement Fig S4.2). One of the three clusters detected with the 5% space-time parameter setting in northwestern Minnesota clusters indicated a known outbreak of Anthrax occurred in 2006. However, the detection of clusters of <1 km in southeastern Kazakhstan, with the 5% space-time setting, did not indicate meaningful outbreaks. Therefore, the sensitivity analysis supported the choice of maximum window sizes of 10% spatial window and 5% temporal window signaling the value of providing a sensitivity analysis per region to optimize detection of clusters.

The spatiotemporal cluster radii may be useful when understanding the potential extent of Anthrax spread as well as determining control and surveillance zones in endemic settings. Based on the distribution of spatiotemporal cluster radii (Table 4.1 and Supplement Fig. S4.1), we

suggest that the surveillance or control zones for Minnesota may consider a range of 2 to 40 km with a median of 12 km for preventive activities such as vaccination. As per the current recommendations of Minnesota Board of Animal Health, spring vaccination is recommended in the Northwestern part of the state of Minnesota. Once a confirmed case is reported, the recommendations include: vaccination and administration of antibiotics as needed to safeguard exposed animals in the herd, notification and recommendation for vaccination of susceptible animals within 10 miles (~16 km) and recommendation to monitor the herds within 10 to 30 miles (16 to 48 km). Further details regarding the vaccine recommendations for Midwestern USA and adjacent Canadian states, i.e. The Unified Anthrax Recommendations, are found elsewhere (Stoltenow 2012). Vaccine recommendations are suggested considering the potential underreporting or delayed reporting scenarios where contamination of Tabanid fly mouth parts with the pathogen and thereby transport spores to unvaccinated susceptible animals on neighboring areas (Stoltenow, 2012). Similarly, the ideal surveillance or control zones for Kazakhstan may have a range between 18 to 310 km, with a median of 67 km. The current recommendations against animal Anthrax in Kazakhstan involve mass vaccination of all the susceptible animals since 1961 and twice-annual vaccination recommended since 1982 (Popov et al., 2002).

The limitations of this study include: the use of officially reported data (therefore potentially missing certain cases, and leading to an uncontrolled reporting bias), the lack of genetic information that prevents discriminating between *B. anthracis* strains (Van Ert et al., 2007), and the absence of information on neighboring states or countries in the analysis. Moreover, the recordkeeping of historic cases inevitably varies by country. Therefore, the findings presented here should be interpreted in light of these limitations. We recognize the importance of genetic diversity in determining Anthrax spread. Still, the comparison made here may still be useful given that Anthrax strains are relatively homogenous within affected regions (Van Ert et al., 2007; Aikembayev et al., 2010), and both the Eurasian and North American sub-lineages belongs to the major lineage A and the two sub-lineages are closely related to each other (Van Ert et al., 2007).

Through this analysis, we have demonstrated that spatiotemporal patterns of anthrax occurrence in two endemic regions may be attributable to specific trigger events. Events such as river floods contributing to the spread of the pathogen and presence of disease hot spots were found at both the sites. However, there were also considerably different factors involved in shaping the site-specific spatiotemporal patterns (Tables 4.2 and 4.3) attributable to the site-

specific environmental, demographic, and agroecological changes driving occurrence of Anthrax in endemic settings, despite the expected similarities of the pathogen's behavior relative to the environmental characteristics. Ultimately, results suggest that it is important to consider the probability of recurrence, locations of disease hot spots, and the extents of climatic, environmental, agricultural, and demographic trigger events, both over space and time, when planning strategies for the prevention and control of the disease in endemic settings.

#### **4.6 Conclusions**

Over the last hundred years Anthrax occurrence in Minnesota was sporadic whereas Kazakhstan experienced a long-term epidemic. The seasonality was similar between sites, with a peak in August. Declining number of cases at both sites is attributable to vaccination and control measures. This study detects Anthrax hotspots in both the settings, with spatially overlapping clusters years apart. Distribution of the spatiotemporal cluster radii between study sites supported suggestion of site-specific control zones. Spatiotemporal patterns of Anthrax occurrence in the two endemic regions were attributed to multiple potential trigger events including major river floods, changes in land use, agriculture, and susceptible livestock populations. Results here may help to understand the long-term epidemiological dynamics of Anthrax while providing suggestions on the surveillance or control zones in endemic settings.

#### **4.7 Acknowledgements**

This study was funded in part by: 1) the Minnesota Discovery, Research, and Innovation Economy (MnDRIVE) program of the Office of the Vice President for Research (OVPR) of the University of Minnesota and 2) Scientific Thematic “Zonification of Kazakhstan according to biosecurity categories with regard to dangerous infectious animal diseases” under the Program #249 of funding scientific researches in agro-industry and environmental management by the Ministry of Agriculture of Kazakhstan.

## 4.8 References

1. Abdrakhmanov SK, Mukhanbetkaliyev YY, Korennoy FI, Sultanov AA, Kadyrov AS, Kushubaev DB, et al. 2017a. Maximum entropy modeling risk of anthrax in the Republic of Kazakhstan. *Preventive Veterinary Medicine*. 144:149-57. doi: 10.1016/j.prevetmed.2017.06.003. PubMed PMID: WOS:000407406600018.
2. Abdrakhmanov SK, Mukhanbetkaliyev YY, Korennoy FI, Karatayev BS, Mukhanbetkaliyeva AA, Abdrakhmanova AS. 2017b. Spatio-temporal analysis and visualisation of the anthrax epidemicsituation in livestock in Kazakhstan over the period 1933-2016. *Geospatial Health*. (12):2. doi: 10.4081/gh.2017.589.
3. Adamovich VL and Nikonov NN. 1970. The importance of landscape-ecological factors in the epizootology of anthrax. *Message 2: Comparative method of assessment of epizootic tension of aterritory. The journal of Microbiology*. 8:113 – 7.(in Russian)
4. Aikembayev AM, Lukhnova L, Temiraliyeva G, Meka-Mechenko T, Pazylov Y, Zakaryan S, et al. 2010. Historical Distribution and Molecular Diversity of *Bacillus anthracis*, Kazakhstan. *Emerging Infectious Diseases*. 16(5):789-96. doi: 10.3201/eid1605.091427. PubMed PMID: WOS:000277209900006.
5. Anonymous. 1980. Lac Qui Parle flood control project, Master plan for public use, Development and resource management. U.S. Army Corps Engineers, St Paul district, Minnesota. AD-A120 745.
6. Arkhipkin OP, Spivak LF, Sagatdinova GN. 2010. Development of Flood Space Monitoring in Kazakhstan. *Geoscience and Remote Sensing, New Achievements*. 419-36. PubMed PMID: WOS:000363695000023.
7. Bellan SE, Gimenez O, Choquet R, Getz WM. 2013. A hierarchical distance sampling approach to estimating mortality rates from opportunistic carcass surveillance data. *Methods in Ecology and Evolution*. 4(4):361-9. doi: 10.1111/2041-210x.12021. PubMed PMID: WOS:000317021300007.
8. Bender J, Cutting K, Miller B, Friendshuh K, Boldingh T, Nelson J, Danila R. 2006. *Historical Anthrax Among Animals and Humans in Minnesota, 1912 – 2005*.

International Conference on Emerging Infectious Diseases, abstr p.156. Atlanta, GA March 19-21.

9. Blackburn JK, McNyset KM, Curtis A, Hugh-Jones ME. 2007. Modeling the geographic distribution of *Bacillus anthracis*, the causative agent of anthrax disease, for the contiguous United States using predictive ecologic niche modeling. *American Journal of Tropical Medicine and Hygiene*. 77(6):1103-10. PubMed PMID: WOS:000252123200022.
10. Bourn D. 2013. Draft guidelines on the enumeration of nomadic and semi-nomadic (transhumant) livestock.
11. Chen X, Wang K. 2017. Geographic area-based rate as a novel indicator to enhance research and precision intervention for more effective HIV/AIDS control. *Preventive Medicine Reports*. 5:301–7.
12. Crickman CW. 1934. Beef cattle production in Minnesota [Internet]. *Minnesota Bulletin: University of Minnesota Agricultural Experiment station and United States Department of Agriculture Bureau of Agriculture Economics*. [8-20]
13. Department of the Interior Bison Leadership Team and Working Group and Biological Resource Management Division, National Park Service. 2014. DOI Bison Report: Looking Forward. Natural Resource Report NPS/NRSS/BRMD/NRR - 2014/821. NPS 909/124952. National Park Service. Fort Collins, Colorado.
14. Driks A. 2009. The *Bacillus anthracis* spore. *Molecular Aspects of Medicine*. 30(6):368-73. doi: 10.1016/j.mam.2009.08.001. PubMed PMID: WOS:000272755000004.
15. Elvander M, Persson B, Lewerin SS. 2017. Historical cases of anthrax in Sweden 1916-1961. *Transboundary and Emerging Diseases*. 64(3):892-8. doi: 10.1111/tbed.12456. PubMed PMID: WOS:000400161300023.
16. Halvorson HO. 1997. Two generations of spore research: from father to son. *Microbiologia*. 13:131-48.

17. Hearne RR, Kritsky CC. 2010. Characteristics of active local water management districts in the Red River Basin. *Water Policy*. 12(6):898-912. doi: 10.2166/wp.2010.145. PubMed PMID: WOS:000282515300008.
18. Hughes-Jones M. 1999. 1996-97 global anthrax report. *Journal of Applied Microbiology*. 87(2):189-91. doi: 10.1046/j.1365-2672.1999.00867.x. PubMed PMID: WOS:000082659900002.
19. Hugh-Jones M, Blackburn J. 2009. The ecology of *Bacillus anthracis*. *Molecular Aspects of Medicine*. 30(6):356-67. doi: 10.1016/j.mam.2009.08.003. PubMed PMID: WOS:000272755000003.
20. Hugh-Jones ME, de Vos V. 2002. Anthrax and wildlife. *Revue Scientifique Et Technique De L Office International Des Epizooties*. 21(2):359-83. doi: 10.20506/rst.21.2.1336. PubMed PMID: WOS:000174946300011.
21. Kenefic LJ, Pearson T, Okinaka RT, Schupp JM, Wagner DM, Ravel J, et al. 2009. Pre-Columbian Origins for North American Anthrax. *Plos One*. 4(3):21. doi: 10.1371/journal.pone.0004813. PubMed PMID: WOS:000265496300003.
22. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. 2005. A space-time permutation scan statistic for disease outbreak detection. *Plos Medicine*. 2(3):216-24. doi: 10.1371/journal.pmed.0020059. PubMed PMID: WOS:000228382500013.
23. Kunkel KE, Easterling DR, Redmond K, Hubbard K. 2003. Temporal variations of extreme precipitation events in the UnitedStates: 1895–2000. *Geophys Res Lett*. 30(1900):10.1029/2003GL018052.
24. Lukhnova LY, Ajkimbaev AM, Pazylov EK. 2004. Epidemic process of Anthrax in Kazakhstan. *Bull. Agric. Sci. Kazakhstan, Bastau*. 7:44. (in Russian).
25. McCauley M. 1976. Khrushchev and the development of Soviet agriculture. *The Virgin Land Programme 1953 – 1964*. Palgrave Macmillan UK. Online ISBN 978-1-349-03059-0. Doi: 10.1007/978-1-349-03059-0

26. Morgan ER, Torgerson PR, Shaikenov BS, Usenbayev AE, Moore ABM, Medley GF, et al. 2006. Agricultural restructuring and gastrointestinal parasitism in domestic ruminants on the rangelands of Kazakhstan. *Veterinary Parasitology*. 139(1-3):180-91. doi: 10.1016/j.vetpar.2006.02.016. PubMed PMID: WOS:000238439100021.
27. Morris LR, Proffitt KM, Asher V, Blackburn JK. 2016. Elk resource selection and implications for anthrax management in Montana. *Journal of Wildlife Management*. 80(2):235-44. doi: 10.1002/jwmg.1016. PubMed PMID: WOS:000369817500007.
28. Mullins JC, Garofolo G, Van Ert M, Fasanella A, Lukhnova L, Hugh-Jones ME, et al. 2013. Ecological Niche Modeling of *Bacillus anthracis* on Three Continents: Evidence for Genetic-Ecological Divergence? *Plos One*. 8(8):8. doi: 10.1371/journal.pone.0072451. PubMed PMID: WOS:000323425700170.
29. Ottorino C, David A, Wolfgang B, Beniyamin C, Mehmet D, Dan D, et al. 2008. Anthrax in humans and animals. ISBN 978 92 4 154753 6.
30. Popov YA, Mikshis NI. 2002. Anthrax vaccines. Problems of especially dangerous infections. *Saratov. № 1 (83)*. - P. 21-36 (in Russian).
31. Robinson S, Milner-Gulland EJ. 2003. Political change and factors limiting numbers of wild and domestic ungulates in Kazakhstan. *Human Ecology*. 31(1):87-110. doi: 10.1023/a:1022834224257. PubMed PMID: WOS:000181621600005.
32. Rowe WC. 2009. Turning the Soviet Union into Iowa: The Virgin Lands Program in the Schuch R, Fischetti VA. *The Secret Life of the Anthrax Agent Bacillus anthracis: Bacteriophage-Mediated Ecological Adaptations*. *Plos One*. 4(8). doi: 10.1371/journal.pone.0006532. PubMed PMID: WOS:000268935900002.
33. Sorge US, Moon RD, Stromberg BE, Schroth SL, Michels L, Wolff LJ, et al. 2015. Parasites and parasite management practices of organic and conventional dairy herds in Minnesota. *Journal of Dairy Science*. 98(5):3143-51. doi: 10.3168/jds.2014-9031. PubMed PMID: WOS:000353267900033. Soviet Union. Brunn S, editor. 2011. *Engineering Earth*. Springer, Dordrecht. doi.org/10.1007/978-90-481-9920-4\_15.

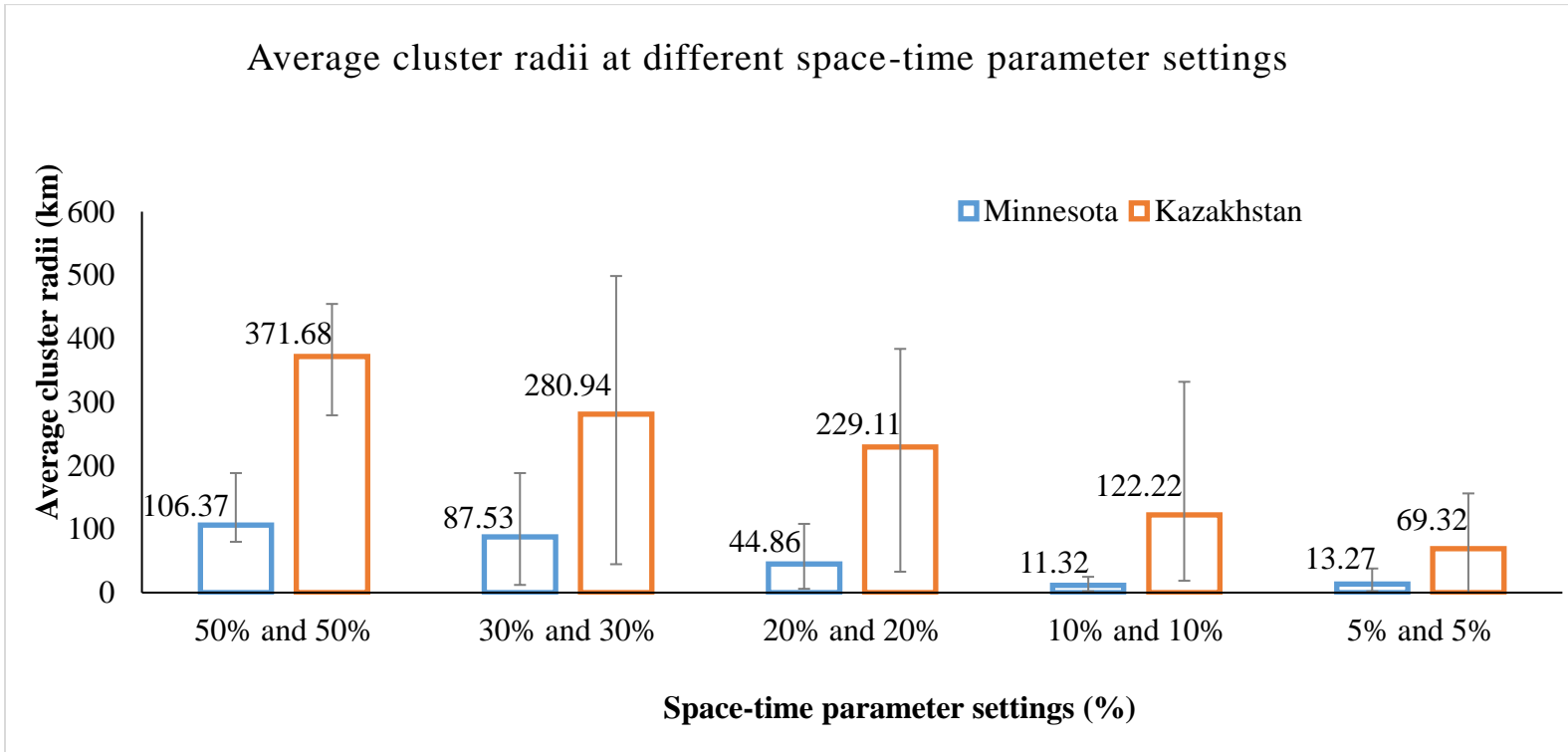
34. Sternbach G. 2003. The history of anthrax. *Journal of Emergency Medicine*. 24(4):463-7. doi: 10.1016/s0736-4679(03)00079-9. PubMed PMID: WOS:000182902600019.
35. Stoltenow CL. 2012. Unified Anthrax Recommendations for MB, MN, MT, ND, & SD. Available online: [https://www.nd.gov/ndda/sites/default/files/legacy/resource/UnifiedAnthraxRecommendations\\_0.pdf](https://www.nd.gov/ndda/sites/default/files/legacy/resource/UnifiedAnthraxRecommendations_0.pdf)
36. Turnbull PCB, Lindeque P, Le Roux J, Bennett AM, Parks SR. 1998. Airborne movement of anthrax spores from carcass sites in the Etosha National Park, Namibia. *Journal of Applied Microbiology*. 84(4):667-76. doi: 10.1046/j.1365-2672.1998.00394.x. PubMed PMID: WOS:000073738600026.
37. Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, Ravel J, et al. 2007. Global Genetic Population Structure of *Bacillus anthracis*. *Plos One*. 2(5). doi: 10.1371/journal.pone.0000461. PubMed PMID: WOS:000207446000009.
38. Ward SH. 1914. Minnesota State Livestock Sanitary Board. 11th Annual report. Syndicate Printing co., Minneapolis. pp 31-33.



