Bayesian spatio-temporal evaluations of spotted fever group rickettsioses with socio-economic and environmental factors: 2013 – 2018

by

Ram K. Raghavan

B.Sc., Mahatma Phule Agricultural University, 1998

M.S., Kansas State University, 2005

Ph.D., Kansas State University, 2011

A THESIS

submitted in partial fulfillment of the
requirements for the degree

MASTER OF PUBLIC HEALTH

Department of Diagnostic Medicine and Pathobiology
College of Veterinary Medicine

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Dr. Justin Kastner

Copyright

Abstract

Recent advances in disease mapping allow for the simultaneous evaluation of space-time dynamics of diseases and the drivers of such dynamics, which are useful for designing public health campaigns and surveillance systems. This study determined the space-time patterns of spotted fever group rickettsioses (SFR), a group of tick-borne Rickettsial diseases widely prevalent in the U.S, and further evaluated the associations of socio-economic and environmental (land cover, climate) factors with SFR. County-level SFR cases reported to the Kansas Department of Health and Environment between years 2013 – 2018 and publicly available covariate data were used in a Bayesian hierarchical modeling framework to quantify trends and associations. The results show a steady increase in space-time trend for SFR in Kansas, the spread of SFR to newer counties over the study period, and two clusters of high-risk areas in the southeast and northeastern parts of Kansas. The space-time pattern of SFR is influenced by poverty status, the number of older homes in a county, and higher relative humidity conditions. The relevance of these findings is discussed in the context of public health and climate change implications on health.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I thank my major advisor, Dr. Justin Kastner, and committee members, Drs. Gary Anderson, John Harrington, and T.G. Nagaraja for their kind support and high sense of humor – or, in graduate student language, they were "just fine". Justin knows precisely where to place the comma, one need only ask (and don't forget to call his mobile). Dr. Ellyn Mulcahy and Barta Stevenson will be happy that I am done with this program, albeit in a short time – they have promised this degree will land me a job at the CDC rather soon. Daniel Neises, epidemiologist at the Kansas Department of Health and Environment, facilitated data acquisition for this work, he is a perfectly *well-adjusted* gentleman who is hard to *out-match* when it comes to collaborative projects. Heath Ritter at the University Compliance Office is, well, a great Compliance Officer – give Heath a raise, please! Heath and his staff facilitated IRB approval for this project. And oh, I must not forget to thank my middle/high school English teacher, Mr. Amalraj, for making English learning an enjoyable endeavor (or endeavour, is it?!). He was a madman who beat many children blue, but the English stayed, nevertheless.

Chapter 1

Introduction

Rocky Mountain spotted fever (RMSF) is a serious tick-borne illness with a mortality rate of up to 20-25% if not treated on time with an appropriate antibiotic [1]. Even with an appropriate antibiotic and timely treatment, the mortality rate due to this disease in the U.S currently ranges from 4-6% every year [1], [2]. RMSF is caused by the bacterium, *Rickettsia rickettsii*, which is spread through the bite of an infected tick. *Rickettsia rickettsii* is an obligately intracellular, Gram-negative bacteria that has evolved in close association with arthropods and are mostly found in ticks, mites, and fleas but can also be found associated with small mammals [1], [3]. *Rickettsia rickettsii* is transmitted to humans when an infected tick bites and subsequently transfers contents of its salivary gland and midgut at the bite site under the skin. From the bite site (skin cells), the rickettsia will spread through the bloodstream and infect other skin cells, brain, lungs, heart, kidneys, liver, gastrointestinal tract, and other organs [1], [3].

As of January 2010, the Centers for Disease Control and Prevention (CDC) required all RMSF cases to be reported together with other spotted fever group rickettsioses (SFR) since most routine laboratory tests are unable to distinguish RMSF from other antigenically similar rickettsia species such as *R. parkeri*, *R. akari*, and Rickettsia 364D [4]. Therefore, in this research, I will use SFR to refer to RMSF and potentially other Spotted Fever Rickettsioses

infections that are reported to the health departments in the U.S. Like RMSF, following infection, most people with SFR will have fever, headache, nausea, vomiting, stomach ache, muscle pain, lack of appetite, and rash [2]. Anecdotal information strongly suggests that routine patient examinations do not readily present clinical symptoms that can help detect SFR infection, and patients are often misdiagnosed and prescribed treatments that do not provide cure (personal communication, Chris Paddock, CDC).

At least three tick species are currently associated with the transmission of *R. rickettsii* to humans in North America, but other ticks may be involved in the epidemiology of SFR that are yet to be discovered [5]. The CDC currently considers the American dog tick (*Dermacentor variabilis*), Rocky Mountain wood tick (*D. andersoni*), and the brown dog tick (*Rhipicephalus sanguineus*) to be the transmitting vectors of *R. rickettsii* in different regions of the U.S [5]. All these ticks are in the Ixodidae Family, commonly referred as "hard ticks", and they have a 3-host lifecycle that involves different mammals as blood sources for their three post-emergent life-stages (larva, nymph, adult) [6], [7]. The geographic distribution of SFR is spatially concordant with that of the infected tick species. In general, the American dog ticks are distributed in the central and eastern regions of the U.S, covering the entire eastern half of the nation, the Rocky Mountain wood tick is distributed in the Rocky Mountain states extending towards the Pacific northwest and west coast of the U.S, and the brown dog tick has been known to transmit *R. rickettsii* in the southwestern U.S and Mexico [5].

Ticks are poikilothermic arthropods, whose geographic distribution and seasonal activities are largely influenced by long-term climatic conditions of a region and the local weather [8], [9]. In the recent decades, it has been widely observed that different ticks have either expanded or have shifted their geographic distribution ranges [10–13], and as well as have widened their activity periods early in Spring and late into the Fall seasons [14], [15]. Increased abundance of ticks have also been noted in the northern latitudes [16]. Understandably, such changes will have implications for the diseases they transmit. For instance, illnesses caused by tick-borne pathogens to humans and animals in general have increased

over the past several years in the Midwestern U.S [17], [18], including in the state of Kansas where tick-borne diseases have expanded to newer areas over the years [19], [20]. Some of the increase in the space-time expansion of tick-borne diseases in the Midwestern region may be attributed to geographic expansion of tick populations [13], [14], [21] and ongoing climate change. Other factors such as suburbanization, improved diagnostic methods, and higher awareness of tick-borne diseases by patients and physicians may have also contributed to the apparent space-time expansion; however, these factors are difficult to determine quantitatively.

One of the primary considerations for public health management of tick-borne diseases is knowing the level of intensity and disparities in the distribution of disease occurrences over time and space, and the various factors that potentially determine such dynamics. This knowledge will then allow for designing appropriate management strategies and for proper resource allocation. Risk mapping is a common approach to identifying spatial risks and disparities in disease events. However, plotting frequency measures of diseases (e.g., counts, incidence, prevalence) assigned to different spatial units by itself can be non-informative and potentially misleading [22], [23]. Factors that determine the temporal prevalence and spatial distribution of tick-borne diseases such as SFR are tightly linked; they could differ from one region to another due to natural changes in the geography and human population distribution, and they are quite scale-dependent (e.g., [24], [22]). Furthermore, the common statistical methods applied when determining the relevance of influential factors of a disease may be inadequate as the importance of some variables often become evident only in the presence of others. For instance, climatic variables are often confounded or interactive with other influential factors viz., land cover/land use and socio-economic properties. [19], [25].

