

Statistical methods for modeling infectious disease outbreaks using disease  
surveillance data

by

Nelson B. Walker

B.S., Brigham Young University, 2014

M.S., Kansas State University, 2018

---

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the  
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics  
College of Arts and Sciences

KANSAS STATE UNIVERSITY  
Manhattan, Kansas

2021

# Abstract

Disease surveillance data are commonly used by epidemiologists, veterinary and plant pathologists, and wildlife and plant ecologists to identify, understand, mitigate, and prevent the spread of infectious disease. We develop three statistical methods that may be applied to spatio-temporal disease surveillance data to understand different aspects of an infectious disease outbreak.

First, we develop a method that provides individual-level inference on spatial covariates despite using several types of spatially aggregated binary disease surveillance data. Our method provides individual-level inference on spatial covariates by applying a series of transformations, including a change of support, to a bivariate point process model. The series of transformations preserves the convenient interpretation of desirable binary regression models that are commonly applied to individual-level disease surveillance data. Using a simulation experiment, we compare the performance of the proposed method under varying types of spatial aggregation against the performance of standard approaches using the original individual-level data. We illustrate our method by modeling individual-level probability of infection using a disease surveillance data set that has been aggregated to protect several at-risk or threatened species of bats in the northeastern U.S.

Second, we develop a staged approximate Bayesian model averaging (SABMA) method to estimate the spatio-temporal origins of an epidemic. Specifically, we estimate the number, locations, and times that a pathogen was introduced into a population using spatio-temporal binary disease surveillance data. We employ an ensemble of simple ecological diffusion processes to model the spatio-temporal spread of the pathogen from multiple locations. We study the statistical properties of the SABMA method, in terms of credible interval coverage for parameters and out-of-sample prediction performance, using a simulation experiment.

We then apply our SABMA method to two sets of binary disease surveillance data in white-tailed deer (*Odocoileus virginianus*); the first in the lower peninsula of Michigan in the U.S., and the second in southern Wisconsin and northern Illinois in the U.S.

Third, we develop a Bayesian hierarchical mixture of ecological diffusion models (BHMEDM) that provides inference on the number, locations, and times of pathogen introduction during an epidemic, using spatio-temporal binary disease surveillance data. Our model incorporates a mixture of ecological diffusion processes that account for both the growth and diffusion of the pathogen. As part of the hierarchical framework, we obtain inference on the spatio-temporal process that produced the pathogen introductions, and predict where new pathogen introductions are likely to occur in the future. We demonstrate the BHMEDM using binary disease surveillance data in white-tailed deer from southern Wisconsin and northern Illinois in the U.S.