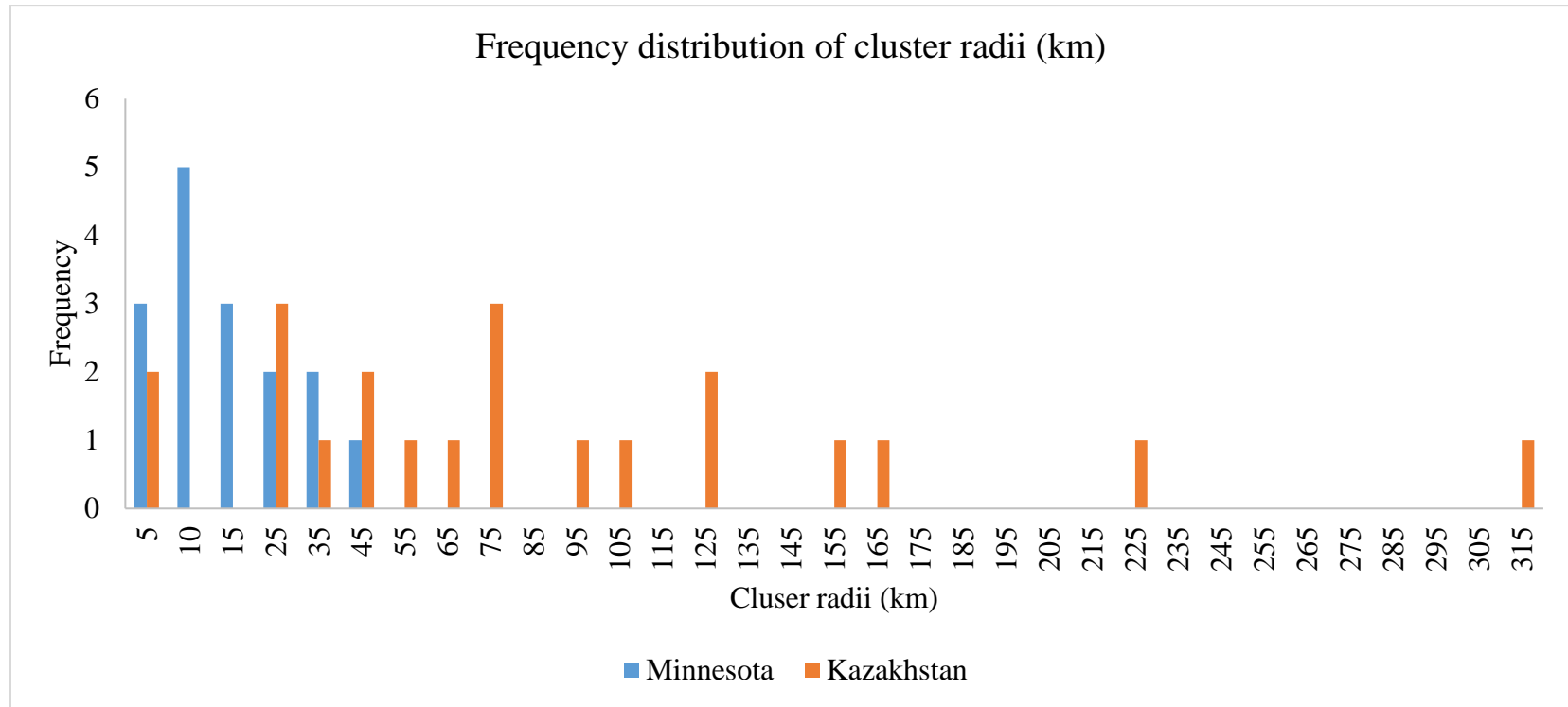


**Supplement Table S4.2.** Summary of the historic events in Kazakhstan related to the patterns how the trigger events i.e., climatic, anthropogenic, agricultural, and environmental changes, which may have led to the recognized spatiotemporal clusters of Anthrax.

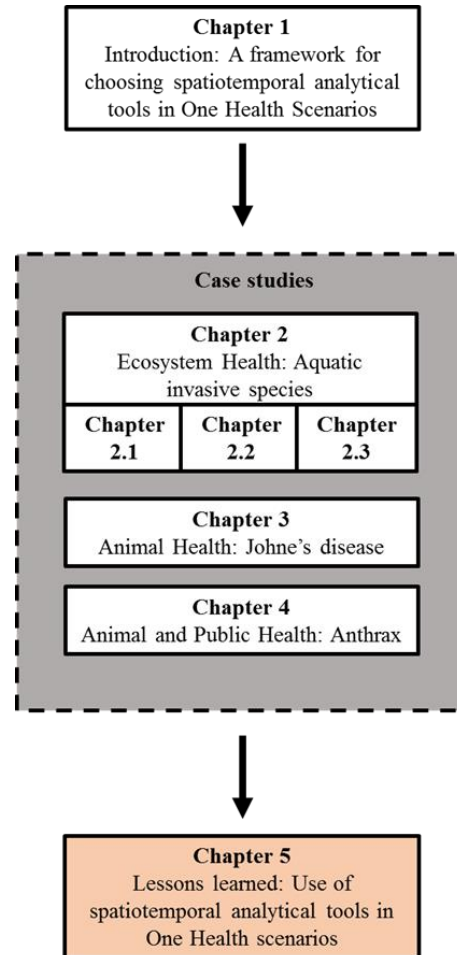
Potential risk factors and trigger event/s	References	#KZ-1	#KZ-2	#KZ-3	#KZ-4	#KZ-5	#KZ-6	#KZ-7	#KZ-8	#KZ-9	#KZ-10	#KZ-11	#KZ-12	#KZ-13	#KZ-14	#KZ-15	#KZ-16	#KZ-17
		1934-1937	1935-1938	1940-1940	1944-1947	1945-1947	1949-1952	1951-1954	1954-1957	1956-1958	1956-1959	1968-1969	1971-1971	1981-1983	1982-1983	1985-1987	1989-1992	2000-2001
<b>1. Vaccine coverage:</b> Insufficient or absent vaccination of susceptible species	<a href="#">Abdrakhmanov et al., 2017a</a> ; <a href="#">Abdrakhmanov et al., 2017b</a>	X	X	X	X	X	X	X	X	X	X	X	X					
<b>2. Livestock populations:</b> Sharp growth of livestock population	<a href="#">Abdrakhmanov et al., 2017a</a> ; <a href="#">Abdrakhmanov et al., 2017b</a>	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
<b>3. Favorable soils:</b> chernozems and kastanozem	<a href="#">Abdrakhmanov et al., 2017a</a>	X	X	X	X		X	X	X		X	X	X	X	X	X	X	X
<b>4. Land Use:</b> Virgin Lands Campaign (between 1954 and 1963)	<a href="#">Robinson and Milner-Gulland, 2003</a>							X	X	X	X	X	X					
<b>5. River flooding:</b> Syrdarya river floods	<a href="#">Abdrakhmanov et al., 2017a</a> ; <a href="#">Abdrakhmanov et al., 2017b</a>					X				X								
<b>6. Intensive agriculture</b>	<a href="#">Rowe, 2009</a>	X				X				X			X			X	X	X



**Supplement Fig. S4.1.** Sensitivity analysis of the spatiotemporal window sizes used in the space-time permutation model of scan statistics.



**Supplement Fig. S4.2** Frequency distribution of the spatiotemporal cluster radii of historic animal Anthrax in Minnesota, USA and the Republic of Kazakhstan. Spatiotemporal clusters were detected using the space-time permutation model of the spatial scan statistic with the spatial- and time-windows set to 10% and 5%, respectively.



## Chapter 5: Lessons learned

### *Risk based management of One Health scenarios using spatiotemporal analytical tools: Lessons learned*

This work has not been published/in-review elsewhere

## 5.1 Overview

The use of spatiotemporal analytical tools to generate risk maps and scores that facilitate early detection of health threats is becoming increasingly popular in many countries and international organizations around the world. The traditional approach of spatial epidemiology focuses on a two stage-process of mapping, to visualize and hypothesize on possible spatial dependence in disease distribution, and, subsequently, test those hypotheses using techniques for clusters detection and identification of associated explanatory factors. The advances in the field of spatial epidemiology are focused on the analysis of (big) data to establish spatiotemporal relations between disease and epidemiological factors to ultimately inform the implementation of strategies for disease prevention and mitigation, and providing risk estimates in a timely manner to support the decision-making process. Supporting decision-making and policy involves an iterative process with stakeholders, aimed at improving and refining the steps of data collection, analysis, discussion of results, and ultimately, design and implementation of risk management and policy measures. However, the process of researcher-stakeholder interactions are poorly documented in the veterinary literature and, in general, is driven by the good will and common sense of researchers and stakeholders without the involvement of a formal methodological process for evaluation. The study here summarizes and compares researcher-stakeholder interactions related to three case studies, which made use of spatiotemporal analytical tools, to address three One Health scenarios in the State of Minnesota. Therefore, the comparison between the case studies was focused on the lessons learned through the researcher-stakeholder interactions and identification of the opportunities and challenges in the use of spatiotemporal eco-epidemiological tools in One Health processes. Ultimately, the discussion of the three case studies presented here, serves as precedent for establishing a protocol of “good practices” when promoting researcher-stakeholder interactions to inform the implementation of science-driven management and policy solutions to One Health issues.

## 5.2 Introduction

### 5.2.1 One Health and spatial epidemiology

The concept of One Health encompasses the interdisciplinary approach of unifying common scientific approaches of health-related research at the interface of ecosystems, animals, and humans (Murtaugh et al., 2017). The objective of One Health research is to generate

integrated knowledge in the assessment of health threats and implementation of mitigation strategies that address all those three elements (ecosystem, animal, and human health)([Rabinowitz et al., 2013](#)). One of the commonalities among One Health scenarios is their patterns of translocation across geographical spaces ([Cunningham et al., 2017](#)). These translocations are partially attributable to increasing volumes and rates of human travel and globalized trade ([Morse, 1993; Cunningham et al., 2017](#)). Spatiotemporal epidemiological tools are useful in identifying those translocation patterns, quantifying the association of the translocation with underlying risk factors, and eventually recognizing high-risk areas or predicting the adverse outcomes in non-sampled locations ([Clements and Pfeiffer, 2009](#)).

Spatial epidemiology is defined as “the description and analysis of geographic variations in disease with respect to demographic, environmental, behavioral, socioeconomic, and infectious risk factors” ([Elliot and Wartenberg 2004](#)). Integration of epidemiological concepts, statistics, spatial analysis, and geographic information system (GIS) enables achieving the objectives of spatial epidemiology, by analysis of adverse events to eventually provide risk estimates in a timely manner to support the design and implementation of decision and policy strategies for disease prevention and control measures ([Cromley and McLafferty, 2002; Wagner Robb et al., 2016; Pfeiffer and Stevens, 2015](#)).

### **5.2.2 Improvement of data quality while using existing data**

A variety of large and complex databases, referred to as “big data”, on environmental, animal, and public health threats are routinely collected by multiple stakeholder institutions. Compared to traditional small research data sets, big data are complex, dynamic, and often spatiotemporally explicit, therefore, demanding the use of sophisticated analytical tools ([Perez 2015; Asokan and Asokan, 2015; VanderWaal et al., 2017](#)). Analysis of data routinely collected by public institutions often results in periodical reports and summaries, but those dataset are only rarely used to discover long-term spatiotemporal trends that could potentially benefit the decision-making processes. Such limited use of big data is attributable to the logistical constraints in maintaining these databases, and to the difficulty of training personnel to collect, analyze, monitor, update, and interpret the overall outcomes of the big data ([VanderWaal et al., 2017](#)). However, recent technological advancements have made sophisticated spatiotemporal analytical tools accessible and user-friendly to the researchers and policy makers that could support the mitigation of the environmental and health threats ([Carroll et al., 2014; Asokan and Asokan, 2015; VanderWaal et al., 2017](#)).

Although there are numerous tools (i.e. analytical methods and software) available, their use alone do not lead necessarily to meaningful conclusions if poor-quality data on risk factors, disease status, or population data were used in the first place (Beale *et al.* 2008). Yet, existing research data and/or big data collected by ecosystem, animal, or public health institutions offer an affordable and relatively simple starting point to understand the distribution and determinants of a variety of One Health scenarios. Demonstrating the use and utility of pre-existing data to produce preliminary results also serves to plan for the improvement of the quality of the data to be collected in the future.

### 5.2.3 Estimation of risks using spatial epidemiological tools

Risk is defined as the probability that an event with negative consequences occurs, as well as the magnitude of those negative consequences (Dufour *et al.*, 2011). Spatiotemporal analyses enable hypothesizing and predicting of the intensity of adverse events in relation to underlying factors (Lawson *et al.* 1999; Elliot and Wartenberg 2004). In addition to the recognition of various risk factors and the estimation of the health risk associated with them, spatial epidemiology may also play a role in improving risk perception and risk communication by providing a basis for better surveillance and prioritization of control measures through risk regionalization.

Outputs of spatiotemporal analyses include risk estimates and risk maps, depicting areas of high risk for disease, often referred to as “disease hot spots”, and risk scores over space and time. Risk maps are visualization tools that serve as effective tools for communicating decision-makers on the probability that a disease outbreak or environmental threat may occur in a given space and time.

Risk regionalization has multiple advantages when designing surveillance programs or management strategies for prevention and control of harmful agents (Moore and Carpenter, 1993). Moreover, defining and isolation of “disease-free areas” in endemic countries through risk assessment, as was done for Foot and Mouth Disease (FMD) (Shanafelt and Perrings, 2018), is recognized as an important aspect of animal and by product trade at the international level (OIE Terrestrial Animal Health Code Chapter 11.4, 2018).

Decision makers could use risk scores for status quo evaluation of the current strategies and compare those to alternative risk-based surveillance or mitigation strategies. For example, the comparison of current distribution of resources among administrative divisions may be based on



human population density, whereas, the risk maps based approach would provide a different suggestion for the distribution of the resources. Admittedly, it is ideal to have these status quo comparisons evaluated for cost-effectiveness of specific management strategies. One approach to increase the efficiency of active surveillance is targeting high-risk areas in the population, commonly known as “risk-based surveillance” or “targeted sampling”. Similarly, risk-based management, defined as the “decision-making process for identifying, evaluating, selecting, prioritizing, and implementing control measures” (CRARM 1997), may benefit from the design of risk maps. Thus, it is essential that stakeholders who are involved in the mitigation of health threats, understand the science behind the risk mapping process. Similarly, the researchers conducting investigations would benefit from understanding the policy/decision-making process and the applicability of diverse One Health approaches through extensive and iterative with scientists and stakeholders/decision-makers.

#### **5.2.4 Communication with stakeholders in the decision-making process**

Stakeholders in the policy/decision-making process may be defined as “any person or group who has an interest in the research topic and/or who stands to gain or lose from a possible policy change that, directly or indirectly, might be influenced by the research findings” (Slunge et al., 2017). The relationship between researches and stakeholders is critical for the translation of hypothesis-driven research into effective and efficient management decisions and policy to prevent or mitigate disease impact in the field. Use of spatiotemporally explicit (big) data and quantitative approaches in spatiotemporal epidemiology to address long-term health and economic benefits is only achievable if relevant stakeholders in the environmental, veterinary, and public health sectors understand each other’s values.

For convenience, this study uses the term “research/er” to denote spatiotemporal analysis and modeling outputs of the studies here and the scientists involved in conducting them onwards. Whereas, the organizations or personnel who would potentially use resulting risk maps and risk estimates are referred to as “stakeholders.” Stakeholders involved in the decision-making process of One Health scenarios represent multi-sectoral and multidisciplinary backgrounds. Communicating the spatial analytical outcomes with stakeholders, receiving their feedback, discussing potential use of the scientific outcomes, and improving the data collection process and use of relevant analytical tools is an iterative process (Cartwright et al., 2016; Goodman and Thompson, 2017). Therefore, it is imperative to get stakeholders involved in the research from the beginning, when the research studies are planned, to help understand and account for field or

practical challenges (Slunge et al., 2017). The iterative nature of researcher-stakeholder interactions occur through updates, discussions, feedback, and re-discussion.

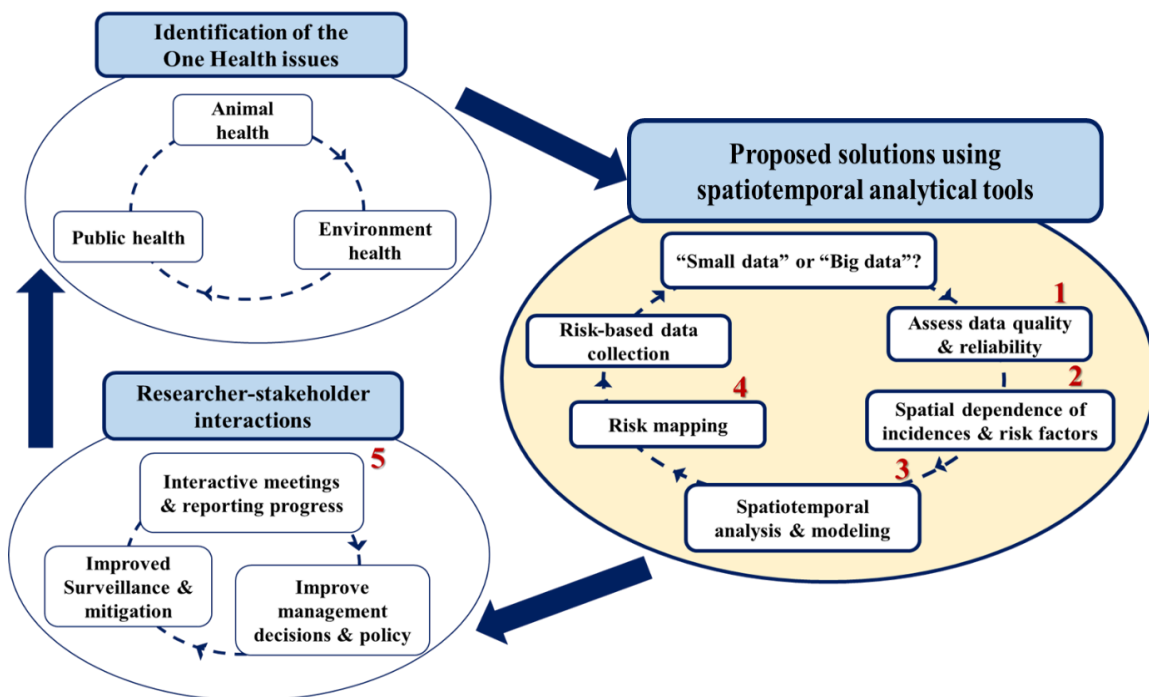
It is essential to understand that maps are representations of an analyst's view of data rather than a precise representation of the reality (Cromley and McLafferty, 2002). Depending on the modeling criteria, assumptions, and the performance of the method (i.e. measures of prediction powers such as sensitivity and specificity), the risk maps produced from the same data may look significantly different from each other (Wagnet and Huston, 1996; Cromley and McLafferty, 2002). Therefore, it is crucial for both researchers and stakeholders to have a mutual understanding of the methods, assumptions, and risk classification criteria used to generate risk maps. Additionally, the communication provides opportunities to understand biases associated with reporting and sampling and the implications of those biases when interpreting the risk estimates (Cromley and McLafferty, 2002). These communications would further benefit by defining threshold risk values when identifying high-risk areas, based on expert opinions of both researchers and stakeholders.