Such challenges in disease risk mapping warrant robust methods and a systematic approach to model building, so that their applications in public health contexts can be made with high confidence. Bayesian hierarchical modeling is one such approach that has been recognized as a powerful analytical technique to provide robust estimates of space, and space-

time trends. Unlike most frequentist approaches, Bayesian models allow for incorporating errors that may arise from mean or median estimates of the independent covariates and observed data through the use of prior probability distributions [26], [27].

In this study, I utilized contemporary Bayesian hierarchical methods to evaluate the space-time patterns of SFR occurrences in Kansas during a 6-year period from 2013 to 2018. Using these models, I further determined the influential factors that are potentially driving forces of SFR occurrence patterns in Kansas. A previous publication in 2016 [28] identified similar epidemiological aspects of RMSF but that study considered a four-state region including Kansas, and considered a temporally older data (2005 – 2014).

Chapter 2

Materials and Methods

2.1 Study area

The spatial extent studied in this research included all 105 counties in Kansas, many of which have recorded varied numbers of SFR cases since the first identification of the disease. Cases of SFR are typically recorded in the eastern half of the state with more cases generally occurring in the southeastern counties, which also has relatively higher tick density and abundance (Raghavan, unpublished). Climate in the study area is transitional from east to west, and as well as south to north, with the southeastern region receiving progressively more rainfall than the west and northern parts.

2.2 Data

2.2.1 Ethics statement.

The SFR case diagnosis data used in this research was nationally notifiable by the local and state health departments to the CDC, and were aggregated annually to their respective administrative units, counties. No individually–identifiable information was used for this study. The use of publicly available, archived SFR data for this research was approved by the Internal Review Board at Kansas State University's Office of Research Compliance (IRB

7465).

2.2.2 Epidemiological data.

Annual county–level SFR cases between years 2005 and 2018 were obtained from Kansas Department of Health and Environment (KDHE). SFR cases in this dataset are classified into three categories, 'confirmed', 'probable', and 'suspected' per CDC guidelines. Case classification for SFR changed once during the study period in 2008 and included the 'suspected' criteria for diagnosis. According to the CDC, confirmed cases of SFR are clinically compatible cases (meets clinical evidence criteria) that is laboratory confirmed. Probable SFR cases are those clinically compatible cases (meets clinical evidence criteria) and has supportive laboratory results. And, suspected SFR cases are those laboratory cases with a laboratory evidence of past or present infection but no clinical information available (e.g., laboratory report). In this research, all three categories of SFR cases were considered to indicate positive SFR diagnosis.

2.2.3 Covariate data.

Table 2.1: *Land cover/land use variables considered in the study*

Source	Independent variables
National Land Cover Dataset (NLCD)	Open water, Developed – open space, Developed – low intensity, Developed – medium intensity, Developed – high intensity, barren land, deciduous forest, evergreen forest, mixed forest, scrub/shrub, grassland, herbaceous, pasture/hay, cultivated crops, woody wetlands, emergent herbaceous wetland.

Since arthropod-borne diseases in general are influenced by the physical environment, including climate and land cover/land use, and socio-economic status of people, data from

these themes were considered as potential explanatory factors (covariates) that could determine the space-time distribution of SFR in Kansas. For land cover/land use data, the percentage land occupied by various cover types in each county was estimated from the publicly available 2011 National Land Cover Dataset [29] (Table 2.1).

Climate variables included were averages of annual mean temperature; the maximum normalized vegetation index (NDVI); minimum land surface temperature (LST); mean LST; precipitation, and relative humidity. The LST and NDVI estimates were obtained from the MODIS (Moderate Resolution Imaging Spectroradiometer) imagery [30], while precipitation and relative humidity data were derived from the Prediction of Worldwide Renewable Energy (POWER) web portal of the NASA Langley Research Center [31] (Table 2.2). Climate variables were extracted for a period roughly corresponding to the active tick–season in the region (May through August). All geospatial data were processed in a geographic information systems (GIS) environment using ArcGIS software.

Table 2.2: *Climate variables considered in the study*

Source	Independent variables
NASA Moderate Resolution Imaging Spectroradiometer (MODIS)	Daytime land surface temperature (≥35 °C, 28– 34.9 °C, 24.9–27.9 °C, ≤25 °C), Night time land surface temperature (≤16 °C, 15.9–19.9 °C, ≥20 °C, Diurnal temperature range.
NASA Prediction of Worldwide Renewable Resources (POWER)	Normalized Difference Vegetation Index (NDVI), Daily maximum temperature, Daily minimum temperature, Daily average temperature, Dew point, Relative humidity, Diurnal temperature range.

For socio-economic variables, the U.S. Census 2010 data on population and housing were obtained from the National Historical Geographic Information System (NHGIS), a publicly

available online resource for U.S. Census Bureau's historical and current population data [32] (Table 2.3). From the tables, 20 variables related to housing conditions and 23 variables related to population level socio-economic conditions were extracted for each county by spatial query and joined to the census shapefiles using the common GIS codes.

Table 2.3: *Socio-economic variables considered in the study*

Source	Independent variables
Population	Housing units (total housing units), Tenure (owner occupied, renter occupied), Tenure (Historic or Latino Householder) (owner occupied, renter occupied), Race of householder (white alone, Black or African American alone, Asian alone), Household size (1–person, 2–person, 3–person, 4–person, 5–person, 6–person, or 7–more person household), Year structure built (Built 2005 or later, 2000 to 2004, 1990 to 1999, 1980 or earlier. (20 variables).
Housing	Population (total population), Race (White alone, Black or African American alone, Asian alone), Household income in the past 12 months (Less than $10,000, $10,000 to $14,999, and thirteen other variables representing $49,999 incremental income thereof up to $199,999, and $200,000 or more), Poverty status in the previous 12 months (income in the past 12 months below poverty level, income in the past 12 months at or above poverty level. (23 variables).

2.3 Statistical analysis

2.3.1 Covariate selection.

In order to avoid model overfitting issues that could result in Type-I and/or Type-II statistical errors, the explanatory variables to be included in the Bayesian hierarchical models were screened *a priori* to formal analysis. To achieve this, several frequentist bivariate regression models were fitted, and each variable was independently evaluated. Only variables that were significantly associated with SFR status at a liberal $p \leq 2$ value (candidate variables) were kept for further analysis. A bivariate regression takes the form,

$$Y_{ij} = (\beta_0 + \beta_k.vk_{ij}).$$

where Y_{ij} is SFR relative risk, β_0 the intercept coefficient, and β_k the coefficient for the explanatory variable $vk_{ij}(k = 1, ..., n)$ and (i=1,..., 105 county; and j = year 2005,..., 2018).

While screening variables, care was taken not to remove candidate variables that were deemed clinically relevant [33]. Following this, among the screened variables the presence of multicollinearity was tested by estimating the variance inflation factor (VIF). All variables with a ($VIF \leq 10$) were deemed to be indicative of multicollinearity and variables were dropped one at a time until no multicollinearity was present. Finally, non–linearity among the selected candidate variables was quantified, and those variables with non–linearity were categorized using cut-offs based on scatter–plots.