The multiple benefits of communicating scientific findings with stakeholders include: (1) sharing the scientific knowledge with stakeholders while receiving their feedback on the potential use of findings which may improve the decisions in mitigating health threats; (2) understanding the constraints in comparison to the existing strategies of decision and policy making and how difficult (logistically, financially, and legislatively) is to change existing approaches; and (3) planning strategic and systematic involvement of stakeholders in the process of scientific research from the beginning and focus on the improvement of decision and policymaking process (Slunge et al., 2017). However, the step that involves communication between researcher and stakeholders regarding scientifically based decision-making has been poorly documented in the scientific literature.

### **5.2.5 Objective**

The objective of the work here was to discuss three examples, or case studies, on One Health scenarios (Table 5.1) addressed using spatiotemporal eco-epidemiological tools to support the design and implementation of management and policy tools in the State of Minnesota. For each case study, we offer a discussion on the choice of method given the research question, type of data, and the potential decisions stakeholders may draw using the analytical outputs. Each of the case studies involved frequent interactions with relevant stakeholders who would use the risk

maps and epidemiological findings to improve the quality of decisions in prevention and control of these three One Health issues endemic to Minnesota. Therefore, an essential component of this work is the understanding of the “lessons learned” during the researcher-stakeholder interactions related to the three case studies.



**Fig. 5.1.** The use of spatiotemporal analytical tools in the process of risk-based management of One Health scenarios. The steps labelled 1: 5 are discussed in this paper using three case studies representing One Health scenarios.

## 5.3 Data and Methods

### 5.3.1 Description of the three examples/case studies

**Ecosystem health:** A study that involved spatial cluster analysis and cokriging was used to understand reporting bias of aquatic invasive zebra mussels and Eurasian water milfoil spread in Minnesota waters, and to improve the detection and control. ([Kanankege et al., 2018a](#); [Kanankege et al., 2018b](#)).

**Animal health:** A study that involved Getis Ord  $G_i^*$  analysis and Conditional Autoregression modeling was used to understand the extent of Johne's disease distribution in Minnesota dairy cattle and its potential association with underlying risk factors ([Kanankege et al, in review, BMC Vet. Res.](#)).

**Public health:** A study that involved spatiotemporal cluster analysis was used to recognize the existence of "hot spots" for Anthrax in Minnesota ([Kanankege et al, in review-PLOS ONE](#)). The analysis recognizes the historic spread of Anthrax and the risk factors that could lead to the historic occurrence and recurrences while determining the optimal sizes for vaccination/surveillance zones.

**Table 5.1.** Comparison between the three case studies

Case study/ Specific aim	Modeling the spatial dynamics of invasive zebra mussels and Eurasian watermilfoil in Minnesota	Epidemiological characterization of Johne's disease in Minnesota dairy cattle	Spatiotemporal patterns of historic animal Anthrax outbreaks in Minnesota (in comparison to Kazakhstan)
<b>Similarities</b>			
<b>All the subjects included in the case studies:</b>	<ul style="list-style-type: none"> <li>• Are endemic in the state of Minnesota</li> <li>• Cause harm to ecosystem, animal, or public health</li> <li>• Cause substantial economic losses</li> <li>• Require attention to improve mitigation strategies</li> <li>• Are in alignment with One Health objectives</li> <li>• Have no existing scientific method to quantify the risk of spread</li> <li>• Have existing databases collecting incidence data through passive surveillance or voluntary reporting</li> </ul>		
<b>Differences</b>			
<b>Causative agent</b>	Zebra mussels (a bivalve) and Eurasian watermilfoil (an aquatic plant)	<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i> (a bacteria)	<i>Bacillus anthracis</i> (a bacteria)
<b>Host (Relevant field)</b>	Waterbodies (Ecosystem health)	Dairy cattle (Animal health)	Livestock and wildlife (Animal and public health)
<b>Regulations</b>	Reportable	Non-reportable	Reportable
<b>Acute vs. chronic nature of the condition within the host</b>	Chronic (Host: Waterbodies)	Chronic (Host: Dairy cattle)	Acute (Host: Animals and human)
<b>Seasonal/non-seasonal nature of the agent spread</b>	Non-seasonal	Non-seasonal	Seasonal
<b>Data source</b>	Minnesota Department of Natural Resources	Minnesota Dairy Herd Improvement Association	Minnesota Board of Animal Health and <a href="#">Bender et al. 2006</a>
<b>Research question / Specific aim</b>	Introduction of a spatial risk score system	Characterization of the spatial epidemiology	Recognition of "disease hot spots" and suggestion of radii for vaccination/surveillance zones
<b>Specific analytical tools used</b>	Spatial cluster analysis, Network analysis, Co-kriging, and Species distribution modeling using MaxENT algorithm	Getis Ord Gi* analysis and Conditional Autoregression modeling	Spatiotemporal cluster analysis, Spatiotemporal directionality test, and G-statistics of spatial heterogeneity

### 5.3.2 How, when, and by whom the risk maps may be used in One Health scenarios: Discussions with stakeholders related to the three case studies

All the stakeholders were involved in the data collection process and were provided with research progress updated during the period between 2014 through 2018. The discussions that are summarized as “lessons learned” here, were based on three separate in-person meetings, i.e. the research team met with the relevant stakeholders for pre-planned meetings that lasted for 3 to 4 hours. During those meetings the research methods, assumptions, strengths and limitations of the available data and spatiotemporal methodologies used were discussed. Furthermore, the output risk maps were shared with the stakeholders. In alignment with the objective of the work presented here, to evaluate the use of spatiotemporal epidemiological tools in the process of risk-based management of One Health scenarios, the three discussions/meetings between researchers and stakeholders were conducted to identify the following three major topics:

- **Potential improvements** in the research and data collection approaches
- **Opportunities** in using risk estimates and risk maps to inform decision making
- **Challenges** stakeholders may face when changing current management strategies

### 5.3.3 Ecosystem health: Aquatic invasive species in Minnesota

Zebra mussels (ZMs) (*Dreissena polymorpha*) and Eurasian watermilfoil (EWM) (*Myriophyllum spicatum*) are aggressive aquatic invasive species (AIS) affecting Minnesota waterbodies. In addition to the negative impact on the ecosystem, ZM and EWM reduce water flow, clog agricultural, industrial, and drinking water supplies, and therefore, cause considerable economic burden due to the need of preventive, and control measures (Connelly et al., 2007). The early detection of ZM and EWM invasion has been challenging, due in part to the absence of a surveillance program, thus relying on public reporting, and which results in reporting bias (Kanankege et al., 2018a). The existing databases used in the analysis included 1) reported and confirmed invasions during 1987-2015 and 2) boater traffic between waterbodies (during summer 2013) collected from selected lakes as an active effort by Minnesota Department of Natural Resources (MNDNR). To predict the areas at high risk for invasions, while accounting for underreporting, we combined network analysis and probability co-kriging to estimate the risk of ZM and EWM invasions. We used network analysis to generate a waterbody-specific variable

representing boater traffic because human-mediated transportation of invasive species is a known high-risk activity. In addition, co-kriging was used to estimate the probability of species introduction, using waterbody-specific variables, including the variable representing the boater traffic network. Further details about the methods and the results are found elsewhere ([Kanankege et al., 2018b](#)).

The Minnesota Aquatic Invasive Species Research Center (MAISRC) of the University of Minnesota (UMN) has collaborated with MNDNR, County representatives of the MNDNR, and Watershed managers of Minnesota to conduct multiple research activities with the intention of supporting the prevention, timely detection and control of AIS. Therefore, as part of an ongoing research to inform a risk-based management approach to manage AIS, a meeting was held by the research team with the participation of representatives of MAISRC, MNDNR, county representatives and watershed managers. This project represents an initial step of an ongoing risk estimation of aquatic invasions conducted by MAISRC and the ultimate goal of our researcher-stakeholder meeting was to share the research findings and identify opportunities and/or concerns for management implementation that can be useful in the next steps of the project.

### **5.3.3.1 Potential improvements**

Potential improvements to the databases on invasions and boater traffic included the following: a) Planning exploratory random surveys to validate both databases because the survey based boater network is subjected to biases in survey effort and b) Use of correlated datasets to improve the value of boater network data (Examples: use of mobile applications for geofencing, tracking license plates, use of bird's eye cameras, and use of fish registries to recognize highly visited lakes). Measurement of the intensity of invasions was another factor that was pointed out as essential because the severity of invasion may define the potential to act as a source of invasion to the connected waterbodies.

Potential improvements to the analytical approaches were: 1) Making adjustments in the analytical phase to account for the biases; 2) Incorporation of a measurement of cost-effectiveness and comparison among multiple management scenarios into the analysis; 3) Bringing together other epidemiological analytical and modeling tools, such as compartmental modeling and cost-effective analysis, to combine with spatial analysis.

### 5.3.3.2 Opportunities

Recognition of high-risk areas using a scientific/evidence-based approach was viewed as useful to provide rationale for changes in management decisions and strategies. Specifically, the risk mapping process allowed recognition of which area to prioritize when investing on control and preventive measures. It was also essential to determine which preventive/control activity to prioritize (Eeg, controlling incoming or outgoing boater traffic) and the incorporation of other analytical methods was suggested as a solution. Additionally, the importance of quantifying the success of the risk-based new mitigation approach through simulations and modeling was emphasized. Overall, the involvement of scientific research in the management decision process was recognized as an opportunity to innovate discussions on improvements and changes in AIS prevention, management, and control.

### 5.3.3.3 Challenges

The limitations that were recognized by stakeholders included the following: 1) distribution of limited resources among the Minnesota administrative divisions (counties); 2) difficulties in the redistributing resources based on research performed only on certain invasive species; 3) inability to evaluate both inspector and boater compliance to inspection hours and following recommendations for the removal of AIS from contaminated surfaces and fishing gear, and therefore the inability to measure the effectiveness of preventive and control measures; and 4) challenges in engaging law enforcement to emphasize the consequences in AIS control violations. These challenges were also discussed in the light of inability to measure the success of control measures as a tool for “convincing” the public and emphasize the importance of the suggested control activities and inability to answer the question of “who takes the responsibility?” for any changes in the resource allocation.

## 5.3.4 Animal health: Paratuberculosis in cattle

One of the key steps in the management and control of chronic diseases, such as Johne’s disease (JD), caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP), is the ability to track disease incidence over time (Collins et al., 2003). Although subjected to debate within the scientific community, MAP may be involved in the development of human Crohn’s disease. Therefore, the potential for foodborne transmission of MAP would highlight the importance of tracking JD. JD surveillance in U.S. dairy cattle is challenging due to lack of regulatory



requirements, imperfect diagnostic tests, and associated expenses, including time and labor. A proportion of the industry still participates in voluntary JD testing programs, therefore, an alternative approach to assess the distribution and dynamics of JD in the U.S. cattle industry is to use these voluntary testing programs to monitor JD. We conducted a cross-sectional study using data from a voluntary JD testing program conducted by the Minnesota Dairy Herd Improvement Association (MNDHIA) to: a) explore whether such program provides representative information on the prevalence of JD in dairy herds in Minnesota, b) estimate JD distribution in Minnesota, and, c) identify herd and environmental factors associated with finding JD-positive herds. Participant representativeness was assessed by comparing the number of JD-tested herds with the number of herds required to estimate the true disease prevalence per county based on official statistics from the National Agricultural Statistical Services. Multivariable logistic regression models, with and without considering spatial dependence between observations, were then used to investigate the association between herd status to JD (positive/negative), as indicated by milk ELISA results, and available covariates at the herd level. Further details about the methods and the results related to JD analysis are found elsewhere ([Kanankege et al., in review- BMC Vet. Res.](#)).

This applied study exemplified the potential use of MNDHIA's voluntary JD testing databases as a passive surveillance tools to monitor JD status and its association with epidemiological factors while identifying opportunities for improvement of the database. Therefore, once the study was finalized and as part a of an ongoing effort to potentially inform risk-based testing and management of JD in Minnesota, a meeting participating UMN research team and MNDHIA representatives was held with the goal to share project outputs and identify opportunities and/or concerns for management implementation.

#### **5.3.4.1 Potential improvements**

The use of the spatial analysis was considered a potential tool to define “disease-free areas” in the process of JD prevention and control. The discussed improvement of the current MNDHIA database included: 1) Considering the fact that animals may have housed in different locations, collection of the actual geocoordinates of the animal housing in addition to the farm addresses would improve the accuracy of spatial analysis (Farm addresses were used to geocode the locations in the JD study discussed here); and 2) Collection of farm characteristics such as herd size, turnover rates, age structure and management strategies from farms regarding their approach to JD positive cows based on a survey or by observation. Conducting exploratory

random tests in Minnesota dairy farms to validate the test results obtained through the voluntary program was also suggested as an opportunity to improve the existing database.

Improvements suggested by our stakeholders at MNDHIA regarding the analytical process and presentation of the results included: 1) Presentation of the analytical results in an informative manner that could help producer understanding of the impact of the risk factors/covariates; 2) Identifying key epidemiological factors that collectively can be used to inform individual risk-based testing ; and 3) incorporation of other modeling and analytical approaches to evaluate and improve herd level testing and management strategies.

#### **5.3.4.2 Opportunities**

The recognition of risk and disease-free areas was a potential step forward when introducing a producer driven, flexible JD testing index that inform farmers on “which areas, when to test, how many animals to test, and what are the options of decisions towards the JD positive animals, such as quarantine or culling?”. Expansion of the project area to other Midwestern and northeastern states, i.e. New York, Pennsylvania, and Vermont, where there are other National Dairy Herd Improve Associations offering JD testing. This expansion of the area was suggested to facilitate determining similarities and differences by state and improve the generalizability of the research approaches.

Establishing communication portals with the dairy producers was recognized as a key factor to establish a sustainable practice of timely testing and management of JD. The communication portals suggested included adding science pages summarizing the research findings in the MNDHIA annual reports and establishing online portals where dairy farmers could access the testing criterion for animal selection based on their farm location and farm characteristics.

#### **5.3.4.3 Challenges**

The challenges MNDHIA perceived as a stakeholder included: 1) Inability to measure the success of control measures as a tool for “convincing” the herd owners and emphasize the importance of suggested testing and control activities; 2) Farmer’s concern on privacy scenarios regarding JD test results and its use in a commercial platform; and 3) Difficulty of convincing those producers who aren’t testing because, although importance of the JD is recognized, testing is not prioritized legislatively

The challenges that the dairy farming community face were also subjected to the discussion. These challenges can be categorized into two major categories: related to the motive for testing and related to management practices for JD prevention and control. The factors related to the motive for testing included: 1) Limited availability of resources for testing; and 2) Farmers using/interpreting the research findings of disease-free areas and the association with epidemiological factors differently. Challenges related to JD management included: 1) Variability of management practices and attitude towards JD positive cows based on their production capacity (i.e. removal of high producing cows even if they are JD positive would be a challenging management decision); and 2) Large herds have too many animals and therefore they usually have the perception that they have the disease, decreasing their motivation to take measures about it (too costly, no great diagnostic tests, chronic disease etc.).

Reliability and generalizability of the MNDHIA data when representing the JD status of Minnesota was discussed as a challenge because MNDHIA data are based on voluntary testing and therefore potentially biased. Planning exploratory random surveys to validate the JD status derived based on MNDHIA data was suggested as a solution to ensure the generalizability and validity of MNDHIA to represent the JD status of Minnesota. In general, the importance of the willingness to continue JD testing, funding, surveillance, and research on the disease by both animal health authorities and the livestock industry was recognized as a key factor in planning a JD surveillance and control program.

### **5.3.5 Animal and public health: Anthrax in an endemic areas**

Anthrax is a zoonotic disease affecting livestock, wildlife, and humans caused by the spore-forming bacterium *Bacillus anthracis* (Sternbach, 2003). Our goal was to compare the spatiotemporal patterns in Anthrax progression, intensity, direction, and recurrence (disease hot spots), in relation to epidemiological factors and potential trigger events, in two endemic areas, located at northern latitudes in the western and eastern hemispheres. Distribution of the spatiotemporal cluster radii between study sites supported suggestion of site-specific control zones. Spatiotemporal patterns of Anthrax occurrence in both endemic regions were attributed to multiple potential trigger events including major river floods, changes in land use, agriculture, and susceptible livestock populations. Results helped to understand the long-term epidemiological dynamics of Anthrax while providing suggestions for the design and implementation of prevention and control programs in endemic settings. Further details about the

methods and the results related to spatiotemporal analysis of historic Anthrax is found elsewhere ([Kanankege et al., in review- PLOS ONE](#)).

To share the research findings and potentially inform a risk-based surveillance and management of Anthrax in Minnesota, a meeting was held by the UMN research team with the participation of representatives of the Minnesota Board of Animal Health (MNBOAH) with the goal to review project results and identify opportunities and/or concerns for management implementation.

#### **5.3.5.1 Potential improvements**

The recent improvements incorporated to the MNBOAH database such as real-time collection of disease data including diagnostics, number and type of animals affected, herd characteristics was expected to facilitate a robust and efficient data collection and storage process. Additionally, through the meeting, the importance of the improvement of the data entry protocols and precision on the exact location information where the infected animals or dead carcasses were found in the field, rather than farm addresses, were recognized as key components that would facilitate spatiotemporal analysis of Anthrax. Involvement of field staff and veterinarians was discussed as a solution to this data discrepancy. Furthermore, conducting a surveillance program or data collection process to gather vaccination information was recognized as a complementary approach because vaccination of the susceptible animals was a non-compulsory, a producer driven decision, and also because there were a few historical vaccine related Anthrax cases in Minnesota, during the 1930's ([Personal communication: Bender, 2018](#)).

The changes discussed regarding the spatiotemporal cluster analysis included the definition of scan window sizes and how the space-time window parameters may change by each endemic site. Therefore, the importance of exercising different cluster definition criteria based on stakeholder, researcher, and expert opinion was discussed.

#### **5.3.5.2 Opportunities**

Current recommendations for Anthrax in the two endemic sites were different. Kazakhstan recommended mass vaccination and our case study suggested that the size of Kazakhstan's natural anthrax outbreaks may range between 18 and 310 km, with a median of 67 km ([Kanankege et al., in review-PLOS ONE](#)). Whereas, in Minnesota the vaccination policy and recommendations included: spring vaccination of northwestern Minnesota considering the recent outbreak in 2000's and 2005 were in that region, and vaccination of susceptible animals with in

16 km (10 miles) from the premise where there is at least one reported case. Furthermore, the Unified Anthrax Recommendations (Stoltenow, 2010) on vaccination of herds with a history of Anthrax within the last 10 years and vaccination of herds that are within 10 km from those herds was recommended. Our case study showed that the range of cluster radii where Anthrax outbreaks were detected was between 2 and 40 km with a median of 12 km, and an average of 14 km. These radii were the recommendations that were suggested when defining surveillance or vaccination zones. The average and median radii were within the range of the existing policy and recommendations that were based on expert opinion.

### **5.3.5.3 Challenges**

The challenges MNBOAH may encounter in improving the reporting and promoting vaccination or antibiotic treatments included: vaccination being a producer driven decision, lack of indemnity even during an outbreak, low priority given to vaccination due to the absence of recent outbreaks, and potential of missing cases of wild animal-anthrax cases, which are not gathered by MNBOAH which may have led to delayed detection of potential livestock outbreaks.

### **5.3.6 Summary of case studies**

The case studies were evaluated in relation to the existence of steps enabling risk-based management of the three scenarios (Table 5.2). The openness or willingness to implement the scientifically-based risk scores to inform decisions is a key component in risk-based management of One Health scenarios. The other factors were related to the existence of an appropriate database, the technical capacity, the availability of the resources for the surveillance, management, and sustainability of the mitigation process. Furthermore, in relation to the spatiotemporal analysis, the existence of management decisions that can be drawn using spatial analysis was more prominent for aquatic invasive species and Anthrax case studies compared to the JD study.

**Table 5.2.** Summary of the assessment of researcher-stakeholder interactions related to Ecosystem health (Aquatic invasive species, AIS), animal health (Johne's disease), and public health (Anthrax) case studies: Lessons learned.

<b>Steps enabling the risk-based management</b>	<b>Aquatic invasive species</b>	<b>Johne's Disease</b>	<b>Anthrax</b>
1. Openness to the scientifically-based risk scores to inform decisions			
2. Existence of preventive decisions that can be drawn based on spatial analysis			
3. Existence of a suitable database and capability of spatial data management			
4. Technical capacity to incorporate spatiotemporal analytical outputs into regular reports			
5. Availability of resources for the surveillance, management, and sustainability			
6. Existence of policy related to prevention, control, and eradication			

Existing
  In progress
  Not existing

## 5.4 Discussion and conclusions

### 5.4.1 Lessons learned from the three One Health studies: suggestions for “good practices”

The major lessons learned through examining the three case studies, which can be suggestions for “good practices” in the future research activities, include: 1) identification of the opportunities to improve data quality and sustainability through preliminary analyses of existing data; 2) involvement of the stakeholders from the planning stage of research activities to better serve the requirements of the primary and secondary stakeholders; 3) understanding the biases associated with data and analysis, thus their influence on the outputs; 4) identification of risk thresholds through communication to differentiate between data driven thresholds and expert/stakeholder opinion based risk thresholds; 5) translation of those analytical and modeling outputs/information into a “decision-making oriented” formats, such as disease testing indexes or guidelines for disease surveillance activities; and 6) to be able to influence the decisions on resource allocation, conducting of cost-effectiveness analysis was recognized as an essential component of the analysis to “convince” the relevant stakeholders involved in the decision-making process.

Efficacy, effectiveness, and efficiency are the three terms commonly used when evaluating the success of health interventions ([Gordis, 2009](#)). A similar pattern was observed during all three discussions with the stakeholders during which, proof of risk-mapping based prioritization of the resources was emphasized as a key to “convince” the funding sources to change the existing policy and decisions. Model validation techniques and predictive powers of the models were discussed as potential measures of efficacy and effectiveness of the use of geospatial model outputs, i.e. risk maps. However, the subjective nature of the acceptable sensitivity and specificity of a spatiotemporal modeling approach was recognized as a key limiting factor which leads to restrict the use of these modeling outputs in the decision making process. Smoothing techniques and spatial model outcomes often result in a trade-off between sensitivity and specificity. This is because, with higher sensitivity true high-risk areas are more easily and correctly identified, whereas, with a higher specificity areas without high-risk are the ones more prone to be correctly identified ([Elliot and Wartenberg, 2004](#)). The decision of the threshold level of risk was a subject of discussion where the expert opinion from both researchers and stakeholders played a role in the decision-making.

The backlash against evidence-based science is in part due to the challenge of scientific communication where scientific knowledge is not translated into stakeholders and policy-makers in an understandable and compelling fashion, i.e. science communication is often referred to as “too technical, jargon and science”. Science communication to stakeholders and the public is extremely important and this highlights scientists to become equipped with outstanding storytelling skills for sustainable researcher-stakeholder relationship.

The power of juxtaposition, i.e. side by side comparisons, was considered as useful in the process when “convincing” stakeholders, decision-makers, and funding sources to use scientifically-based approaches over the traditional methods of decision making. Moreover, the use of technical mediums such as audio-visual and use of platforms such as internet blogging, YouTube, or social media for mass communication were seen as powerful resources. The impact of this type of mass communication may last for extended periods compared to traditional researcher-stakeholder communication platforms such as workshops or conferences.

#### **5.4.2 The potential weaknesses of spatiotemporal analytical tools**

Common weaknesses associated with the spatial analysis and risk mapping are related to shortcoming in the accuracy and mapping of data in the geographical space, choice of the analytical/ modeling tools and relevant assumptions, and the decisions related to the representation of the risk maps to the end users (Ocaña-Riola, 2010; Loth *et al.* 2011; ). When spatial analytics and models are conducted based on available and potentially biased data, the resulting risk maps are invariably subjected to the negative impact of the data quality. However, we emphasize the use of existing data, bringing several databases together, and initiating the process of improving data quality can be benefitted by the spatiotemporal eco-epidemiological tools.

The choice of analytical/modeling tool used depends on multiple factors: 1) characteristics of the disease/adverse event; 2) the study design; 3) the spatial explicitness of data; 4) Data quality and availability; 5) research question and hypothesis; 6) stakeholder involvement; and 7) the existence of regulations, resources, and policy related to the adverse event (Thesis Chapter1: Kanankege, 2019). All analytics and models involve certain assumptions on statistical properties of variables and often these assumptions are violated in natural environments. For example, spatial continuity of risk is a common assumption in risk-mapping process while there can be natural (Ex. Mountain range acting as a physical barrier) or infrastructural barriers (Ex.



Urban vs. rural neighborhoods) that violate the continuity assumption resulting in step changes of risk between adjacent areas (Rushworth et al, 2017). Therefore, clarity on the choice of analytical or modeling tool, underlying assumptions is an essential component when using spatiotemporal eco-epidemiological tools in the One Health process.

The representations risk associations and risk maps, both qualitatively and quantitatively, are often based on arbitrary decisions of risk thresholds or cut-off values that should be based upon agreements between researchers and stakeholder. Therefore, having standards for these risk thresholds becomes a challenging task. Moreover, when comparing risk maps on infectious diseases with multiple hosts or vectors, it is likely that there may be disagreements (Loth et al. 2011). The differences between risk maps may largely be attributed to the distribution of the population at risk, influence of changing environmental factors, and the data collection discrepancies (Loth et al. 2011).

There are several hierarchies of stakeholders and the primary stakeholders chosen here were the data holders who were involved in the data collection and maintenance. In both AIS and Anthrax studies, the stakeholders were also the relevant authorities for state level decisions on preventive and control of the events. State governmental agency was not chosen in the JD study because JD was not regulated and testing for JD was not required. Admittedly, it is ideal to bring all the relevant stakeholders into these conversations and understand their contributions in the process, and how the outcomes of the analyses such as risk maps would support their decisions and recognize the opportunities for improvement.

### 5.4.3 Conclusions

The knowledge generated through well-designed, effectively executed research is critical to our future capacity building. The overarching goals of preparedness and response in One Health scenarios include: preventing, controlling, and eradication of environmental, animal, and human health threats. Spatial analysis supports understanding and modeling of the spread of ecosystem, animal and human health threats, which are essential component in risk-based surveillance and management. The risk-based paradigm involves evaluating whether the relative efforts of regulation, allocation of resources for monitoring and mitigation devoted to reducing risks are in reasonable proportion to the seriousness of the risk being compared.

The relative significance of the different components of complex One Health problems needs to be understood to help further refine policy and intervention measures to target the right populations and spread pathways. Compared to the traditional surveillance programs, risk-based

surveillance and management programs, which uses spatial analytical outcomes, may provide a focused index to follow when utilizing finite monetary resources.

## 5.5 Acknowledgements

This study was funded in part by the Minnesota Discovery, Research, and Innovation Economy (MnDRIVE) program and Office of the Vice President for Research (OVPR) of the University of Minnesota. We acknowledge Dr. Meggan Craft, Dr. Eva Enns, and members of the Minnesota Department of Natural Resources for the participation at the meeting/workshop. We also extend our acknowledgement to Bruce Dokkebakken, Val Schumann, and the members of Minnesota Dairy Herd Improvement Association, as well as Dr. Linda Glaser, Dr. Beth Thompson and Dr. Stacey Schwabenlander of Minnesota Board of Animal Health for their kind participation in the meetings and sharing their expert knowledge.

## 5.6 References

1. Asokan GV, Asokan V. 2015. Leveraging "big data" to enhance the effectiveness of "one health" in an era of health informatics. *Journal of Epidemiology and Global Health*. 5(4):311-4. doi: 10.1016/j.jegh.2015.02.001. PubMed PMID: WOS:000214972400002.
2. Banerjee S, Carlin BP, Gelfand AE. 2014. *Hierarchical Modeling and Analysis for Spatial Data*, Second Edition, Monographs on Statistics and Applied Probability (2nd ed.), Chapman and Hall/CRC, ISBN 9781439819173.
3. Beale L, Abellan JJ, Hodgson S, Jarup L. 2008. Methodologic issues and approaches to spatial epidemiology. *Environmental Health Perspectives*. Aug; 116:1105-1110.
4. Cartwright SJ, Bowgen KM, Collop C, Hyder K, Nabe-Nielsen J, Stafford R, Stillman RA, Thorpe RB, Sibly RM. 2016. Communicating complex ecological models to non-scientist end users. *Ecological Modelling*. Oct;338:51-59.
5. Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. 2014. Visualization and analytics tools for infectious disease epidemiology: A systematic review. *Journal of Biomedical Informatics*. Oct;51:287-298.

6. Clements ACA, Pfeiffer DU. 2009. Emerging viral zoonoses: Frameworks for spatial and spatiotemporal risk assessment and resource planning. *Veterinary Journal*, 182(1), 21-30. doi:10.1016/j.tvjl.2008.05.010.
7. Collins MT. 2003. Update on paratuberculosis: 1. Epidemiology of Johne's disease and the biology of *Mycobacterium paratuberculosis*. *Ir. Vet. J.* 56(11): 565-574.
8. Connelly NA, O'Neill CR, Knuth BA, Brown TL. 2007. Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. *Environ. Management.* 40, 105-112.
9. CRARM: The Presidential/ Congressional Commission on Risk Assessment and Risk Management. 1997. Framework for Environmental Health Risk Management. Commission on Risk Assessment and Management, Washington DC. Vol.
10. Cromley EK, McLafferty SL. 2002. GIS and Public Health. The Guilford Publications Inc., New York. ISBN: 1-57230-707-2.
11. Cunningham AA, Daszak P, Wood JLN. 2017. One Health, emerging infectious diseases and wildlife: two decades of progress? *Philos Trans R Soc Lond B Biol Sci.* 2017 Jul 19; 372(1725): 20160167. doi: 10.1098/rstb.2016.0167.
12. Dufour B, Plee L, Moutou F, Boisseleau D, Chartier C, Durand B, et al. 2011. A qualitative risk assessment methodology for scientific expert panels. *Revue Scientifique Et Technique-Office International Des Epizooties*; 30: 673-681.
13. Elliott P, Wartenberg D. 2004. Spatial epidemiology: Current approaches and future challenges. *Environmental Health Perspectives* 112:998-1006.
14. Goodman MS, Thompson VLS. 2017. The science of stakeholder engagement in research: classification, implementation, and evaluation. *Translational Behavioral Medicine.* Sep;7:486-491.
15. Kanankege K, Machado G, Zhang L, Dokkebakken B, Schumann V, Wells S, Perez A, Alvarez J. Use of a voluntary testing program to study the spatial epidemiology of Johne's

- disease in Minnesota: A cross-sectional study. (In Review-BMC Veterinary Research: 2018).
16. Kanankege KST, Abdrakhmanov SK, Alvarez J, Glaser L, Bender JB, Mukhanbetkaliyev YY, Korennoy FI, Kadyrov AS, Abdrakhmanova AS, Perez AM. Comparison of spatiotemporal patterns of historic animal Anthrax outbreaks in Minnesota and Kazakhstan. (In Review-PLOS One: 2018).
  17. Kanankege KST, Alkhamis MA, Perez AM, Phelps NBD. 2018a. Zebra mussels and Eurasian watermilfoil reporting patterns in Minnesota. *Journal of Great Lakes Research*. doi: 10.1016/j.jglr.2018.02.008.
  18. Kanankege KST, Alkhamis MA, Phelps NBD, Perez AM. 2018b. A probability co-kriging model to account for reporting bias and recognize areas at high risk for zebra mussels and Eurasian watermilfoil invasions in Minnesota. *Frontiers in Veterinary Science*. doi: 10.3389/fvets.2017.00231.
  19. Kirby RS, Delmelle E, Eberth JM. 2017. Advances in spatial epidemiology and geographic information systems. *Annals of Epidemiology*. Jan;27:1-9.
  20. Lawson AB, Bohning D, Biggeri A, et al. eds. 1999. *Disease mapping and its uses. disease mapping and risk assessment for public health*. Chichester: Wiley.
  21. Loth L, Gilbert M, Wu G, Czarnecki C, Hidayat M, Xiao X. 2011. Identifying risk factors of highly pathogenic avian influenza (H5N1 subtype) in Indonesia. *Preventive Veterinary Medicine*. Volume 102, Issue 1, 1 October 2011, Pages 50–58.
  22. March D, Susser E. 2006. The eco- in eco-epidemiology. *International Journal of Epidemiology*. Dec;35:1379-1383.
  23. Moore DA, Carpenter TE. 1999. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev*. 21(2):143-6.
  24. Morse SS. 1993. *Emerging viruses*. New York, NY: Oxford University Press.

25. Murtaugh MP, Steer CJ, Sreevatsan S, Patterson N, Kennedy S, Sriramarao P. 2017. The science behind One Health: at the interface of humans, animals, and the environment. *Annals of the New York Academy of Sciences*. May;1395:12-32.
26. Office International Des Epizootis (OIE). *Terrestrial Animal Health Code*. 2018. Chapter 11.4. Bovine Spongiform Encephalopathy. Article 11.4.1. Available online at: [http://www.oie.int/fileadmin/Home/eng/Health\\_standards/tahc/current/chapitre\\_bse.pdf](http://www.oie.int/fileadmin/Home/eng/Health_standards/tahc/current/chapitre_bse.pdf). Cited: 12.24.2018.
27. Perez A. 2015. Past, Present, and Future of Veterinary Epidemiology and Economics: One Health, Many Challenges, No Silver Bullets. *Specialty Grand Challenge*. *Front Vet Sci*. 2: 60. doi: 10.3389/fvets.2015.00060.
28. Personal communications: Bender JB. 2018. Vaccine originated cases during late 1990's in few farms in Minnesota.
29. Pfeiffer DU, Stevens KB. 2015. Spatial and temporal epidemiological analysis in the Big Data era. *Preventive Veterinary Medicine*, 122(1-2), 213-220. doi:10.1016/j.prevetmed.2015.05.012.
30. Rabinowitz PM, Kock R, Kachani M, Kunkel R, Thomas J, Gilbert J, et al. 2013. Toward proof of concept of a one health approach to disease prediction and control *Emerging Infectious Diseases*, 19(12).
31. Rezaeian M, Dunn G, St Leger S, Appleby L. 2007. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *Journal of Epidemiology and Community Health*. 61(2):98-102. doi: 10.1136/jech.2005.043117. PubMed PMID: WOS:000243597800005.
32. Ricardo Ocaña-Riola. 2010. Common errors in disease mapping. *Geospatial Health* 4(2): 139-154.
33. Rushworth A, Lee D, Sarran C. 2017. An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk. *Journal of the Royal Statistical Society Series C-Applied Statistics*. 66(1):141-57. doi: 10.1111/rssc.12155. PubMed PMID: WOS:000392808300002.

34. Slunge D, Drakenberg O, Ekbohm A, Göthberg M, Knaggård A, Sahlin U. 2017. Stakeholder interaction in research process - A guide for researcher and research groups. Available online at: [https://gmv.gu.se/digitalAssets/1619/1619929\\_stakeholder-interaction-in-research-processes---guide---final-march-15-2017.pdf](https://gmv.gu.se/digitalAssets/1619/1619929_stakeholder-interaction-in-research-processes---guide---final-march-15-2017.pdf).
35. Shanafelt DW, Perrings C. 2018. The Effect of the Post 2001 Reforms on FMD Risks of the International Live Animal Trade. *Ecohealth*. Jun;15:327-337.
36. Sternbach G. 2003. The history of anthrax. *Journal of Emergency Medicine*. 24(4):463-7. doi: 10.1016/s0736-4679(03)00079-9. PubMed PMID: WOS:000182902600019.
37. Stevens KB, Pfeiffer DU. 2016. The Role of Spatial Analysis in Risk-Based Animal Disease Management. *Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
38. Stoltenow CL. 2012. Unified Anthrax Recommendations for MB, MN, MT, ND, and SD. Available online: [https://www.nd.gov/ndda/sites/default/files/legacy/resource/UnifiedAnthraxRecommendations\\_0.pdf](https://www.nd.gov/ndda/sites/default/files/legacy/resource/UnifiedAnthraxRecommendations_0.pdf).
39. Susser E. 2004. Eco-epidemiology: Thinking outside the black box. *Epidemiology*. Sep;15:519-520.
40. VanderWaal K, Morrison RB, Neuhauser C, Vilalta C, Perez AM. 2017. Translating big data into smart data for veterinary epidemiology. *Front Vet Sci*. 4:110. doi: 10.3389/fvets.2017.00110.
41. Wagner Robb S, Bauer SE, Vena JE. 2016. Integration of Different Epidemiologic Perspectives and Applications to Spatial Epidemiology. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.