2.3.2 Bayesian model specification

The cases of SFR reported in each county in the study were notated as Y_{ij} among N_{ij} individuals that are at risk for SFR infection in the population of county i, in year j. Y_{ij} was assumed to follow a Poisson approximation, $(Y_{ij}) \sim Poisson(E_{ij}\theta_{ij})$ where E_{ij} is the expected number of the population at risk for SFR and θ_{ij} is the relative risk. I first

considered the Binomial approximation for this relationship since n is known. However, relatively few SFR infections occur randomly over time within different age-groups, whose members vary throughout the study period due to aging. Therefore, I chose Poisson function over Binomial since the latter is better suited for situations where study population is rather constant. Additionally, SFR prevalence, particularly RMSF prevalence is disproportionate among different age groups [34]; therefore, I calculated standardized rates by assuming five age classes l, (< 5, 5–19, 20–45, 46–65 and > 65). Mapping crude observations can be non–informative or misleading when population in some areal units are small, resulting in large posterior estimates, which in turn render it difficult to distinguish chance variability from genuine differences. The expected number of SFR cases was therefore calculated by

$$\sum_{ij} = \sum_{ij} n_{ijl} \frac{\sum_i \sum_{ijl} Y_{ijl}}{\sum_i \sum_{ijl} N_{ijl}}.$$

Following the estimation of E_{ij} I used a logit link function in an extended generalized linear model (GLM) structure for modeling the spatiotemporal pattern of SFR in Kansas. The GLM models incorporated stochastic spatial and temporal functions and as well as different covariate effects. I first fitted a parametric spatiotemporal model with different random terms to act as surrogate indicators of unobserved risk factors that vary over time, space or both. Subsequently, in a second model, in sequential steps, the parametric terms were replaced with nonparametric terms with different priors to assess any model improvements. Finally, in a third step, I extended the second model with different covariates to quantify unexplained spatiotemporal patterns.

The parametric model was notated as following,

$$Log(\theta_{ij}) = \alpha + u_i + v_i + \gamma_j + \psi_{ij}.$$

Where, α (intercept term) is the mean number of SFR in all counties in all years, and u_i is a random term accounting for spatially structured variation, and v_i is a random term that accounts for spatially unstructured heterogeneity in SFR distribution. In this construct, no

interaction effect was assumed to exist between u_i and v_i. I used a conditionally autoregressive (CAR) approximation for the spatially structured variation, such that $u_i \sim CAR$, and for the spatially unstructured heterogeneity I used a $v_i \sim Normal(0, \sigma_v^2)$ prior. Spatial dependence in u_i was applied by assuming a conditional autoregressive model $(CAR)(\gamma)$ with a Gaussian distribution, which implies that each u_i is conditional on its neighboring u_j with variance σ_i^2 dependent on the number of neighboring counties n_i of county i , i.e.,

$$u_i|u, j \text{ neighbor of } i \sim N\left[\frac{1}{n_i}\gamma \sum_{j=1}^{n} u_j, \frac{\sigma_i^2}{n_i}\right].$$

The γ_i term measured purely temporal and a linear time trend in the data, which assumed no temporal structure a priori. An independent, Gaussian prior with unknown variance σ_γ^2 was used for γ .

The spatiotemporal interaction effects in SFR distribution in the study region was quantified using a spatially structured term, modeled as an intrinsic Gaussian Markov random field (IGMRF) [35]. This term had a joint prior density for $\psi = (\psi_1, ..., \psi_l)$ written as

$$\pi(\psi|\sigma_\psi^2) \propto exp\left[-\frac{1}{2\sigma_\psi^2}\sum_{i\sim i'}(\psi_i - \psi_{i'})^2\right].$$

where the sum includes all pairs of adjacent counties i and i'.

A major limitation with the IGMRF model, however, is the assumption that there is a linear time trend for the observed variable (mean SFR case number) in each spatial unit (county). Therefore, in a second model I removed the parametric linearity assumption and included a nonparametric term, β_j, that quantified the overall time trend for the region (instead of linear trend per county). For the β term, a prior representing first-order random walk pattern (RW1) was specified. Higher order space-time interaction effects in the data can be quantified by specifying a δ term, with one of four types of interaction effects discussed by Knorr-Held (2000) [36], [37]. These models were named 2a, 2b, 2c, and 2d, respectively.

$$\pi(\beta|\sigma_\beta^2) \propto exp\left[-\frac{1}{2\sigma_\beta^2} \sum_{t=2}^{T} (\beta_t - 1\beta_{t-1})^2 \right].$$

The Type-II prior was notated as

$$\pi(\beta|\sigma_\beta^2) \propto exp\left[-\frac{1}{2\sigma_\beta^2} \sum_{t=3}^{T} (\beta_t - 2\beta_{t-1} + \beta_{t-2})^2 \right].$$

The Type-III prior was notated as

$$\pi(\delta|\sigma_\delta^2) \propto exp\left[-\frac{1}{2\sigma_\delta^2} \sum_{t=3}^{T} \sum_{i=1}^{I} (\delta_{it} - 2\delta_{i,t-1} + \delta_{i,t-2})^2 \right].$$

The Type-IV prior was notated as

$$\pi(\delta|\sigma_\psi^2) \propto exp\left[-\frac{1}{2\sigma_\delta^2} \sum_{t=3}^{T} \sum_{i\sim i'} \left\{ \left(\delta_{it} - 2\delta_{i,t-1} + \delta_{i,t-2} \right) - \left(\delta_{i',t-2} - 2\delta_{i',t-1} + \delta_{i't} \right) \right\}^2 \right].$$

where v, ψ, γ, and β, explained above are main effect terms. The δ term in these priors represent space-time interactions I - IV [36].

For the extended models with covariate terms, different covariates were included to the second non-parametric model in several sequential steps, starting with a model that included all covariates screened in the bivariate procedure that retained significance at a liberal ($p \leq 2$) value followed by the removal of one variable at each step. Individual covariates were retained in the model unless their removal resulted in the increase of Deviance Information Criterion (DIC) value by 5 units or more. The removed covariates did not re–enter the model and all covariates were assigned noninformative priors, $Normal(0, \sigma_\gamma^2)$.

The spatiotemporal interaction term, ψ_{ij} in the covariate models measured the trend after accounting for purely spatial ($\alpha + u_i + v_i$) and purely temporal effects (β_{ij}) in addition to the covariate effects, This reveals the variation in SFR incidence trends across counties

over time. The differential trend (the difference between overall trend and local trend) of a county i for a given year j can also be estimated from ψ_i with $\psi_i > 0$ indicating steeper than overall trend and $\psi_{ij} < 0$ indicating a less-steeper the overall trend. Counties where $\psi_{ij} = 0$ are areas where the trends are equal [38].

I applied two model selection criteria in this study, the DIC value and the logarithmic score (LS). The former is the tool of Bayesian model choice for selecting the most parsimonious model after penalizing for model complexity [39], however, recent studies have shown that DIC can be problematic when used to evaluate models that consider many random effects [40]. Therefore, in addition to DIC, I computed the LS of each model to assess predictive quality of models [41],[35], which is represented by $LS = -log(\pi_i j)$, where $\pi_i j = pr(Y_i j = y_i j | y\_ij)$ denotes the cross-validated predictive probability. A smaller LS indicates better predictive quality of a model.

All model posterior parameters were estimated using a Bayesian framework implemented using R–INLA software [42], and the median estimates from the posterior distribution and their corresponding uncertainty measures [95% Credible Intervals (CrI)] were recorded.

Chapter 3

Results

3.1 Descriptive summary

There were $n = 1129$ SFR cases reported to KDHE during January 1, 2013 to December 31, 2018 period, considered for this study. The cases were categorized into three groups viz., "probable", "suspect", and "confirmed" according to CDC guidelines provided in 2010 [4]. Of these, only 1.1% of cases met the criteria for "confirmed". Most cases were classified as "probable" (83.20%) and "suspect" (15.64%). A summary of SFR cases reported to KDHE under different summary categories (diagnosis, gender, race, and age-group) for the study period is present in Tables A.