42. Wagnet RJ, Huston JL. 1996. Scale-dependency of solute transport modeling/GIS applications. *Journal of Environmental Quality*. 25 (3): 499-510.
43. Ward MP, Carpenter TE. 2000. Analysis of time-space clustering in veterinary epidemiology. *Prev Vet Med*. 43:225–37.10.1016/S0167-5877(99)00111-7.

## Comprehensive reference list

1. Abdou M, Hamill L, Gilbert N. 2012. Designing and Building an Agent-Based Model. Eds. Heppenstall AJ, Crooks AT, See LM, Batty M. Springer Dordrecht Heidelberg, NY. DOI 10.1007/978-90-481-8927-4.
2. Abdrakhmanov SK, Mukhanbetkaliyev YY, Korennoy FI, Karatayev BS, Mukhanbetkaliyeva AA, Abdrakhmanova AS. 2017b. Spatio-temporal analysis and visualisation of the anthrax epidemicsituation in livestock in Kazakhstan over the period 1933-2016. *Geospatial Health*. (12):2. doi: 10.4081/gh.2017.589.
3. Abdrakhmanov SK, Mukhanbetkaliyev YY, Korennoy FI, Sultanov AA, Kadyrov AS, Kushubaev DB, et al. 2017a. Maximum entropy modeling risk of anthrax in the Republic of Kazakhstan. *Preventive Veterinary Medicine*. 144:149-57. doi: 10.1016/j.prevetmed.2017.06.003. PubMed PMID: WOS:000407406600018.
4. Adamovich VL and Nikonov NN. 1970. The importance of landscape-ecological factors in the epizootology of anthrax. *Message 2: Comparative method of assessment of epizootic tension of aterritory. The journal of Microbiology*. 8:113 – 7.(in Russian)
5. Agterberg FP. 1984. Trend surface analysis. *Spatial statistics and models*, Gaile, G. L. and C. J. Willmott, eds., Reidel, Dordrecht, Holland, 147–171.
6. Aikembayev AM, Lukhnova L, Temiraliyeva G, Meka-Mechenko T, Pazylov Y, Zakaryan S, et al. 2010. Historical Distribution and Molecular Diversity of *Bacillus anthracis*, Kazakhstan. *Emerging Infectious Diseases*. 16(5):789-96. doi: 10.3201/eid1605.091427. PubMed PMID: WOS:000277209900006.
7. Aikio S, Duncan RP, Hulme PE. 2010. Herbarium records identify the role of long-distance spread in the spatial distribution of alien plants in New Zealand. *J. Biogeogr*. 37, 1740-1751.
8. Akaike, H. 1977. On entropy maximization principle. *Applications of Statistics*, P. R. Krishnaiah, ed., North-Holland Publishing Co., Amsterdam, Holland, 27-41.
9. Alkhamis M, Moore BR, Perez AM. 2015. Phylodynamics of H5N1 highly pathogenic avian influenza in Europe, 2005-2010.: potential for molecular surveillance of new outbreaks. *Viruses*. 7: 3310-28. SoI: 10.3390/v7062773



10. Alkhamis MA, Hijmans RJ, Al-Enezi A, Martínez-López B, Perez AM. 2016. The use of spatial and spatiotemporal modeling for surveillance of H5N1 highly pathogenic avian influenza in poultry in the middle east. *Bio One. Avian diseases.* 60:146-155.
11. Alkhamis MA, VanderWaal KL. 2016. Spatial and Temporal Epidemiology of Lumpy Skin Disease in the Middle East, 2012-2015. *Frontiers in Veterinary Science*; 3:19.
12. Anonymous. 1980. Lac Qui Parle flood control project, Master plan for public use, Development and resource management. U.S. Army Corps Engineers, St Paul district, Minnesota. AD-A120 745.
13. Anselin L. 1995. Local indicators of spatial association – LISA. *Geographical Analysis.* 27(2):93-115. doi: 10.1111/j.1538-4632.1995.tb00338.x. PubMed PMID: WOS:A1995QU19800001.
14. Anselin L. 1996. The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In *Spatial Analytical Perspective on GIS*, 111–25, edited by M. Fischer, H. Scholten, and D. Unwin. London: Taylor & Francis.
15. Arbia G. 2006, *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*, Springer-Verlag, Berlin.
16. Arkhipkin OP, Spivak LF, Sagatdinova GN. 2010. Development of Flood Space Monitoring in Kazakhstan. *Geoscience and Remote Sensing, New Achievements.* 419-36. PubMed PMID: WOS:000363695000023.
17. Asokan GV, Asokan V. 2015. Leveraging "big data" to enhance the effectiveness of "one health" in an era of health informatics. *Journal of Epidemiology and Global Health.* 5(4):311-4. doi: 10.1016/j.jegh.2015.02.001. PubMed PMID: WOS:000214972400002.
18. Auchincloss AH, Gebreab SY, Mair C, Roux AVD. 2012. A Review of Spatial Methods in Epidemiology, 2000-2010. *Annual Review of Public Health, Vol 33.* 33:107-+. doi: 10.1146/annurev-publhealth-031811-124655. PubMed PMID: WOS:000304202700008.
19. Baldwin RA. 2009. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy.* Dec;11:854-866.
20. Banerjee S, Carlin BP, Gelfand AE. 2014. *Hierarchical Modeling and Analysis for Spatial Data, Second Edition, Monographs on Statistics and Applied Probability (2nd ed.)*, Chapman and Hall/CRC, ISBN 9781439819173.
21. Banerjee S. 2016. Spatial survival models. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology.* Eds: Lawson AB, Banerjee S, Haining RP, Ugarte

- MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
22. Banha F, Gimeno I, Lanao M, Touya V, Duran C, Peribanez MA, et al. 2016. The role of waterfowl and fishing gear on zebra mussel larvae dispersal. *Biol. Invasions*. 18, 115-125.
  23. Barnes MA, Jerde CL, Wittmann M., Chadderton WL, Ding JQ, Zhang JL, Purcell M, Budhathoki M, Lodge DM. 2014. Geographic selection bias of occurrence data influences transferability of invasive *Hydrilla verticillata* distribution models. *Ecol. Evol.* 4, 2584-2593.
  24. Barrios JM, Verstraeten WW, Maes P, Aerts JM, Farifteh J, Coppin P. 2012. Using the Gravity Model to Estimate the Spatial Spread of Vector-Borne Diseases. *Int. J. Environ. Res. Public Health*. 9, 4346-4364.
  25. Batty M. 1997. Cellular Automata and Urban Form: A Primer. *Journal of the American Planning Association*, 63, 266-274.
  26. Beale L, Abellan JJ, Hodgson S, Jarup L. 2008. Methodologic issues and approaches to spatial epidemiology. *Environmental Health Perspectives*. Aug; 116:1105-1110.
  27. Bellan SE, Gimenez O, Choquet R, Getz WM. 2013. A hierarchical distance sampling approach to estimating mortality rates from opportunistic carcass surveillance data. *Methods in Ecology and Evolution*. 4(4):361-9. doi: 10.1111/2041-210x.12021. PubMed PMID: WOS:000317021300007.
  28. Bender J, Cutting K, Miller B, Friendshuh K, Bolding T, Nelson J, Danila R. 2006. Historical Anthrax Among Animals and Humans in Minnesota, 1912 – 2005. International Conference on Emerging Infectious Diseases, abstr p.156. Atlanta, GA March 19-21.
  29. Benhamou S. 2011. Dynamic approach to space and habitat use based on biased random bridges. *PLoS ONE* 6(1): e14592. Doi:10.1371/journal.pone.0014592
  30. Benjamin LA, Fosgate GT, Ward MP, Roussel AJ, Feagin RA, Schwartz AL. 2010. Attitudes towards biosecurity practices relevant to Johne's disease control on beef cattle farms. *Prev. Vet. Med.* 94(3-4): 222-230.
  31. Besag J, York J and Mollié A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*. 43(1): 1–20.
  32. Besag J. 1974. Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion), *J. Royal Stat. Soc. Series B*. 36(2): 192–236.

33. Besag, J, Newell, J. 1991. The detection of clusters in rare diseases. *Jornal of the Royal Statistic Society A*, (154), 154, 143–55.
34. Bihrmann K, Nielsen SS, Toft N, Ersboll AK. 2012. Spatial differences in occurrence of paratuberculosis in Danish dairy herds and in control programme participation. *Prev. Vet. Med.*103(2-3): 112-119.
35. Biondi F. 2013. Space-time kriging extension of precipitation variability at 12 km spacing from tree-ring chronologies and its implications for drought analysis. *Hydrology and Earth System Sciences Discussussions*, 10:4301–4335, 2013. URL: <http://dx.doi.org/10.5194/hessd-10-4301-2013>.
36. Birch CPD, Oom SP, Beecham JA. 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*. 206, 347-359.
37. Bivand R, Piras G. 2015. Comparing Implementations of Estimation Methods for Spatial Econometrics. *J. Stat. Softw.* 63: 1-36.
38. Bivand RS, Hauke J, Kossowski T. 2013. Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geogr. Anal.* 45:150-179.
39. Blackburn JK, McNyset KM, Curtis A, Hugh-Jones ME. 2007. Modeling the geographic distribution of *Bacillus anthracis*, the causative agent of anthrax disease, for the contiguous United States using predictive ecologic niche modeling. *American Journal of Tropical Medicine and Hygiene*. 77(6):1103-10. PubMed PMID: WOS:000252123200022.
40. Bobeldyk AM, Bossenbroek JM, Evans-White MA, Lodge DM, Lamberti GA. 2005. Secondary spread of zebra mussels (*Dreissena polymorpha*) in coupled lake-stream systems. *Ecosci.* 12, 339-346.
41. Bosch I, Makarewicz JC, Bonk EA, Ruiz C, Valentino M. 2000. Responses of lake macrophyte beds dominated by Eurasian watermilfoil (*Myriophyllum spicatum*) to best management practices in agricultural sub-watersheds: Declines in biomass but not species dominance. *J Great Lakes Research*; 35: 99-108.
42. Bossenbroek JM, Johnson LE, Peters B, Lodge DM. 2007. Forecasting the expansion of zebra mussels in the United States. *Conservation Biology*. Jun;21:800-810.
43. Bourn D. 2013. Draft guidelines on the enumeration of nomadic and semi-nomadic (transhumant) livestock.

44. Breiman L. 2001. Random forests. *Machine Learning* 45(1):5–32. DOI 10.1023/A:1010933404324
45. Bright EA, Rose AN, Urban ML. 2013. LandScan 2012™. High Resolution Global Population Data. UT-Battelle, LLC. Oak Ridge National Laboratory. Contract No. DE-AC05-00OR22725. U.S. Department of Energy. <http://www.ornl.gov/landscan/>.
46. Brown JE, Stepien CA. 2010. Population genetic history of the dreissenid mussel invasions: expansion patterns across North America. *Biol. Invasions*. 12, 3687-3710.
47. Brunson C, Fotheringham AS, Charlton ME. 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*. 28(4):281-98. doi: 10.1111/j.1538-4632.1996.tb00936.x. PubMed PMID: WOS:A1996VL03500001.
48. Buchan LAJ, Padilla DK. 1999. Estimating the probability of long-distance overland dispersal of invading aquatic species. *Ecological Applications*. Feb;9:254-265.
49. Buchan LAJ, Padilla DK. 2000. Predicting the likelihood of Eurasian watermilfoil presence in lakes, a macrophyte monitoring tool. – *Ecol. Appl.* 10, 1442-1455.
50. Buckland, S.T. & Elston, D.A. (1993) Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology* 30: 478–95.
51. Cai B, Lawson AB, Hossain MM, Choi J. 2012. Bayesian latent structure models with space-time dependent covariates. *Statistical Modelling* 12(2), 145–164.
52. Capen DE, Fenwick JW, Inkley DB, Boynton AC. 1986. Multivariate models of songbird habitat in New England forests. In: *Wildlife 2000: Modelling Habitat Relationships of Terrestrial Vertebrates*, ed. J.A. Verner, M.L. Morrison and C.J. Ralph, pp. 171–75. Madison, WI, USA: University of Wisconsin Press.
53. Carlton JT. 2008. The zebra mussel *Dreissena polymorpha* found in North America in 1986 and 1987. *J. Great Lakes Res.* 34, 770–773.
54. Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. 2014. Visualization and analytics tools for infectious disease epidemiology: A systematic review. *Journal of Biomedical Informatics*. Oct;51:287-298.
55. Carter MA. 2012. Prevalence and Prevention of Paratuberculosis in North America. *Jpn. J. Vet. Res.* 60: S9-S18.
56. Cartwright SJ, Bowgen KM, Collop C, Hyder K, Nabe-Nielsen J, Stafford R, Stillman RA, Thorpe RB, Sibly RM. 2016. Communicating complex ecological models to non-scientist end users. *Ecological Modelling*. Oct;338:51-59.

57. Chase ME, Bailey RC. 1999. The ecology of the zebra mussel (*Dreissena polymorpha*) in the lower Great Lakes of North America: I. Population dynamics and growth. *Journal of Great Lakes Research*; 25: 107-121.
58. Chen X, Wang K. 2017. Geographic area-based rate as a novel indicator to enhance research and precision intervention for more effective HIV/AIDS control. *Preventive Medicine Reports*. 5:301–7.
59. Cheruvilil KS, Soranno PA, Madsen JD, Roberson MJ. 2002. Plant architecture and epiphytic macroinvertebrate communities: the role of an exotic dissected macrophyte. *J. N. Amer. Benthol. Soc.* 21, 261-277.
60. Claudi R, Mackie, GL. 1993. *Practical manual for zebra mussel monitoring and control*. CRC Press Inc. Florida. ISBN 0-8771-985-9.
61. Clayton D, Kaldor J. 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. 43:671–681. doi: 10.2307/2532003.
62. Clements ACA, Pfeiffer DU. 2009. Emerging viral zoonoses: Frameworks for spatial and spatiotemporal risk assessment and resource planning. *Veterinary Journal*, 182(1), 21-30. doi:10.1016/j.tvjl.2008.05.010
63. Collins MT, Eggleston V, Manning EJB. 2010. Successful control of Johne's disease in nine dairy herds: Results of a six-year field trial. *J. Dairy Sci.* 93(4):1638-1643.
64. Collins MT, Sockett DC, Goodger WJ, Conrad TA, Thomas CB, Carr DJ. 1994. Herd prevalence and geographic distribution of, and risk factors for, bovine paratuberculosis in Wisconsin. *J. Am. Vet. Med. Assoc.* 204(4): 636-641.
65. Collins MT. 2003. Update on paratuberculosis: 1. Epidemiology of Johne's disease and the biology of *Mycobacterium paratuberculosis*. *Ir. Vet. J.* 56(11): 565-574.
66. Connelly NA, O'Neill CR, Knuth BA, Brown TL. 2007. Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. *Environ. Management*. 40, 105-112.
67. CRARM: The Presidential/ Congressional Commission on Risk Assessment and Risk Management. 1997. *Framework for Environmental Health Risk Management*. Commission on Risk Assessment and Management, Washington DC. Vol. 1.
68. Cressie N, Wikle CK. 2011. *Statistics for spatio-temporal data*. Wiley Series in Probability and Statistics. ISBN: 978-0-471-69274-4
69. Cressie NAC. 1993. *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

70. Crickman CW. 1934. Beef cattle production in Minnesota [Internet]. Minnesota Bulletin: University of Minnesota Agricultural Experiment station and United States Department of Agriculture Bureau of Agriculture Economics. [8-20]
71. Cromley EK, McLafferty SL. 2002. GIS and Public Health. The Guilford Publications Inc., New York. ISBN: 1-57230-707-2.
72. Crooks AT, Heppenstall AJ. 2012. Agent-based models of geographical systems. Eds. Heppenstall AJ, Crooks AT, See LM, Batty M. Springer Dordrecht Heidelberg, NY. DOI 10.1007/978-90-481-8927-4.
73. Cunningham AA, Daszak P, Wood JLN. 2017. One Health, emerging infectious diseases and wildlife: two decades of progress? *Philos Trans R Soc Lond B Biol Sci.* 2017 Jul 19; 372(1725): 20160167. doi: 10.1098/rstb.2016.0167.
74. Cuzick J, Edwards R. 1990. Spatial clustering for inhomogenous populations. *Journal of the Royal Statistical Society. Series B.* 52, 73-104.
75. Davis JC. 1973. *Statistics and Data Analysis in Geology.* New York: John Wiley & Sons, 550p.
76. Davis ROE, Bennett HH. 1927. Grouping of soils on the basis of mechanical analysis, Department Circular 419, U.S. Dept. of Agric., Washington, D. C.
77. DEM: Minnesota LiDAR based Digital Elevation Model. 2012. 30-meter resolution. Published by Management Resources Management Information Systems Unit of Minnesota Department of Natural Resources. ID: elev\_dem03ibra4.
78. Department of the Interior Bison Leadership Team and Working Group and Biological Resource Management Division, National Park Service. 2014. DOI Bison Report: Looking Forward. Natural Resource Report NPS/NRSS/BRMD/NRR - 2014/821. NPS 909/124952. National Park Service. Fort Collins, Colorado.
79. Dhand NK, Eppleston J, Whittington RJ, Toribio J. 2009. Association of farm soil characteristics with ovine Johne's disease in Australia. *Prev. Vet. Med.* 89(1-2), 110-120.
80. Diggle P, Rowlingson B, Su T. 2005. Point process methodology for on-line spatiotemporal disease surveillance. *Environmetrics* 16, 423–434.
81. Diggle P. 2007. Spatio-temporal point processes: methods and applications. In Finkenstadt B., L. Held, and V. Isham (eds.), *Statistical Methods for Spatio-Temporal Systems*, pp. 1–45. London: CRC Press.
82. Ditzler C, Scheffe K, Monger HC.(eds.). 2017. Soil Science Division Staff. Soil survey manual. USDA Handbook 18. Government Printing Office, Washington, D.C.

83. Dixon PM. 2002. Ripley's K function. *Encyclopedia of Environments*. Eds. El-Shaarawi AH and Piegorsch WW. 3: 1796-1803. ISBN: 0471899976.
84. Dore E, Pare J, Cote G, Buczinski S, Labrecque O, Roy JP, Fecteau G. 2012. Risk Factors Associated with Transmission of *Mycobacterium avium* subsp *paratuberculosis* to Calves within Dairy Herd: A Systematic Review. *J. Vet. Intern. Med.* 26: 32-45.
85. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. 36(1):27-46. doi: 10.1111/j.1600-0587.2012.07348.x. PubMed PMID: WOS:000315892600003.
86. Downing JA, Anderson MR. 1985. Estimating the standing biomass of aquatic macrophytes. *Can. J. Fish. Aquat. Sci.* 42:1860-1869.
87. Drake JM, Bossenbroek JM. 2009. Profiling ecosystem vulnerability to invasion by zebra mussels with support vector machines. *Theoretical Ecology*. Dec;2:189-198.
88. Driks A. 2009. The *Bacillus anthracis* spore. *Molecular Aspects of Medicine*. 30(6):368-73. doi: 10.1016/j.mam.2009.08.001. PubMed PMID: WOS:000272755000004.
89. Dufour B, Plee L, Moutou F, Boisseleau D, Chartier C, Durand B, et al. 2011. A qualitative risk assessment methodology for scientific expert panels. *Revue Scientifique Et Technique-Office International Des Epizooties*; 30: 673-681.
90. Eldeiry AA, Garcia LA. 2010. Comparison of Ordinary Kriging, Regression Kriging, and Cokriging Techniques to Estimate Soil Salinity Using LANDSAT Images. *Journal of Irrigation and Drainage Engineering*. 136(6):355-64. doi: 10.1061/(asce)ir.1943-4774.0000208. PubMed PMID: WOS:000277745200001.
91. Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*; 17: 43-57.
92. Elliott GN, Hough RL, Avery LM, Maltin CA, Campbell CD. 2015. Environmental risk factors in the incidence of Johne's disease. *Crit. Rev. Microbiol.* 41(4): 488-507.
93. Elliott P, Wakefield J. 2001. Disease clusters: should they be investigated, and, if so, when and how? *J. R. Stat. Soc. A Stat. Soc.* 164, 3-12.
94. Elliott P, Wartenberg D. 2004. Spatial epidemiology: Current approaches and future challenges. *Environmental Health Perspectives* 112:998-1006.
95. Elvander M, Persson B, Lewerin SS. 2017. Historical cases of anthrax in Sweden 1916-1961. *Transboundary and Emerging Diseases*. 64(3):892-8. doi: 10.1111/tbed.12456. PubMed PMID: WOS:000400161300023.

96. ESRI. 2007. Environmental System Research Institute. ArGIS Desktop: Geometric interval. Available online at:  
[http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?topicname=geometrical\\_interval](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?topicname=geometrical_interval).
97. ESRI. 2016. Environmental System Research Institute. ArGIS Desktop: Release 10.3.1. Redlands, CA, USA.
98. ESRI: Environmental Systems Research Institute. ArcMap Version 10.3.4. 2017. Environmental Research Institute, Inc. Redlands, CA, USA.
99. Federal register 65. 2000. Available online at:  
[https://www.aphis.usda.gov/animal\\_health/animal\\_diseases/johnes/downloads/docket98-037-2.pdf](https://www.aphis.usda.gov/animal_health/animal_diseases/johnes/downloads/docket98-037-2.pdf).
100. Fielding AH, Haworth PF. 1995. Testing the generality of bird-habitat models. *Conservation Biology* 9: 1466–81.
101. Fielding AL, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38-49
102. Fisheries Stream Survey Manual. 2007. Stream Survey Methods. Published by Minnesota Department of Natural Resources. Special publication No. 165. Version 2.1. May, 2007. Available from: [http://files.dnr.state.mn.us/publications/fisheries/special\\_reports/165.pdf](http://files.dnr.state.mn.us/publications/fisheries/special_reports/165.pdf).
103. Fotheringham AS, Charlton ME, Brunsdon C. 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ. Plann. A.* 30:1905–1927.
104. Fotheringham AS, Wong DWS. 1991. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environ Plan A.* 23:1025–1044
105. Frome EL, Checkoway H. 1985. Use of Poisson Regression Models in Estimating Incidence Rates and Ratios. *Am J Epidemiol.* 121:309–323
106. Gallardo B, Ermgassen P, Aldridge DC. 2013. Invasion ratcheting in the zebra mussel (*Dreissena polymorpha*) and the ability of native and invaded ranges to predict its global distribution. *J Biogeogr.* 40, 2274-2284.
107. Gallardo B. 2014. Europe's top 10 invasive species: relative importance of climatic, habitat and socio-economic factors. *Ethology Ecol. Evol.* 26(2-3), 130-151.
108. Garcia AB, Shalloo L. 2015. Invited review: The economic impact and control of paratuberculosis in cattle. *J. Dairy Sci.* 98(8): 5019-5039.
109. Geary RC. 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician* 5(3): 115–145.



110. Gelman A, Price PN, Lin C. 2000. A method for quantifying artefacts in mapping methods illustrated by application to headbanging. *Statistics in Medicine* 19: 2309-20
111. Gesler WM, Albert DP. 2000. How spatial analysis can be used in medical geography. In: Albert D.P., Gesler W.M., Levergood B. (Eds.), *Spatial Analysis, GIS and remote sensing applications in the health sciences*. Ann Arbor Press, Chelsea, MI, pp.11-38.
112. Getis A, Ord JK. 1992. The analysis of spatial association by distance statistics. *Geogr. Anal.* 24 (3), 189–206
113. Gonda MG, Chang YM, Shook GE, Collins MT, Kirkpatrick BW. 2007. Effect of *Mycobacterium paratuberculosis* infection on production, reproduction, and health traits in US Holsteins. *Prev. Vet. Med.* 80: 103-119.
114. Goodman MS, Thompson VLS. 2017. The science of stakeholder engagement in research: classification, implementation, and evaluation. *Translational Behavioral Medicine*. Sep;7:486-491.
115. Goovaerts P. 1997. *Geostatistics for natural resource evaluation*, Oxford University Press, New York, New York.
116. GRASS Development Team. 2017. *Geographic Resources Analysis Support System (GRASS) Software, Version 7.2*. Open Source Geospatial Foundation. Electronic document. <http://grass.osgeo.org>
117. Grewal SK, Rajeev S, Sreevatsan S, Michel FC. 2006. Persistence of *Mycobacterium avium* subsp *paratuberculosis* and other zoonotic pathogens during simulated composting, manure packing, and liquid storage of dairy manure. *Appl. Environ. Microbiol.* 72(1): 565-574.
118. Griffith DA, Amrhein CG. 1983. An Evaluation of Correction Techniques for Boundary Effects in Spatial Statistical Analysis: Traditional Methods. *Geographical Analysis*, Vol. 15(4): 352–360.
119. Griffith DA. 1985. An Evaluation of Correction Techniques for Boundary Effects in Spatial Statistical Analysis: Contemporary methods. *Geographical Analysis*. 17(1): 81-88.
120. Haining R. 2003. *Spatial data analysis: theory and practice*. Cambridge University Press, Cambridge.
121. Halvorson HO. 1997. Two generations of spore research: from father to son. *Microbiologia*. 13:131-48.

122. Hart TC, Zandbergen PA. 2013. Reference data and geocoding quality Examining completeness and positional accuracy of street geocoded crime incidents. *Policing-an International Journal of Police Strategies & Management*. 36(2):263-94. doi: 10.1108/13639511311329705. PubMed PMID: WOS:000320741900003.
123. Havel JE, Kovalenko KE, Thomaz SM, Amalfitano S, Kats LB. 2015. Aquatic invasive species: challenges for the future. *Hydrobiol*. 750,147-170.
124. Hazelton ML. 2016. Kernel smoothing methods. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
125. Hearne RR, Kritsky CC. 2010. Characteristics of active local water management districts in the Red River Basin. *Water Policy*. 12(6):898-912. doi: 10.2166/wp.2010.145. PubMed PMID: WOS:000282515300008.
126. Hernán MA, Robins JM. 2006. Estimating causal effects from epidemiological data, *J Epidemiol Community Health*. vol. 60 7(pg. 578-586)
127. Hijmans RJ, Elith J. 2016. Species distribution modeling with R. 0.9-1. [cited: June 20, 2016]. Available from: <https://cran.rProject.org/web/packages/dismo/vignettes/sdm.pdf>.
128. Hijmans RJ. 2012. Cross validation of species distribution models: removing spatial sorting bias and calibration with null model. *Ecology*; 93: 679-688.
129. Hohle M. 2016. Infectious Disease Modeling. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
130. Huanxin W, Nan H. 1991. Trend-surface analysis was applied to the study of the geographic distribution for LBWI in China. *Chinese J Hlth Stat*, 8, 12
131. Hughes-Jones M. 1999. 1996-97 global anthrax report. *Journal of Applied Microbiology*. 87(2):189-91. doi: 10.1046/j.1365-2672.1999.00867.x. PubMed PMID: WOS:000082659900002.
132. Hugh-Jones M, Blackburn J. 2009. The ecology of *Bacillus anthracis*. *Molecular Aspects of Medicine*. 30(6):356-67. doi: 10.1016/j.mam.2009.08.003. PubMed PMID: WOS:000272755000003.

133. Hugh-Jones ME, de Vos V. 2002. Anthrax and wildlife. *Revue Scientifique Et Technique De L Office International Des Epizooties*. 21(2):359-83. doi: 10.20506/rst.21.2.1336. PubMed PMID: WOS:000174946300011.
134. Humphry RW, Cameron A, Gunn GJ. 2004. A practical approach to calculate sample size for herd prevalence surveys. *Prev. Vet. Med.* 65: 173-188.
135. Invasive Species Program. 2016. Invasive Species of Aquatic Plants and Wild Animals in Minnesota; Annual Report for 2015, Minnesota Department of Natural Resources, St. Paul, MN. Available online at:  
[http://files.dnr.state.mn.us/natural\\_resources/invasives/ais-annual-report.pdf](http://files.dnr.state.mn.us/natural_resources/invasives/ais-annual-report.pdf).
136. Isaaks EH, Srivastava RM. 1989. *Applied Geostatistics*. Oxford University Press. New York. ISBN-13: 978-0195050134.
137. Jacquez GM, Estberg L, Long A, Do E, Rommel B, Durbeck H, Greiling DA et al. 2012. Clusterseer v.2.05. ClusterSeer User Manual 2: Software for the detection and analysis of event clusters. Biomedware, Inc. Ann Arbor, MI.
138. Jacquez GM, Greiling D, Durbeck H, Estberg L, Do E, Long E, Rommel B. 2002. ClusterSeer. User Guide 2: Software for Identifying Disease Clusters. TerraSeer Press, Ann Arbor, MI. 316 pp
139. Jacquez GM. 1996. Statistical software for the clustering of health events. *Stat. Med.* 15, 951-952.
140. Jacquez GM. 2008. Spatial Cluster Analysis (Chapter 22). *The Handbook of Geographic Information Science*. S. Fotheringham and J. Wilson (Eds.). Blackwell Publishing, pages 395-416
141. Jenks GF. 1967. The data model concept in statistical mapping, *International Yearbook of Cartography*. 7: 186–190.
142. Jensen BB, Ersboll AK, Korsholm H, Skall HF, Olesen NJ. 2014. Spatio-temporal risk factors for viral haemorrhagic septicaemia (VHS) in Danish aquaculture. *Diseases of Aquatic Organisms*, 109(2), 87-97. doi:10.3354/dao02706.
143. Jirsa MA, Boerboom TJ, Chandler VW, Mossler JH, Runkel AC, Setterholm DR. 2011. S-21 Geologic Map of Minnesota-Bedrock Geology. Minnesota Geological Survey. Retrieved from the University of Minnesota Digital Conservancy,  
<http://purl.umn.edu/101466>.
144. Johnson-Ifearulundu YJ, Kaneene JB. 1997. Relationship between soil type and *Mycobacterium paratuberculosis*. *J. Am. Vet. Med. Assoc.* 210: 1735-1740.

145. Jolley DJ, Jerman B, Elliot P. 1992. Socio-economic confounding, in Geographical and Environmental Epidemiology. In: Elliot, P., Cuzick, J., English, D., Stern, R. (Eds.), Methods for small-area studies. Oxford University Press, Oxford, pp. 3-13.
146. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. 2008. Global trends in emerging infectious diseases. *Nature*. 451(7181):990-U4. doi: 10.1038/nature06536. PubMed PMID: WOS:000253313100048.
147. Jung I, Kulldorff M, Richard OJ. 2010. A spatial scan statistics for multinomial data. *Statistic. Med.* 29, 1910–1918.
148. Kafadar K. 1996. Smoothing geographical data, particularly rates of disease. *Statistics in Medicine*. 15:2539–2560
149. Kanankege K, Machado G, Zhang L, Dokkebakken B, Schumann V, Wells S, Perez A, Alvarez J. Use of a voluntary testing program to study the spatial epidemiology of Johne's disease in Minnesota: A cross-sectional study. (In Review-BMC Veterinary Research: 2018).
150. Kanankege KST, Abdrakhmanov SK, Alvarez J, Glaser L, Bender JB, Mukhanbetkaliyev YY, Korennoy FI, Kadyrov AS, Abdrakhmanova AS, Perez AM. Comparison of spatiotemporal patterns of historic animal Anthrax outbreaks in Minnesota and Kazakhstan. (In Review-PLOS One: 2018).
151. Kanankege KST, Alkhamis MA, Perez AM, Phelps NBD. 2018a. Zebra mussels and Eurasian watermilfoil reporting patterns in Minnesota. *Journal of Great Lakes Research*. doi: 10.1016/j.jglr.2018.02.008.
152. Kanankege KST, Alkhamis MA, Phelps NBD, Perez AM. 2018b. A probability co-kriging model to account for reporting bias and recognize areas at high risk for zebra mussels and Eurasian watermilfoil invasions in Minnesota. *Frontiers in Veterinary Science*. doi: 10.3389/fvets.2017.00231
153. Karatayev AY, Burlakova LE, Mastitsky SE, Padilla DK. 2015. Predicting the spread of aquatic invaders: insight from 200 years of invasion by zebra mussels. *Ecol. Appl.* 25, 430-440.
154. Keckler D. 1995. *The Surfer Manual*. Golden, CO.: Golden Software, Inc.
155. Kelsall JE, Diggle PJ. 1995. Kernel estimation of relative risk. *Bernoulli*. 1(1–2), 3–16.
156. Kelting DL, Laxson CL. 2010. Cost and Effectiveness of Hand Harvesting to Control the Eurasian Watermilfoil Population in Upper Saranac Lake, New York. *J. Aquat. Plant Manage.* 48,1-5.

157. Kenefic LJ, Pearson T, Okinaka RT, Schupp JM, Wagner DM, Ravel J, et al. 2009. Pre-Columbian Origins for North American Anthrax. *Plos One*. 4(3):21. doi: 10.1371/journal.pone.0004813. PubMed PMID: WOS:000265496300003.
158. Kennedy DJ, Benedictus G. 2001. Control of *Mycobacterium avium* subsp paratuberculosis infection in agricultural species. *Rev. Sci. Tech. Office International Des Epizooties*. 20 151-179.
159. Kirby RS, Delmelle E, Eberth JM. 2017. Advances in spatial epidemiology and geographic information systems. *Annals of Epidemiology*. 27(1):1-9. doi: 10.1016/j.annepidem.2016.12.001. PubMed PMID: WOS:000393268100001.
160. Knox EG. 1964. The detection of space-time interactions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 13, 25-30.
161. Kopecky KE. 1977. Distribution of paratuberculosis in Wisconsin, by soil regions. *J. Am. Vet. Med. Assoc.* 170(3): 320-324.
162. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. 2005. A space-time permutation scan statistic for disease outbreak detection. *Plos Medicine*. 2(3):216-24. doi: 10.1371/journal.pmed.0020059. PubMed PMID: WOS:000228382500013.
163. Kulldorff M, Mostashari F, Duczmal L, Katherine YW, Kleinman K, Platt R. 2007. Multivariate scan statistics for disease surveillance. *Stat. Med.* 26, 1824-1833.
164. Kulldorff M, Nagarwalla N. 1995. Spatial disease clusters and inference. *Stat. Med.* 14, 1. 799–810.
165. Kulldorff M, Nagarwalla N. 1995. Spatial disease clusters and inference. *Stat. Med.* 14, 799-810.
166. Kulldorff M. 1997. A spatial scan statistic. *Commun. Stat.: Theory Methods* 26, 1481–1496.
167. Kulldorff M. 2009. SaTScanTMv9.4.1: software for the spatial and space-time scan statistics. <http://www.satscan.org/>.
168. Kunkel KE, Easterling DR, Redmond K, Hubbard K. 2003. Temporal variations of extreme precipitation events in the United States: 1895–2000. *Geophys Res Lett.* 30(1900):10.1029/2003GL018052.
169. LandScan™. 2012. High Resolution Global Population Data. UT-Battelle, LLC. Oak Ridge National Laboratory. Contract No. DE-AC05-00OR22725. U.S. Department of Energy. <http://www.ornl.gov/landscan/>.

170. Latzka AW, Hansen GJA, Kornis M, Vander Zanden MJ. 2016. Spatial heterogeneity in invasive species impacts at the landscape scale. *Ecosphere*. Mar;7:15.
171. Lawson AB, Bohning D, Biggeri A, et al. eds. 1999. *Disease mapping and its uses. disease mapping and risk assessment for public health*. Chichester: Wiley.
172. Lawson AB, Choi J, Cai B, Hossain M, Kirby RS, Liu JH. 2012. Bayesian 2-Stage Space-Time Mixture Modeling With Spatial Misalignment of the Exposure in Small Area Health Data. *Journal of Agricultural Biological and Environmental Statistics*. Sep;17:417-441.
173. Lawson AB. 2006. *Statistical methods in Spatial Epidemiology*. 2nd Ed. John Wiley & Sons Ltd., The Atrium, Southern Gate, Chichester, West Sussex, England. ISBN-13: 978-0-470-01484-4
174. Lawson AB. 2013. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. Chapman and Hall/CRC press.
175. Leroux BG, Lei X, Breslow N. 2000. Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer. pp. 179–191.
176. Les DH, Mehrhoff L. 1999. Introduction of nonindigenous aquatic vascular plants in southern New England: A historical perspective. *Biol. Invasions* 1, 281–300.
177. Leung B, Roura-Pascual N, Bacher S, Heikkila J, Brotons L, Burgman MA, Dehnen-Schmutz K, Essl F, Hulme PE, Richardson DM, Sol D, Vila M. 2012. TEASIng apart alien species risk assessments: a framework for best practices. *Ecol. Letters*. 15, 1475-1493.
178. Lefever DW. 1926. Measuring geographic concentration by means of the standard deviation ellipse. *Am. J. Sociol.* 32, 88–94.
179. Levine N. 2010. *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.3)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. July.
180. Lindahl JF, Grace D. 2015. The consequences of human actions on risks for infectious diseases: a review. *Infect Ecol Epidemiol* ;5:30048. doi: 10.3402/iee.v5.30048. eCollection 2015.
181. Lodge DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA, Carlton JT, McMichael A et al. 2006. Biological invasions: Recommendations for US policy and management. *Ecol. Appl.* 16, 2035-2054.