Not all counties reported SFR cases to KDHE during the study period. Cases were distributed predominantly in counties to the east and central regions and as well as southern counties in Kansas (Fig 1). There was a large variability in the average annual number of cases aggregated at the respective counties, ranging from 25 reports in 2018 originating from residents of Labette county to mostly 1 or more cases in other counties. Case reports were filed to KDHE from 42, 30, 42, 40, 38 and 50 counties, in respective years, from 2013 through December 2018.

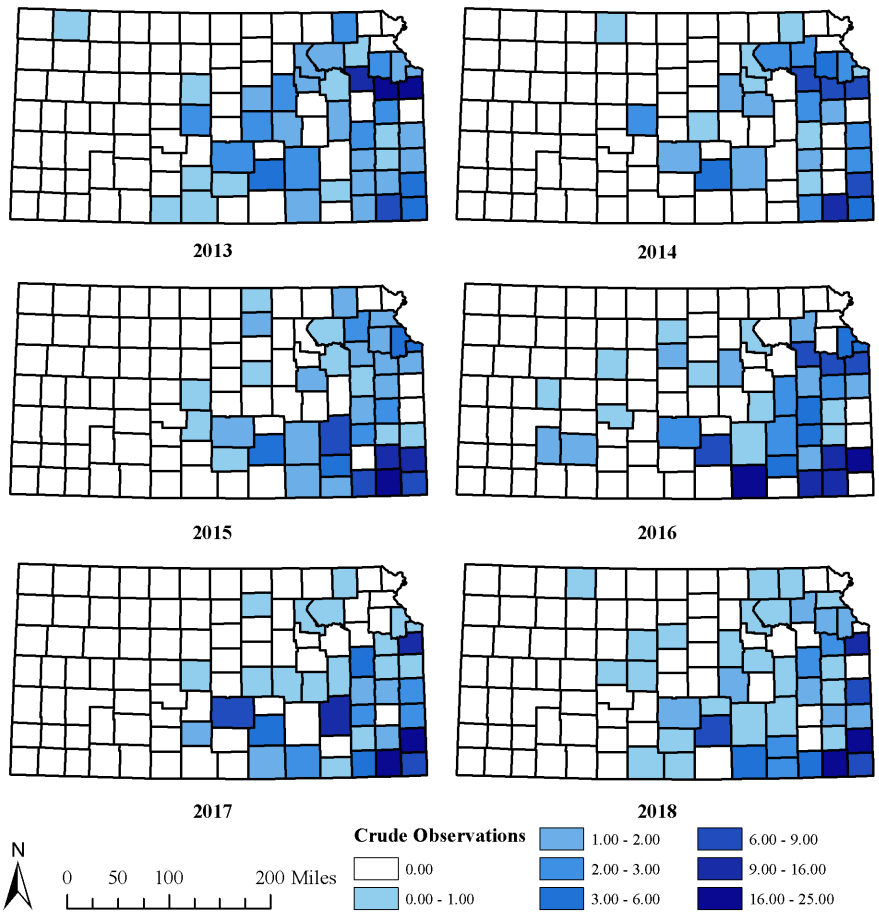

Figure 3.1: *County–level crude rate estimates of spotted fever group rickettsioses prevalence reported to the Kansas Department of Health and Environment for the study period, 2013 – 2018.*

3.2 Univariate analysis

There were 16 variables representing land cover/land use in Kansas, 10 variables representing average estimates of different climate parameters, and 43 housing and population related socio-economic variables considered as covariates in this study. Of all the covariates, four retained significance in the univariate screening with a liberal $p-value$ of 0.2, and they were considered candidate variables for Bayesian hierarchical modeling. No multicollinearity or nonlinearity among the candidate variables was observed. The covariates and the character-

istics of their univariate relationship with SFR counts in Kansas counties is present in Table 3.1

Table 3.1: *Results of univariate analysis and candidate variables ($p \leq 0.2$)*

Covariate	Estimate	S.E	p-value
Poverty status (income in the past 12 months below poverty level)	1.94	0.02	0.00
Year structure built (1980 or earlier)	0.89	0.10	0.01
Relative humidity	1.32	0.21	0.02
Normalized Difference Vegetation Index (NDVI)	0.14	0.82	0.18

3.3 Bayesian hierarchical model

Of the several progressively additive parametric and non-parametric space-time models (partial space-time models without covariates) evaluated in the study, model-2d had the lowest DIC value, indicating the presence of non-linear time trend and space-time interaction effect in the SFR cases reported to KDHE during the 2013 – 2018 period. When different covariates were added to this model (model-2d), further covariate effects were evident.

The partial space-time model with lowest DIC value (model-2d) and the best-fitting covariate model differed by 477 DIC units and also showed a reduction in the range of LS values (Table 3.2); indicating that the covariates competed with the Bayesian terms to explain additional variability in the SFR occurrence data in Kansas. Therefore, all interpretations were made based on the covariate model-4 alone. The best fitting covariate model indicated a significant spatiotemporal effect for all years, and two socioeconomic variables; income

in the past 12 months below poverty level, henceforth referred simply as 'poverty status', and year structure built (1980 or earlier), henceforth referred simply as 'older housing', and one climate variable (relative humidity) are significant variables that determined the SFR prevalence in Kansas (Table 3.3).

Table 3.2: *Model statistics for covariate models evaluating county-level spotted fever group rickettsioses prevalence in Kansas.*

Covariate	Model-3 Estimate, 95% Bayes CrI.	Model-4 Estimate, 95% Bayes CrI.
β_1	0.81 (0.44 – 0.92)	0.87 (0.48 – 0.92)
β_2	0.21 (0.11 – 0.58)	0.26 (0.18 – 0.41)
β_3	0.42 (0.17 – 0.61)	0.41 (0.19 – 0.59)
β_4	0.11 (0.01 – 0.97)	–

β_1 = poverty status, β_2 = year structure built (1980 or earlier), β_3 = relative humidity, β_4 = Normalized Difference Vegetation Index (NDVI).

The odds ratios and 95% CrI for different covariates in the models are present in Table 3.3, and they indicate that those counties with higher relative humidity, poverty status, and more number of structures built in 1980 or earlier significantly increased the odds of positive diagnosis for SFR among humans in Kansas. The county-level reported number of SFR cases per county, per year is present in Figure 1, and the posterior estimates of the final covariate model (that also included terms for space, and space-time interaction effects) is present in Figure 2.

When the space-time parameter estimates and their 95% CrI for the study period was plotted (Fig. 3), it showed a steady progression of reported SFR cases in Kansas over the years. And, the posterior probability of the space-time interaction effect, with values above 0.8 were plotted to highlight those counties with most substantial space-time interactions

Table 3.3: *Model fit and comparison criteria.*

Model	D	p^D	DIC	LS
Partial spatio–temporal models:				
Model 1 (parametric terms)	5721.32	543.87	6265.19	0.34, 0.72
Model 2 (non-parametric terms)				
a	5701.21	218.33	5919.54	0.28, 0.61
b	5105.39	321.27	5426.66	0.21, 0.48
c	4219.56	401.62	4621.18	0.35, 0.57
d	4001.27	328.52	4329.79	0.27, 0.46
Covariate models:				
Model 3	3822.12	311.38	4133.50	0.23, 0.41
Model 4	3571.37	281.21	3852.58	0.18, 0.39

D is the expected deviance, p^D is the deviance derived from the expected values of parameters, DIC is the deviance information criterion, and LS is the logarithmic score. β_1 = poverty status, β_2 = year structure built (1980 or earlier), β_3 = relative humidity, β_4 = Normalized Difference Vegetation Index (NDVI). The removal of β_3, then β_2 one at a time resulted in model DIC values of 3731.24 and 3829,59 and were therefore retained in the covariate model.

(Fig. 4). The plot revealed two clusters of counties; first in the southeastern region of Kansas, and second in the north eastern Kansas, appearing during the last three years of the study. These clusters indicate a progressively higher space-time trend, suggesting that there are common forces that lead to the prevalence pattern of SFR in each of the two clusters.

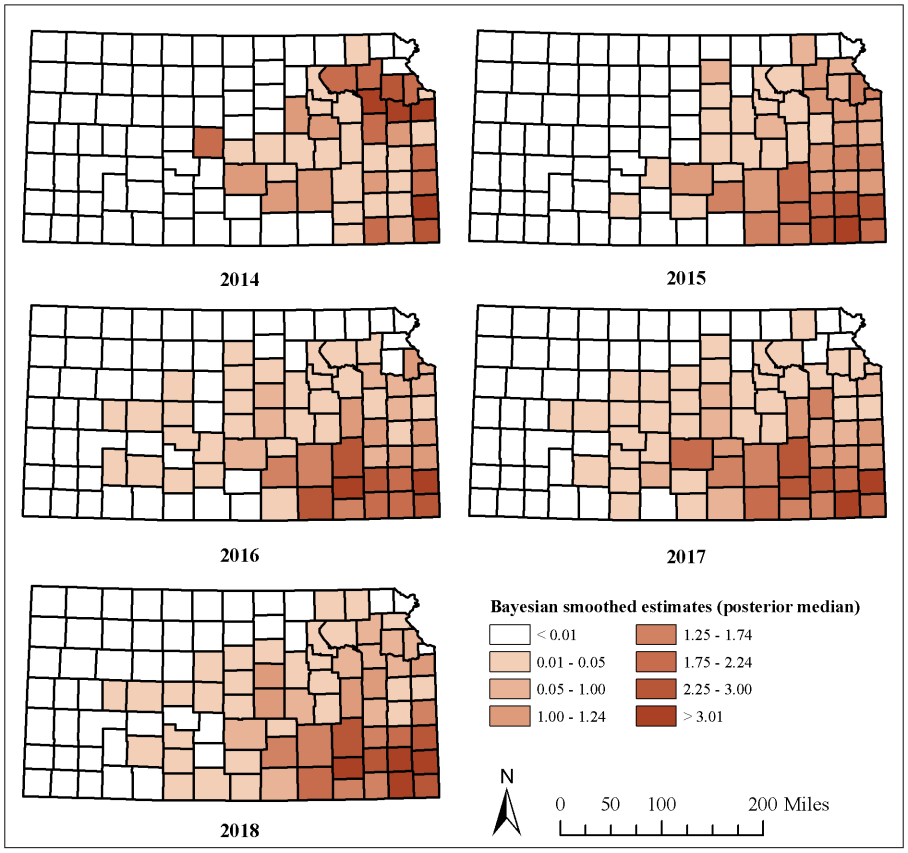

Figure 3.2: *County–specific Bayesian smoothed estimates (posterior median) of spotted fever group rickettsioses prevalence for the study period between years 2013 – 2018.*

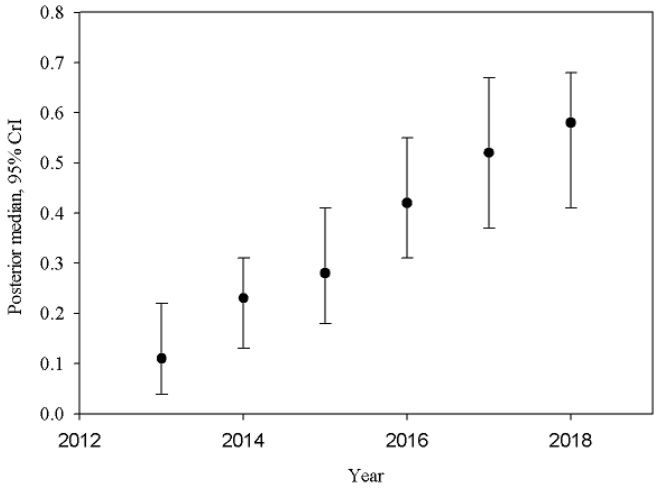

Figure 3.3: *The posterior median and 95% CrI for the overall time trend in the covariate model.*

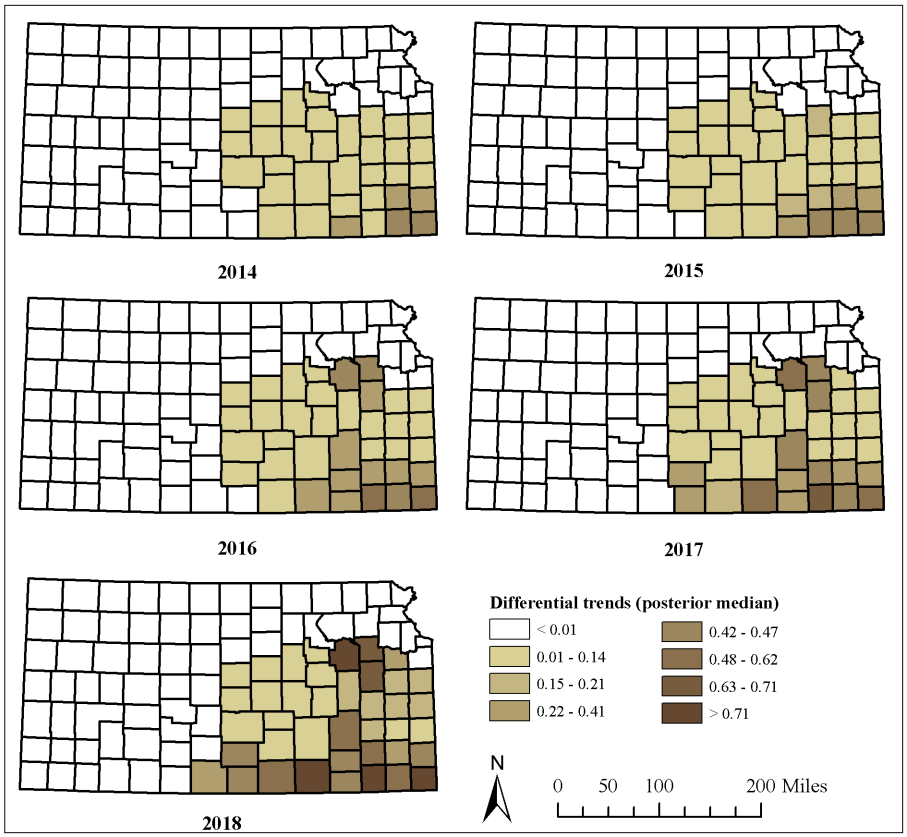

Figure 3.4: *Posterior median of county–specific differential trends. Counties with values farther from 0 indicate a higher risk for SFR.*

Chapter 4

Discussion

Cases of SFR in the U.S have steadily increased over the years [34], [43] and the trend for SFR during 2013 – 2018 period in Kansas is similar to that of the nation's. During this time, more than twice the number of reported cases to KDHE were males ($n = 763$) compared to females ($n = 366$), which is perhaps reflective of potentially higher outdoor recreational and work-related exposure to infected ticks by men more than women. A vast majority of cases reported to KDHE were White, which is expected for the state's racial demographics. And, of all the age-groups considered, people between 45 and 65 years of age were most affected by SFR followed by those above 65 years, a trend consistent with the nation's SFR and RMSF statistics [43]. This trend is likely due to increased outdoor activity for this age-group, and potentially also weakened immune system. SFR prevention and management strategies will benefit by considering these county-level demographic aspects of high-risk groups.