182. Lodge, DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA, Carlton JT, McMichael A. 2006. Biological invasions: Recommendations for US policy and management. *Ecol. Appl.* 16, 2035-2054.
183. Lombard JE, Gardner IA, Jafarzadeh SR, Fossler CP, Harris B, Capsel RT, Wagner BA, Johnson WO. 2013. Herd-level prevalence of *Mycobacterium avium* subsp *paratuberculosis* infection in United States dairy herds in 2007. *Prev. Vet. Med.* 108: 234-238.
184. Lombard JE. 2011. Epidemiology and economics of paratuberculosis. *Vet. Clin. North Am. Food Anim. Pract.* 27: 525–535
185. Loth L, Gilbert M, Wu G, Czarnecki C, Hidayat M, Xiao X. 2011. Identifying risk factors of highly pathogenic avian influenza (H5N1 subtype) in Indonesia. *Preventive Veterinary Medicine.* Volume 102, Issue 1, 1 October 2011, Pages 50–58.
186. Lukhnova LY, Ajkimbaev AM, Pazylov EK. 2004. Epidemic process of Anthrax in Kazakhstan. *Bull. Agric. Sci. Kazakhstan, Bastau.* 7:44. (in Russian).
187. Lund K, Bloodsworth K, Wolbers T, Weling C, Gamble A. 2015. Guidance for conducting aquatic invasive species early detection and baseline monitoring in lakes. Invasive Species Program. Division of Ecologicalk and Water Resources of the Minnesota Department of Natural Resources. Available from: [https://www.ifound.org/files/6714/4745/1209/ais\\_detection-baseline-monitoring.pdf](https://www.ifound.org/files/6714/4745/1209/ais_detection-baseline-monitoring.pdf)
188. Lusting LK. 1969. Trend-surface analysis of the Basin and Range Province, and some geomorphic implications, U.S. Geol. Survey Prof. Paper 500-D, 70p.
189. Machado G, Kanankege K, Schumann V, Wells S, Perez A, Alvarez J. 2018. Identifying individual animal factors associated with *Mycobacterium avium* subsp *paratuberculosis* (MAP) milk ELISA positivity in dairy cattle in the Midwest region of the United States. *BMC Vet. Res.* 14: 28.
190. MacIsaac HJ. 1996. Potential abiotic and biotic impacts of zebra mussels on the inland waters of North America. *Am. Zoologist* 36, 287-299
191. Mackaness WA. 1991. Integration and Evaluation of Map Generalization. In: Buttenfield, B., McMaster, R.B. (Eds.), *Map Generalization: Making rules for knowledge representation.* . Longman, United Kingdom, pp. 217-226.
192. MacNab YC, Dean CB. 2004. Autoregressive Spatial Smoothing and Temporal Spline Smoothing for Mapping Rates. *Biometrics.* 57(3): 949-956. doi.org/10.1111/j.0006-341X.2001.00949.x.

193. MacNab YC, Gustafson P. 2007. Regression B-spline smoothing in Bayesian disease mapping: With an application to patient safety surveillance. *Statistics in Medicine*. 26(24):4455-74. doi: 10.1002/sim.2868. PubMed PMID: WOS:000250158100007.
194. MacQueen JB. 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Symposium on Math, Statistics, and Probability* (pp. 281–297). Berkeley, CA: University of California Press.
195. Madsen JD, Richardson RJ, Wersal RM. 2012. Managing Aquatic Vegetation. *Small Impoundment Management in North America*. In Neal JW, Willis DW, editors. American Fisheries Society, Bethesda. pp 275-305
196. Mandrak NE, Cudmore B. 2015. Risk assessment: Cornerstone of an aquatic invasive species program. *Aquatic Ecosystem Health and Management*; 18: 312-320.
197. Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*. 27 (2): 209–220.
198. March D, Susser E. 2006. The eco- in eco-epidemiology. *International Journal of Epidemiology*. 35(6):1379-83. doi: 10.1093/ije/dyl249. PubMed PMID: WOS:000243806400002.
199. Mariella L, Tarantino M. 2010. Spatial Temporal Conditional Auto-Regressive Model: A New Autoregressive Matrix. *Australian Journal of Statistics*. 39(3):223-244.
200. Martinez M, Perez AM, de la Torre A, Iglesias I, Munoz MJ. 2008. Association between number of wild birds sampled for identification of H5N1 avian influenza virus and incidence of the disease in the European Union. *Transbound. Emerg. Dis*. 55, 393-403.
201. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. 2009. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound. Emerg. Dis*. 56, 109-120.
202. Matheron G. 1971. *The Theory of the Regionalized Variables and its Applications*; Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau No. 5; Published by the École Nationale Supérieure des Mines de Paris, p 211.
203. MATLAB and Statistics Toolbox TM Release. 2012. The MathWorks, Inc., Natick, Massachusetts, United States.
204. Mayer JD. 1983. The role of spatial analysis and geographic data in the detection of disease causation. *Pergamon Press Ltd. Soc. Sci. Med*. Vol.17: (16) 1213-1221



205. McCauley M. 1976. Khrushchev and the development of Soviet agriculture. The Virgin Land Programme 1953 – 1964. Palgrave Macmillan UK. Online ISBN 978-1-349-03059-0. Doi: 10.1007/978-1-349-03059-0
206. McKenna SLB, Keefe GP, Tiwari A, VanLeeuwen J, Barkema HW. 2006. Johne's disease in Canada Part II: Disease impacts, risk factors, and control programs for dairy producers. *Can. Vet. J.* 47: 1089-1099.
207. Mellina E, Rasmussen JB. 1994. Patterns in the distribution and abundance of zebra mussel (*Dreissena polymorpha*) in rivers and lakes in relation to substrate and other physicochemical factors. *Can. J. Fisheries Aquat. Sci.* 51, 024-1036.
208. Merow C, Smith MJ, Silander Jr. JA. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*; 36: 001-012.
209. Mi C, Huettmann F, Guo Y, Han X, Wen L. 2016. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* 5:e2849. DOI 10.7717/peerj.2849
210. Mills EL, Leach JH, Carlton JT, Secor CL. 1993. Exotic species in the Great-Lakes: A history of biotic crisis and anthropogenic introductions. *J Great Lakes Res.* 19,1-54.
211. Minnesota Statute 477A.19. 2016. Aquatic Invasive Prevention Aid. The office of the Reviser of the Statutes. Chapter 308. Article 1. Section 11. Available from: <https://www.revisor.mn.gov/statutes/?id=477A.19&format=pdf>.
212. Mitas L, Mitasova H. 1999. Spatial Interpolation. In: P.Longley, M.F. Goodchild, D.J. Maguire, D.W.Rhind (Eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications*, Wiley
213. MNDNR AIS: Minnesota Department of Natural Resources: Aquatic Invasive Species. 2016. Available from: <http://www.dnr.state.mn.us/invasives/ais/infested.html>
214. MNDNR WIP: Watercraft Inspection Program of the Minnesota Department of Natural Resources. 2014. Available from: [http://files.dnr.state.mn.us/natural\\_resources/invasives/mndnr\\_ais\\_watercraft\\_inspection\\_handbook.pdf](http://files.dnr.state.mn.us/natural_resources/invasives/mndnr_ais_watercraft_inspection_handbook.pdf).
215. MNDR: Minnesota Department of Revenue. Aquatic Invasive Prevention Aid. 2016. [cited: June 10, 2016]. Available from: [http://www.revenue.state.mn.us/local\\_gov/prop\\_tax\\_admin/Pages/ais.aspx](http://www.revenue.state.mn.us/local_gov/prop_tax_admin/Pages/ais.aspx).

216. MNGSC GAP: Minnesota GeoSpatial commons. 2015. Minnesota Department of Natural Resources (DNR) - Division of Forestry. GAP Land Cover. Accessed 15 Aug 2017. Available from: <https://gisdata.mn.gov/dataset/biota-landcover-gap>
217. MNGSC Hydrography: Minnesota GeoSpatial commons. MNDNR Hydrography data layer. 2015. [cited August 15, 2015]. Available from: <https://gisdata.mn.gov/dataset/water-dnr-hydrography>.
218. MNGSC Roads: Minnesota GeoSpatial commons. 2012. Roads, Minnesota 2012 data layer. Published by Minnesota Department of Transportation (MnDOT). [cited: May 15, 2016]. Available online at: <https://gisdata.mn.gov/dataset/trans-roads-mndot-tis>. Last updated: 02/25/2016.
219. MNGSC Stream: Minnesota GeoSpatial commons. Stream Routes with Kittle Numbers and Mile Measures data layer. 2015. Content update: 11/01/2013. [cited: November 15, 2015]. Available from: <https://gisdata.mn.gov/dataset/water-measured-kittle-routes>.
220. MNGSC: Minnesota GeoSpatial commons. 2015. Lake Bathymetric Outlines, Contours, Vegetation, and DEM: Lake Bathymetric Digital Elevation Model (DEM) data layer. Published by MNDNR – Fisheries department. ID: water\_lake\_bathymetry. Last updated: April 24, 2015. [cited: May 15, 2016]. Available from: <https://gisdata.mn.gov/dataset/water-lake-bathymetry>.
221. Moore DA, Carpenter TE. 1999. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev.* 21(2):143-6
222. Moran PAP. 1948. The interpretation of statistical maps. *J. Royal Stat. Soc. Series B.* 37: 243–251.
223. Moran PAP. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37, 17–23.
224. Morgan ER, Torgerson PR, Shaikenov BS, Usenbayev AE, Moore ABM, Medley GF, et al. 2006. Agricultural restructuring and gastrointestinal parasitism in domestic ruminants on the rangelands of Kazakhstan. *Veterinary Parasitology.* 139(1-3):180-91. doi: 10.1016/j.vetpar.2006.02.016. PubMed PMID: WOS:000238439100021.
225. Morris LR, Proffitt KM, Asher V, Blackburn JK. 2016. Elk resource selection and implications for anthrax management in Montana. *Journal of Wildlife Management.* 80(2):235-44. doi: 10.1002/jwmg.1016. PubMed PMID: WOS:000369817500007.
226. Morris RD, Munasinghe RL. 1993. Aggregation of existing geographic regions to diminish spurious variability of disease rates. *Stat Med.* 12:1915–1929.
227. Morse SS. 1993. *Emerging viruses.* New York, NY: Oxford University Press.

228. Mullins JC, Garofolo G, Van Ert M, Fasanella A, Lukhnova L, Hugh-Jones ME, et al. 2013. Ecological Niche Modeling of *Bacillus anthracis* on Three Continents: Evidence for Genetic-Ecological Divergence? *Plos One*. 8(8):8. doi: 10.1371/journal.pone.0072451. PubMed PMID: WOS:000323425700170.
229. Mungiole M, Pickle LW, Simonson KH. 1999. Application of a weighted headbanging algorithm to mortality data maps. *Statistics in Medicine*. 18:3201–3209.
230. Murtaugh MP, Steer CJ, Sreevatsan S, Patterson N, Kennedy S, Sriramarao P. 2017. The science behind One Health: at the interface of humans, animals, and the environment. *Annals of the New York Academy of Sciences*. May;1395:12-32.
231. Nakaya T, Fotheringham AS, Brunsdon C, Charlton M. 2005. Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine*. 24(17):2695-717. doi: 10.1002/sim.2129. PubMed PMID: WOS:000231816900010.
232. NASS: National Agricultural Statistics Service. 2018. Milk Production. ISSN:1949-1557. Available from:  
<http://usda.mannlib.cornell.edu/usda/nass/MilkProd//2010s/2017/MilkProd-02-21-2017.pdf>.
233. NASS: National Agricultural Statistics Service. Census of Agriculture. 2012. United States Department of Agriculture, Washington DC: Government Printing Office. Available from:  
[https://www.agcensus.usda.gov/Publications/2012/Full\\_Report/Volume\\_1,\\_Chapter\\_2\\_US\\_State\\_Level/st99\\_2\\_011\\_011.pdf](https://www.agcensus.usda.gov/Publications/2012/Full_Report/Volume_1,_Chapter_2_US_State_Level/st99_2_011_011.pdf). 2012; Table 11. p. 343
234. Nielsen SS, Toft N. 2008. Ante mortem diagnosis of paratuberculosis: A review of accuracies of ELISA, interferon-gamma assay and faecal culture techniques. *Vet. Microbiol*. 129: 217-235.
235. Oden N. 1995. Adjusting Moran's I for population density. *Statistics in Medicine* 14: 17-26.
236. Office International Des Epizootis (OIE). Terrestrial Animal Health Code. 2018. Chapter 11.4. Bovine Spongiform Encephalopathy. Article 11.4.1. Available online at:  
[http://www.oie.int/fileadmin/Home/eng/Health\\_standards/tahc/current/chapitre\\_bse.pdf](http://www.oie.int/fileadmin/Home/eng/Health_standards/tahc/current/chapitre_bse.pdf). Cited: 12.24.2018.
237. Oliver MA, Webster R, Lajaunie C, Muir KR, Parkes SE, Cameron AH, Stevens MCG, Mann JR. 1998. Binomial co-kriging for estimating and mapping the risk of childhood cancer. *Ima J. Math. Appl. Med. Biol*. 15, 279-297.

238. Olson JR, Hawkins CP. 2012. Predicting natural base-flow stream water chemistry in the western United States. *Water Resources Research*. Feb;48.
239. Openshaw S. 1984. *The Modifiable Areal Unit Problem. Concepts and Techniques in Modern Geography*. Norwich, UK: Geo Books.
240. Ord JK, Getis A. 1995. Local spatial autocorrelation statistics – Distributional issues and an application. *Geogr. Anal.* 1995; 27: 286-306.
241. Osawa T, Mitsuhashi H, Niwa H. 2013. Many alien invasive plants disperse against the direction of stream flow in riparian areas. *Ecol. Complex.* 15, 26-32.
242. Ott SL, Wells SJ, Wagner BA. 1999. Herd-level economic losses associated with Johne's disease on US dairy operations. *Prev. Vet. Med.* 40: 179-192.
243. Ottorino C, David A, Wolfgang B, Beniyamin C, Mehmet D, Dan D, et al. 2008. *Anthrax in humans and animals*. ISBN 978 92 4 154753 6.
244. Papes M, Havel JE, Vander Zanden MJ. 2016. Using maximum entropy to predict the potential distribution of an invasive freshwater snail. *Freshwater Biology*. Apr;61:457-471.
245. Perez A. 2015. Past, Present, and Future of Veterinary Epidemiology and Economics: One Health, Many Challenges, No Silver Bullets. Specialty Grand Challenge. *Front Vet Sci.* 2: 60. doi: 10.3389/fvets.2015.00060.
246. Perez AM, Thurmond MC, Carpenter TE. 2006. Spatial distribution of foot-and-mouth disease in Pakistan estimated using imperfect data. *Prev. Vet. Med.* 76, 280-289.
247. Perez AM, Thurmond MC, Grant PW, Carpenter TE. 2005. Use of the scan statistic on disaggregated province-based data: Foot-and-mouth disease in Iran. *Preventive Veterinary Medicine.* 71(3-4): 197-207. DOI: 10.1016/j.prevetmed.2005.07.005.
248. Personal communications: Bender JB. 2018. Vaccine originated cases during late 1990's in few farms in Minnesota.
249. Peterson AT, Sanchez-Cordero V, Ben Beard C, Ramsey JM. 2002. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. *Emerging Infectious Diseases.* 8(7):662-7. doi: 10.3201/eid0807.010454. PubMed PMID: WOS:000176394200004.
250. Peterson AT, Tian H, Martínez-Meyer E, Soberón J, Sánchez-Cordero V, Huntley B. 2005. Modeling distributional shifts of individual species and biomes. In: Lovejoy TE, Hannah L, editors. *Climate change and biodiversity*. New Haven (CT): Yale University Press. p. 211–28.