My primary objective in this study was to identify if any overall trend in space, time and or space-time existed for SFR cases in Kansas, and to identify high risk counties where SFR case trend may be increasing. To achieve this, I used appropriate random effect terms in a Bayesian framework based on previously published research [36], [38], [39] to first identify an overall time trend, and any space–time interaction effect, followed by the inclusion of differ-

ent covariates that would explain additional variability in the data due to space–time effect. This study has identified a steadily increasing overall space-time trend for SFR during the study period (Fig. 3), indicating that the case numbers have in fact increased year by year in Kansas, which are not readily visible when only the crude numbers are plotted on a map. Therefore, it appears that public health practices will benefit from applying model-based approaches for risk mapping and further strategic planning.

This study has identified a significant space-time progression of SFR cases in Kansas. In simple terms, this means that the risk of SFR infection among people in different Kansas counties over time is strongly related to the level of SFR prevalence occurring in their neighboring counties, in the years prior. This also means that the tick-borne SFR disease has been gradually spreading from one county to the next over the past 6 years in Kansas. This is perhaps an indication of expanding *D. variabilis* populations in Kansas along with an increase in reservoir hosts for Rickettsial pathogens. A recent modeling study showed vast areas of Kansas being highly suitable for *D. variabilis* ticks than areas where they are currently known to occur (Boorgula et al., unpublished). Other likely factor may be that other tick species are involved in the transmission of SFR pathogens that are yet to be identified in Kansas.

It is further evident from this study that the SFR prevalence in Kansas are clustered in two areas; one in the southeast and another in the northeastern region, and since these clusters are geographically isolated, they are potentially driven by different sets of influential factors identified in the study. Further evaluations at the scale of counties around these clusters will help identify which dominant socio-economic and/or environmental drivers are operational in these two clusters. Analyzing patient-level SFR data will additionally help to identify risk-factors that are perhaps masked at a county-scale. It is also possible that these two clusters represent case reports of SFR caused by different Rickettsial pathogens, but the

current SFR dataset is not adequate to verify this effect. It is important to note that the posterior Bayesian estimates calculated in this study are based on reports to counties that indicated positive diagnosis for any one of the Rickettsioses, including RMSF. It is possible that most of the cases referred to the laboratories in Kansas are likely due to RMSF; however, surveys of other SFR pathogens among ticks and wildlife hosts in Kansas is lacking and therefore cannot be confirmed.

Including covariate terms in addition to the nonparametric time and space-time terms further improved model fitness, indicating that these terms competed with the random effects to explain previously hidden variability in the SFR dataset. Poverty status and older housing, are two socio-economic variables found to be significantly associated with the SFR spatiotemporal trend in this study. In two previously published works on tick-borne disease spatial epidemiology, human monocytic ehrlichiosis [17], and RMSF [28], I have similarly detected poverty status to be a strong predictor of spatiotemporal trends. While lower socio-economic conditions have been long associated with tick-borne diseases in Europe, such as tick-borne encephalitis [44], [45], such associations have not been evaluated before in North America. The association of poverty status and the number of older homes in a county with SFR prevalence is therefore significant, and likely an indication of higher exposure of residents in these counties to tick-borne pathogens, who may work outdoors, and have less awareness towards preventing tick bites and seeking treatment.

The identification of higher relative humidity as a driver for SFR space-time trend in Kansas is consistent with other studies that have pointed out its relevance as an important limiting factor for tick survival and spatial distribution (e.g., [46], [47]). Relative humidity is also one of many climate change indices, and meteorological studies have shown that the humidity profile for the Midwestern region has gradually changed since the 1950s [48], [49] with higher atmospheric relative humidity than in the past. Such a change may have improved

the conditions for tick survival and expansion into newer territories in the region. There is heightened interest among public health epidemiologists to understand climate change implications on human health due to vector-borne diseases [25, 50, 51] and it is generally expected that disease burden due to ticks and other arthropod vectors will worsen over the time due to ongoing climate change [52–54]. This association found between SFR space-time trend and higher humidity in this study adds further evidence to support this theory.

Quantifying climate change impacts on vector-borne diseases is complicated and subject to a high degree of uncertainty [55], [56] due to the complex nature of how these diseases are maintained in the nature and the interplay of other significant factors such as social and environmental drivers in addition to climate change. However, climate change is expected to change the exposure frequency to tick borne diseases, sensitivity (the degree to which people or communities are affected, either adversely or beneficially, by climate variability and change), and adaptive-capacity (the ability of communities, institutions, or people to adjust to potential hazards, to take advantage of opportunities, or to respond to consequences) [56].

An important limitation of this study is the use of aggregated SFR data, which is subject to the modifiable area unit problem (MAUP) [22], [57], as well as the aggregation and inclusion of SFR data over a certain set of years *versus* others. Even though such aggregations are typical for epidemiological studies, it is likely for the trends and strength of associations indentified between variables in this study to change when the spatial and temporal units are altered. Therefore, the model results presented in this study are only interpretable at the scales in which they are analyzed.

4.1 Conclusions

This study has shown a steady space-time progression of county-scale SFR cases in Kansas during 2013 – 2018, strongly influenced by socio-economic and high relative humidity conditions. Although much of the eastern half of the state has reported SFR cases over the years, two clusters of counties appear to be worst-affected, needing immediate public health intervention. Risk maps are one of the most commonly used tools in public health agencies at all levels; by counties, regional, and national level agencies, to help aid with decision making and strategic planning [58–60]. The smoothed Bayesian maps of relative risk such as those produced in this study are more reliable than maps of crude numbers and/or crude rates, largely owing to the ability of Bayesian models to efficiently borrow information from neighboring areal units over time [61, 62]. Bayesian spatio-temporal models are also efficient in tackling missing information and the "low–number problem" that are common with epidemiological data [27]. For these reasons, publich health agencies may want to implement ways to utilize Bayesian approaches for disease mapping and subsequent decision making.

Bibliography

[1] D. H. Walker. *Rickettsiae*. University of Texas Medical Branch at Galveston, Galveston (TX), 1996. ISBN 0963117211.

[2] Rocky Mountain Spotted Fever (RMSF). Signs and Symptoms. https://www.cdc.gov/rmsf/info/index.html/, 2019. [Online; accessed 19-March-2019].

[3] D. H. Walker and D. B. Fishbein. Epidemiology of Rickettsial diseases. *European Journal of Epidemiology*, 7(3):237–245, 1991. ISSN 0393-2990.

[4] Rocky Mountain Spotted Fever (RMSF). https://www.cdc.gov/rmsf/info/index.html/, 2019. [Online; accessed 19-March-2019].

[5] Tickborne Diseases of the United States. https://www.cdc.gov/ticks/tickbornediseases/rmsf.html, 2019. [Online; accessed 19-March-2019].

[6] F. Dantas-Torres. Climate change, biodiversity, ticks and tick-borne diseases: The butterfly effect. *International Journal for Parasitology: Parasites and Wildlife*, 4(3): 452–461, 2015. ISSN 2213-2244.

[7] A. S. Varela-Stokes, S. H. Park, S. Kim, and S. C. Ricke. Microbial communities in North American Ixodid ticks of veterinary and medical importance. *Frontiers in Veterinary Science*, 4:179, 2017. ISSN 2297-1769.

[8] A. Estrada-Peña. Ticks as vectors: Taxonomy, biology and ecology. *Revue Scientifique et Technique*, 34(1):53–65, 2015.

[9] D. E. Sonenshine and R. M. Roe. *Biology of ticks*, volume 2. Oxford University Press, 2013. ISBN 0199744068.

[10] E. Lindgren, L. Talleklint, and T. Polfeldt. Impact of climatic change on the northern latitude limit and population density of the disease-transmitting european tick *Ixodes ricinus*. *Environmental Health Perspectives*, 108(2):119–123, 2000. ISSN 0091-6765. doi: 10.2307/3454509. URL <GotoISI>://CABI:20000505787.

[11] E. Lindgren and R. Gustafson. Tick-borne encephalitis in Sweden and climate change. *Lancet*, 358(9275):16–18, 2001. ISSN 0140-6736. doi: 10.1016/s0140-6736(00)05250-8. URL <GotoISI>://WOS:000169710900010.

[12] J. M. Medlock, K. M. Hansford, A. Bormane, M. Derdakova, A. Estrada-Peña, J. C. George, I. Golovljova, T. G. Jaenson, J. K. Jensen, P. M. Jensen, M. Kazimirova, J. A. Oteo, A. Papa, K. Pfister, O. Plantard, S. E. Randolph, A. Rizzoli, M. M. Santos-Silva, H. Sprong, L. Vial, G. Hendrickx, H. Zeller, and W. Van Bortel. Driving forces for changes in geographical distribution of *Ixodes ricinus* ticks in Europe. *Parasites & Vectors*, 6, 2013. ISSN 1756-3305. doi: 10.1186/1756-3305-6-1. URL <GotoISI>://WOS:000313908600001.

[13] R. K. Raghavan, D. G. Goodin, G. A. Hanzlicek, G. Zolnerowich, M. W. Dryden, G. A. Anderson, and R. R. Ganta. Maximum entropy-based ecological niche model and bioclimatic determinants of Lone Star Tick *Amblyomma americanum* niche. *Vector-Borne and Zoonotic Diseases*, 16(3):205–211, 2016.

[14] N. Riccardi, R. M. Antonello, R. Luzzati, J. Zajkowska, S. Di Bella, and D. R. Giacobbe. Tick-borne encephalitis in Europe: a brief update on epidemiology, diagnosis, prevention, and treatment. *European Journal of Internal Medicine*, 2019. ISSN 0953-6205.

[15] N. Hartemink, A. van Vliet, H. Sprong, F. Jacobs, I. Garcia-Martí, R. Zurita-Milla, and W. Takken. Temporal-spatial variation in questing tick activity in the Netherlands:

The effect of climatic and habitat factors. *Vector-Borne and Zoonotic Diseases*, 2019. ISSN 1530-3667.

[16] T. G. T. Jaenson, D. G. E. Jaenson, L. Eisen, E. Petersson, and E. Lindgren. Changes in the geographical distribution and abundance of the tick *Ixodes ricinus* during the past 30 years in Sweden. *Parasites & Vectors*, 5(8):(10 January 2012), 2012. ISSN 1756-3305. URL http://www.parasitesandvectors.com/content/pdf/1756-3305-5-8.pdf.

[17] R. K. Raghavan, D. Neises, D. G. Goodin, D. A. Andresen, and R. R. Ganta. Bayesian spatio-temporal analysis and geospatial risk factors of Human Monocytic Ehrlichiosis. *PloS ONE*, 9(7):e100850, 2014.

[18] R. K. Raghavan, K. Almes, D. G. Goodin, J. A. Harrington Jr, and P. W. Stackhouse Jr. Spatially heterogeneous land cover/land use and climatic risk factors of tick-borne feline cytauxzoonosis. *Vector-Borne and Zoonotic Diseases*, 14(7):486–495, 2014.

[19] R. K. Raghavan, J. Harrington Jr, G. A. Anderson, J. M. S. Hutchinson, and B. M. DeBey. Environmental, climatic, and residential neighborhood determinants of feline tularemia. *Vector-Borne and Zoonotic Diseases*, 13(7):449–456, 2013. ISSN 1530-3667.

[20] G. A. Hanzlicek, R. K. Raghavan, R. R. Ganta, and G. A. Anderson. Bayesian space-time patterns and climatic determinants of bovine anaplasmosis. *PLoS ONE*, 11(3): e0151924, 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0151924. URL http://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0151924.

[21] A. T. Peterson and R. K. Raghavan. The leading edge of the geographic distribution of *Ixodes scapularis* (Acari: Ixodidae). *Journal of Medical Entomology*, 54(5): 1103–1103, 2017. ISSN 0022-2585. doi: 10.1093/jme/tjx097. URL <GotoISI>://WOS:000409133700001.

[22] R. K. Raghavan, K. M. Brenner, J. A. Harrington Jr., J. J. Higgins, and K. R. Harkin. Spatial scale effects in environmental risk-factor modelling for diseases.

Geospatial Health, 7(2):169–182, 2013. ISSN 1827-1987. URL <GotoISI>://WOS:000323554000002.

[23] M. Monmonier. *How to lie with maps*. University of Chicago Press, 2018.

[24] A. S. Fotheringham, C. Brunsdon, and M. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons, 2003. ISBN 0470855258.

[25] J. A. Patz, T. K. Graczyk, N. Geller, and A. Y. Vittor. Effects of environmental change on emerging parasitic diseases. *International Journal for Parasitology*, 30(12-13):1395–1405, 2000. ISSN 0020-7519. doi: 10.1016/s0020-7519(00)00141-7. URL <GotoISI>://WOS:000166144900014.

[26] P. Dorny, I. K. Phiri, J. Vercruysse, S. Gabriël, A. L. Willingham Iii, J. Brandt, B. Victor, N. Speybroeck, and D. Berkvens. A Bayesian approach for estimating values for prevalence and diagnostic test characteristics of porcine cysticercosis. *International Journal for Parasitology*, 34(5):569–576, 2004. ISSN 0020-7519.

[27] A. B. Lawson. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Chapman and Hall/CRC, 2013. ISBN 146650482X.

[28] R. K. Raghavan, D. G. Goodin, D. Neises, G. A. Anderson, and R. R. Ganta. Hierarchical Bayesian spatio-temporal analysis of climatic and socio-economic determinants of Rocky Mountain spotted fever. *PLoS ONE*, 11(3):e0150180, 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0150180. URL http://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0150180.

[29] Multi-Resolution Land Characteristics Consortium. NLCD 2011 Land Cover (CONUS). https://www.mrlc.gov/data/nlcd-2011-land-cover-conus, 2019. [Online; accessed 19-March-2019].

[30] LP DAAC. Land Processing Distributed Active Archive Center. https://lpdaac.usgs.gov/, 2019. [Online; accessed 19-March-2019].

[31] NASA Prediction of Worldwide Energy Resources. https://power.larc.nasa.gov/, 2019. [Online; accessed 19-March-2019].

[32] IPUMS NHGIS. Download U.S. Census Data Tables & Mapping Files. https://www.nhgis.org/, 2019. [Online; accessed 19-March-2019].

[33] D. W. Hosmer and S. Lemeshow. *Model-Building Strategies and Methods for Logistic Regression.*, pages 91 – 142. Second edition edition, 1990.

[34] Rocky Mountain Spotted Fever (RMSF). Epidemiology and Statistics. https://www.cdc.gov/rmsf/stats/, 2019. [Online; accessed 19-March-2019].

[35] Rue H. and L. Held. *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC, 2005. ISBN 9781584884323.

[36] L. Knorr-Held. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19(17-18):2555–2567, 2000. ISSN 0277-6715. doi: 10.1002/1097-0258(20000915/30)19:17/18⟨2555::aid-sim587⟩3.0.co;2-#. URL <GotoISI>://WOS:000089275000025.