251. Pfeiffer DU, Otte MJ, Roland-Holst D, Zilberman D. 2013. A one health perspective on HPAI H5N1 in the Greater Mekong sub-region. *Comparative Immunology Microbiology and Infectious Diseases*, 36(3), 309-319. doi:10.1016/j.cimid.2012.11.005
252. Pfeiffer DU, Stevens KB. 2015. Spatial and temporal epidemiological analysis in the Big Data era. *Preventive Veterinary Medicine*, 122(1-2), 213-220. doi:10.1016/j.prevetmed.2015.05.012.
253. Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.
254. Pickup RW, Rhodes G, Bull TJ, Arnott S, Sidi-Boumedine K, Hurley M, Hermon-Taylor J. 2006. *Mycobacterium avium* subsp paratuberculosis in lake catchments, in river water abstracted for domestic use, and in effluent from domestic sewage treatment works: diverse opportunities for environmental cycling and human exposure. *Appl. Environ. Microbiol.* 72: 4067–77.
255. Pillars RB, Grooms DL, Woltanski JA, Blair E. 2009. Prevalence of Michigan dairy herds infected with *Mycobacterium avium* subspecies paratuberculosis as determined by environmental sampling. *Prev. Vet. Med.* 89: 191-196.
256. Popov YA, Mikshis NI. 2002. Anthrax vaccines. Problems of especially dangerous infections. *Saratov. № 1 (83).* - P. 21-36 (in Russian).
257. Protter MH, Protter PE. 1988. Sequences. In: *Calculus with analytic geometry*. 4th Ed. Jones and Bartlett Publishers, Inc. Boston, MA. pp 386-391.
258. Pysek P, Richardson DM. 2010. Invasive Species, Environmental Change and Management, and Health. Pages 25-55 in A. Gadgil and D. M. Liverman, editors. *Annual Review of Environment and Resources*, Vol 35. Annual Reviews, Palo Alto.
259. QGIS Development Team. 2018. QGIS Geographic Information System. Open Source Geospatial Foundation. URL <http://qgis.org>
260. Quinn A, Gallardo B, Aldridge DC. 2014. Quantifying the ecological niche overlap between two interacting invasive species: the zebra mussel (*Dreissena polymorpha*) and the quagga mussel (*Dreissena rostriformis bugensis*). *Aquatic Conservation-Marine and Freshwater Ecosystems*. Jun;24:324-337.
261. 261. R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

262. Rabinowitz PM, Kock R, Kachani M, Kunkel R, Thomas J, Gilbert J, et al. 2013. Toward proof of concept of a one health approach to disease prediction and control *Emerging Infectious Diseases*, 19(12).
263. Raizman EA, Fetrow J, Wells SJ, Godden SM, Oakes MJ, Vazquez G. 2007. The association between *Mycobacterium avium* subsp. *paratuberculosis* fecal shedding or clinical Johne's disease and lactation performance on two Minnesota, USA dairy farms. *Prev. Vet. Med.* 78:179–195.
264. Raizman EA, Wells SJ, Munoz-Zanzi CA, Tavornpanich S. 2011. Estimated within-herd prevalence (WHP) of *Mycobacterium avium* subsp *paratuberculosis* in a sample of Minnesota dairy herds using bacterial culture of pooled fecal samples. *Canadian Journal of Veterinary Research-Revue Canadienne De Recherche Veterinaire.* 75(2):112-116.
265. Rangel SJ, Pare J, Dore E, Arango JC, Cote G, Buczinski S, Labrecque O, Fairbrother JH, Roy JP, Wellemans V, Fecteau G. 2015. A systematic review of risk factors associated with the introduction of *Mycobacterium avium* spp. *paratuberculosis* (MAP) into dairy herds. *Can. Vet. J.* 56: 169-177.
266. Read JS, Winslow LA, Hansen GJA, Van den Hoek J, Hanson PC, Bruce LC, et al. 2014. Simulating 2368 temperate lakes reveals weak coherence in stratification phenology. *Ecological Modelling*; 291:142-150. doi: 10.1016/j.ecolmodel.2014.07.029
267. Rezaeian M, Dunn G, St Leger S, Appleby L. 2007. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *Journal of Epidemiology and Community Health.* 61(2):98-102. doi: 10.1136/jech.2005.043117. PubMed PMID: WOS:000243597800005.
268. Ricardo Ocaña-Riola. 2010. Common errors in disease mapping. *Geospatial Health* 4(2): 139-154.
269. Ricciardi A. 2001. Facilitative interactions among aquatic invaders: is an "invasional meltdown" occurring in the Great Lakes? *Can. Fish. Aquat. Sci.* 58, 2513-2525.
270. Rideout BA, Brown ST, Davis WC, Gay JM, Giannella RA, Hines II ME, Hueston WD, Hutchinson LJ. et al. 2003. Committee on Diagnosis and Control of Johne's Disease Board on Agriculture and Natural Resources. Division on Earth and Life Studies. National Academy of Sciences. USA. ISBN: 0-309-56641-X, p.244, 6x9
271. Ripley BD. 1977. Modelling spatial patterns. *J. R. Stat. Soc. Series B Stat. Methodol.*, 39, pp. 172-192

272. Ritter C, Kwong GPS, Wolf R, Pickel C, Slomp M, Flaig J, Mason S, Adams CL, Kelton DF, Jansen J, De Buck J, Barkema HW. 2015. Factors associated with participation of Alberta dairy farmers in a voluntary, management-based Johne's disease control program. *J. Dairy Sci.* 98(11):7831-7845.
273. Robinson S, Milner-Gulland EJ. 2003. Political change and factors limiting numbers of wild and domestic ungulates in Kazakhstan. *Human Ecology.* 31(1):87-110. doi: 10.1023/a:1022834224257. PubMed PMID: WOS:000181621600005.
274. Rogers DJ, Sedda L. 2012. Statistical models for spatially explicit biological data. *Parasitol.* 139: 1852-1869.
275. Roley SS, Newman RM. 2008. Predicting Eurasian watermilfoil invasions in Minnesota. *Lake Reserv. Manage.* 24, 361-369.
276. Rowe WC. 2009. Turning the Soviet Union into Iowa: The Virgin Lands Program in the Schuch R, Fischetti VA. *The Secret Life of the Anthrax Agent Bacillus anthracis: Bacteriophage-Mediated Ecological Adaptations.* Plos One. 4(8). doi: 10.1371/journal.pone.0006532. PubMed PMID: WOS:000268935900002.
277. Rushworth A, Lee D, Sarran C. 2017. An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk. *Journal of the Royal Statistical Society Series C-Applied Statistics.* 66(1):141-57. doi: 10.1111/rssc.12155. PubMed PMID: WOS:000392808300002.
278. Salgado M, Collins MT, Salazar F, Kruze J, Bolske G, Soderlund R, Juste R, Sevilla IA, Biet F, Troncoso F, Alfaro M. 2011. Fate of Mycobacterium avium subsp paratuberculosis after Application of Contaminated Dairy Cattle Manure to Agricultural Soils. *Appl. Environ. Microbiol.* 77(6): 2122-2129.
279. Sanson RL, Pfeiffer DU, Morris RS. 1991. Geographic information systems:their application in animal disease control. *Rev. sci. tech. Off. int. Epiz.* 10 (1), 179-195
280. SAS Institute Inc. 2015. SAS/STAT® 14.1 User's Guide. Cary, NC:SAS Institute Inc.
281. Scott HM, Sorensen O, Wu JT, Chow EY, Manninen K. 2007. Seroprevalence of and agroecological risk factors for Mycobacterium avium subspecies paratuberculosis and Neospora caninum infection among adult beef cattle in cow-calf herds in Alberta, Canada. *Can. Vet. J.* 48(4): 397-406.
282. Seekamp E, McCreary A, Mayer J, Zack S, Charlebois P, Pasternak L. 2016. Exploring the efficacy of an aquatic invasive species prevention campaign among water recreationists. *Biological Invasions.* 18: 1745-1758.

283. Shanafelt DW, Perrings C. 2018. The Effect of the Post 2001 Reforms on FMD Risks of the International Live Animal Trade. *Ecohealth*. Jun;15:327-337.
284. Silverman BW. 1986. *Density estimation for statistics and data analysis*, Boca Raton, FL: Chapman & Hall/CRC.
285. Simberloff D. 2009. The Role of Propagule Pressure in Biological Invasions. *Annual Review of Ecology Evolution and Systematics*. 40:81-102. Doi: 10.1146/annurev.ecolsys.110308.120304
286. Slunge D, Drakenberg O, Ekbohm A, Göthberg M, Knaggård A, Sahlin U. 2017. Stakeholder interaction in research process - A guide for researcher and research groups. Available online at: [https://gmv.gu.se/digitalAssets/1619/1619929\\_stakeholder-interaction-in-research-processes---guide---final-march-15-2017.pdf](https://gmv.gu.se/digitalAssets/1619/1619929_stakeholder-interaction-in-research-processes---guide---final-march-15-2017.pdf).
287. Soil Survey Staff. 2017. Gridded Soil Survey Geographic (gSSURGO) Database for State name. United States Department of Agriculture, Natural Resources Conservation Service. Available online at <http://datagateway.nrcs.usda.gov/>. Accessed 10 Feb 2018.
288. Soil Survey Staff. Hydrologic Soil Group. 2018. Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at the following link: <https://websoilsurvey.sc.egov.usda.gov/>. Accessed 10 Jan 2018.
289. Soil survey staff. USDA soil texture triangle classification. Based on: Davis ROE, Bennett HH. 1927. Grouping of soils on the basis of mechanical analysis, Dep. Circ. 419, U.S. Dep. of Agric., Washington, D. C. Available from: [https://www.nrcs.usda.gov/Internet/FSE\\_DOCUMENTS/stelprdb1044818.pdf](https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1044818.pdf). 2017.
290. Song TM, Ryu S. 2015. Big data analysis framework for healthcare and social sectors in Korea. *Healthc Inform Res*. 21:3–9.10.4258/hir.2015.21.1.3
291. Sorge US, Moon RD, Stromberg BE, Schroth SL, Michels L, Wolff LJ, et al. 2015. Parasites and parasite management practices of organic and conventional dairy herds in Minnesota. *Journal of Dairy Science*. 98(5):3143-51. doi: 10.3168/jds.2014-9031. PubMed PMID: WOS:000353267900033. Soviet Union. Brunn S, editor. 2011. *Engineering Earth*. Springer, Dordrecht. doi.org/10.1007/978-90-481-9920-4\_15.
292. Spencer DF, Carruthers RI. 2013. Predicting Eurasian watermilfoil's (*Myriophyllum spicatum*) distribution and its likely response to biological control in a spring-fed river. *J. Aquat. Plant Manage*. 51, 7-14.
293. Stark CR, Mantel N. 1967. Lack of seasonal or temporal spatial clustering of Down's Syndrome births in Michigan. *American Journal of Epidemiology* 86: 199-213.



294. StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC
295. Sternbach G. 2003. The history of anthrax. *Journal of Emergency Medicine*. 24(4):463-7. doi: 10.1016/s0736-4679(03)00079-9. PubMed PMID: WOS:000182902600019.
296. Stevens KB, Pfeiffer DU. 2016. The Role of Spatial Analysis in Risk-Based Animal Disease Management. *Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
297. Stewart-Koster B, Olden JD, Johnson PTJ. 2015. Integrating landscape connectivity and habitat suitability to guide offensive and defensive invasive species management. *J Appl. Ecol.* 52, 366-378.
298. Stockwell D, Peters D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *Int J Geogr Inf Sci* 13: 143–158.
299. Stockwell DRB, Peterson AT. 2002. Effects of sample size on accuracy of species distribution models. *Ecol Modell* 148: 1–13.
300. Stockwell DRB. 1992. Machine learning and the problem of prediction and explanation in ecological modelling. Ph.D. Thesis, Australian National University.
301. Stoltenow CL. 2012. Unified Anthrax Recommendations for MB, MN, MT, ND, & SD. Available online: [https://www.nd.gov/ndda/sites/default/files/legacy/resource/UnifiedAnthraxRecommendations\\_0.pdf](https://www.nd.gov/ndda/sites/default/files/legacy/resource/UnifiedAnthraxRecommendations_0.pdf)
302. Susanto F, de Souza P, He J. 2016. Spatiotemporal Interpolation for Environmental Modelling. Reindl L. Eds. *Sensors (Basel)*. 16(8): 1245. doi: 10.3390/s16081245
303. Susser E. 2004. Eco-epidemiology: Thinking outside the black box. *Epidemiology*. 15(5):519-20. doi: 10.1097/01.ede.0000135911.42282.b4. PubMed PMID: WOS:000223382600004.
304. Swets JA. 1988. Measuring the accuracy of diagnostic systems. *Science*. 240, 1285–1293.
305. Szklo M, Nieto FA. 2007. *Epidemiology Beyond the basics*. 3rd Ed. Jones and Bartlett Learning, LLC. Burlington, MA. pp 153-305.
306. Tamayo M, Olden JD. 2014. Forecasting the Vulnerability of Lakes to Aquatic Plant Invasions. *Invasive Plant Sci. Manage.* 7, 32-45.

307. Taner MU, Carleton JN, Wellman M. 2011. Integrated model projections of climate change impacts on a North American lake. *Ecological Modelling*; 222: 3380-3393.
308. Thrusfield M. 2007. *Veterinary Epidemiology*, 3rd Edition, Blackwell Science, Oxford, UK. pp. 233-234). ISBN: 978-1-405-15627-1
309. Tiwari A, VanLeeuwen JA, McKenna SLB, Keefe GP, Barkema HW. 2006. Johne's disease in Canada - Part I: Clinical symptoms, pathophysiology, diagnosis, and prevalence in dairy herds. *Can. Vet. J.* 47(9): 874-882.
310. Tobler WR. 1970. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* 46: 234 -240.
311. Tomczak M. 1998. Spatial Interpolation and its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW) - Cross-Validation/Jackknife Approach. *Journal of Geographic Information and Decision Analysis*, vol. 2, no. 2, pp. 18-30.
312. Turnbull BW, Iwano EJ, Burnett W, Howe HL, Clark LC. 1990. Monitoring for clusters of disease: Applications to leukaemia incidence in upstate New York. *American Journal of Epidemiology* 132, S136–S143.
313. Turnbull PCB, Lindeque P, Le Roux J, Bennett AM, Parks SR. 1998. Airborne movement of anthrax spores from carcass sites in the Etosha National Park, Namibia. *Journal of Applied Microbiology*. 84(4):667-76. doi: 10.1046/j.1365-2672.1998.00394.x. PubMed PMID: WOS:000073738600026.
314. USAHA: United States Animal Health Association. 1993. Proceedings of the 97th Annual Meeting of the USAHA.
315. USDA: United States Department of Agriculture. 1997. Johne's disease on U.S. dairy operations. In *Johne's disease on U.S. Dairy Operations*. Washington DC: Animal and Plant Health Inspection Service.
316. USDA-APHIS: United States Department of Agriculture - Animal and Plant Health Inspection Service. Disease information. 2015. Preventing Introduction of Johne's disease. Last modified: August 2015. Last cited: March 2018. Available from: [https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/animal-disease-information/cattle-disease-information/sa\\_johnes/ct\\_prevent-intro](https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/animal-disease-information/cattle-disease-information/sa_johnes/ct_prevent-intro).
317. USDA-APHIS-VS: United States Department of Agriculture - Animal and Plant Health Inspection Service - Veterinary Services. 2010. Uniform Program Standards for the Voluntary Bovine Johne's Disease Control Program, APHIS. 91–45–016,

- <http://www.johnesdisease.org>. Accessed 10 Dec 2016. Available online from:  
[https://johnes.org/handouts/files/USDA\\_Program\\_Standards\\_Sept-2010.pdf](https://johnes.org/handouts/files/USDA_Program_Standards_Sept-2010.pdf)
318. Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, Ravel J, et al. 2007. Global Genetic Population Structure of *Bacillus anthracis*. *Plos One*. 2(5). doi: 10.1371/journal.pone.0000461. PubMed PMID: WOS:000207446000009.
  319. Vander Zanden MJ, Olden JD. 2008. A management framework for preventing the secondary spread of aquatic invasive species. *Can. J. Fisheries Aquat. Sci.* 65, 1512-1522.
  320. VanderWaal K, Morrison RB, Neuhauser C, Vilalta C, Perez AM. 2017. Translating big data into smart data for veterinary epidemiology. *Front Vet Sci.* 4:110. doi: 10.3389/fvets.2017.00110.
  321. VanLeeuwen JA, Keefe GP, Tremblay R, Power C, Wichtel JJ. 2001. Seroprevalence of infection with *Mycobacterium avium* subspecies *paratuberculosis*, bovine leukemia virus, and bovine viral diarrhea virus in Maritime Canada dairy cattle. *Can. Vet. J.* 42(3): 193-198.
  322. Vauclin M, Vieira SR, Vachaud G, Nielsen DR. 1983. The use of co-kriging with limited field soil observations. *Soil Sci. Society Am. J.* 47,175-184.
  323. Velasova M, Drewe JA, Gibbons J, Green M, Guitian J. 2015. Evaluation of the usefulness at national level of the dairy cattle health and production recording systems in Great Britain. *Veterinary Record.* 177(12): 304-304
  324. Venables WN, Ripley BD. 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN: 0-387-95457-0
  325. Verbyla DL, Litaitis JA. 1989. Resampling methods for evaluating classification accuracy of wildlife habitat models. *Environmental Management* 13: 783-7.
  326. Vergne T, Paul MC, Chaengprachak W, Durand B, Gilbert M, Dufour B, Roger F, Kasemsuwan S, Grosbois V. 2014. Zero-inflated models for identifying disease risk factors when case detection is imperfect: Application to highly pathogenic avian influenza H5N1 in Thailand. *Prev. Vet. Med.* 114: 28-36.
  327. Wackernagel H. 1994. Co-kriging versus kriging in regionalized multivariate data analysis. *Geoderma*; 62: 83-92.
  328. Waddell L, Rajic A, Stark K, McEwen SA. 2016. *Mycobacterium avium* sub spp. *paratuberculosis* detection in animals, food, water and other sources or vehicles of human exposure: A scoping review of the existing evidence. *Prev. Vet. Med.* 132: 32-48.

329. Wagner Robb S, Bauer SE, Vena JE. 2016. Integration of Different Epidemiologic Perspectives and Applications to Spatial Epidemiology. Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
330. Wagner RS, Bauer SE, Vena JE. 2016. Integration of Different Epidemiologic Perspectives and Applications to Spatial Epidemiology. Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology. Eds: Lawson AB, Banerjee S, Haining RP, Ugarte MD. Chapman & Hall/CRC Press Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
331. Wagnet RJ, Huston JL. 1996. Scale-dependency of solute transport modeling/GIS applications. *Journal of Environmental Quality*. 25 (3): 499-510.
332. Wall MM. 2004. A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, pp. 311-324.
333. Waller LA, Turnbull BW. 1993. The Effects of Scale on Tests for Disease Clustering. *Stat Med*. 12:1869–1884.
334. Wand MP, Jones MC. 1995. Kernel smoothing, Boca Raton, FL: Chapman & Hall/CRC.
335. Wang B, Shi W, Miao Z. 2015. Confidence analysis of standard deviational ellipse and its extension into higher dimensional Euclidean space. *PLOS ONE* 10, DOI: 10.1371/journal.pone. 0118537.
336. Ward MP, Carpenter TE. 2000. Analysis of time-space clustering in veterinary epidemiology. *Prev Vet Med*. 43:225–37.10.1016/S0167-5877(99)00111-7.
337. Ward MP, Perez AM. 2004. Association between soil type and paratuberculosis in cattle herds. *Am. J. Vet. Res*. 65: 10-14.
338. Ward SH. 1914. Minnesota State Livestock Sanitary Board. 11th Annual report. Syndicate Printing co., Minneapolis. pp 31-33.
339. Wells SJ, Wagner BA. 2000. Herd-level risk factors for infection with *Mycobacterium paratuberculosis* in US dairies and association between familiarity of the herd manager with the disease or prior diagnosis of the disease in that herd and use of preventive measures. *J. Am. Vet. Med. Assoc*. 216: 1450-1457.
340. Whittington RJ, Marsh IB, Reddacliff LA. 2005. Survival of *Mycobacterium avium* subsp *paratuberculosis* in dam water and sediment. *Appl. Environ. Microbiol*. 71: 5304-5308.

341. Yadav SK, Chakrapani GJ. 2006. Dissolution kinetics of rock-water interactions and its implications. *Current Science*. Apr;90:932-937.
342. Zandbergen PA. 2008. A comparison of address point, parcel and street geocoding techniques. *Computers Environment and Urban Systems*. 32(3):214-32. doi: 10.1016/j.compenvurbsys.2007.11.006. PubMed PMID: WOS:000257017100005.