[37] B. Schroedle and L. Held. Spatio-temporal disease mapping using INLA. *Environmetrics*, 22(6):725–734, 2011. ISSN 1180-4009. doi: 10.1002/env.1065. URL <GotoISI>://WOS:000295233000004.

[38] L. Bernardinelli, D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini. Bayesian analysis of space—time variation in disease risk. *Statistics in Medicine*, 14(21-22):2433–2443, 1995. ISSN 1097-0258.

[39] D. J. Spiegelhalter, N. G. Best, B. R. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical*

Methodology, 64:583–616, 2002. ISSN 1369-7412. doi: 10.1111/1467-9868.00353. URL <GotoISI>://WOS:000179221100001.

[40] M. Plummer. Penalized loss functions for Bayesian model comparison. *Biostatistics*, 9(3):523–539, 2008. ISSN 1465-4644. doi: 10.1093/biostatistics/kxm049. URL <GotoISI>://WOS:000256977000012.

[41] T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. ISSN 0162-1459. doi: 10.1198/016214506000001437. URL <GotoISI>://WOS:000244361000032.

[42] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009.

[43] Rocky Mountain Spotted Fever (RMSF). Historical trends. https://www.cdc.gov/rmsf/stats/index.html#anchor_1531851121362, 2019. [Online; accessed 19-March-2019].

[44] S. E. Randolph. To what extent has climate change contributed to the recent epidemiology of tick-borne diseases? *Veterinary Parasitology*, 167(2-4):92–94, 2010. ISSN 0304-4017. doi: 10.1016/j.vetpar.2009.09.011. URL <GotoISI>://WOS:000274942600002.

[45] P. Stefanoff, M. Rosinska, S. Samuels, D. J. White, D. L. Morse, and S. E. Randolph. A national case-control study identifies human socio-economic status and activities as risk factors for tick-borne encephalitis in Poland. *Plos One*, 7(9), 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0045511. URL <GotoISI>://WOS:000309388400095.

[46] S. E. Rodgers, C. P. Zolnik, and T. N. Mather. Duration of exposure to suboptimal atmospheric moisture affects nymphal blacklegged tick survival. *Journal of Medical Entomology*, 44(2):372–375, 2007.

[47] J. A. Yoder, B. Z. Hedges, and J. B. Benoit. Water balance of the American dog tick, *Dermacentor variabilis*, throughout its development with comparative observations between field-collected and laboratory-reared ticks. *International Journal of Acarology*, 38(4):334–343, 2012.

[48] M. D. Schwartz. Detecting structural climate change: an air mass-based approach in the North Central United States, 1958–1992. *Annals of the Association of American Geographers*, 85(3):553–568, 1995.

[49] D. Changnon, M. Sandstrom, and C. Schaffer. Relating changes in agricultural practices to increasing dew points in extreme Chicago heat waves. *Climate Research*, 24(3):243–254, 2003.

[50] A. K. Githeko, S. W. Lindsay, U. E. Confalonieri, and J. A. Patz. Climate change and vector-borne diseases: a regional analysis. *Bulletin of the World Health Organization*, 78:1136–1147, 2000.

[51] J. N. Mills, K. L. Gage, and A. S. Khan. Potential influence of climate change on vector-borne and zoonotic diseases: a review and proposed research plan. *Environmental Health Perspectives*, 118(11):1507–1514, 2010.

[52] R. S. Ostfeld and J. L. Brunner. Climate change and Ixodes tick-borne diseases of humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370 (1665):20140051, 2015.

[53] J. S. Gray, H. Dautel, A. Estrada-Peña, O. Kahl, and E. Lindgren. Effects of climate change on ticks and tick-borne diseases in Europe. *Interdisciplinary Perspectives on Infectious Diseases*, 2009, 2009.

[54] L. Eisen. Climate change and tick-borne diseases: a research field in need of long-term empirical field studies. *International Journal of Medical Microbiology*, 298:12–18, 2008.

[55] J. Lederberg, M. A. Hamburg, M. S. Smolinski, et al. *Microbial Threats to Health: Emergence, Detection, and Response*. National Academies Press, 2003.

[56] Climate and health assessment. 5. Vector-borne Diseases. https://health2016.globalchange.gov/vectorborne-diseasesl, 2019. [Online; accessed 19-March-2019].

[57] A. S. Fotheringham and D. W. S. Wong. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, 23(7):1025–1044, 1991.

[58] K. M. Hanafiah, K. H. Jacobsen, and S. T. Wiersma. Challenges to mapping the health risk of Hepatitis A virus infection. *International Journal of Health Geographics*, 10 (57):(18 October 2011)–(18 October 2011), 2011. ISSN 1476-072X. URL <GotoISI>://CABI:20113367672.

[59] C. I. J. Nykiforuk and L. M. Flaman. Geographic information systems (GIS) for health promotion and public health: a review. *Health Promotion Practice*, 12(1):63–73, 2011. ISSN 1524-8399. doi: 10.1177/1524839909334624. URL <GotoISI>://MEDLINE:19546198.

[60] L. Bernardinelli and C. Montomoli. Empirical Bayes versus fully Bayesian-analysis of geographical variation in disease risk. *Statistics in Medicine*, 11(8):983–1007, 1992. ISSN 0277-6715. doi: 10.1002/sim.4780110802. URL <GotoISI>://WOS:A1992JB42900001.

[61] A. Mollié. Bayesian mapping of disease. *Markov Chain Monte Carlo in Practice*, 1: 359–379, 1996.

[62] A. B. Lawson. Disease map reconstruction. *Statistics in Medicine*, 20(14):2183–2204, 2001.

Appendix A

Descriptive Summaries

Table A.1: *Summary of spotted fever group rickettsioses cases reported to Kansas Department of Health and Environment during 2013 – 2018 by case category*

Year	Spotted Fever group Rickettsioses cases			
	Probable	Suspect	Confirmed	$\sum_{group}$
2013	166	0	2	168
2014	110	0	0	110
2015	146	34	2	182
2016	128	32	2	162
2017	216	60	2	278
2018	175	51	5	231
$\sum_{year}$	941	177	13	1129

Table A.2: *Summary of spotted fever group rickettsioses cases reported to Kansas Department of Health and Environment during 2013 – 2018 by gender*

Year	Male	Female	$\sum_{gender}$
2013	119	49	168
2014	78	32	110
2015	125	55	180
2016	111	51	162
2017	166	112	278
2018	164	67	231
$\sum_{year}$	763	366	1129

Table A.3: *Summary of spotted fever group rickettsioses cases reported to Kansas Department of Health and Environment during 2013 – 2018 by race*

Year	American Indian / Alaskan Native	Asian	Black / African American	White	Unknown	$\sum_{race}$
2013	1	1	1	154	11	168
2014	2	2	1	103	3	110
2015	2	0	3	139	36	180
2016	1	0	2	128	31	162
2017	2	1	3	212	60	278
2018	1	2	1	199	28	231
$\sum_{year}$	9	5	11	935	169	1129

Table A.4: *Summary of spotted fever group rickettsioses cases reported to Kansas Department of Health and Environment during 2013 – 2018 by age-group*

Year	< 5	$5 - 19$	$20 - 45$	$46 - 65$	> 65	$\sum_{age-group}$
2013	1	1	1	154	11	168
2014	2	2	1	103	3	110
2015	2	0	3	139	36	180
2016	1	0	2	128	31	162
2017	2	1	3	212	60	278
2018	1	2	1	199	28	231
$\sum_{year}$	9	5	11	935	169	1129