# Bayesian Modeling of Partially Observed Epidemic Count Data

Jonathan Fintzi

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Vladimir Minin, Chair

Jon Wakefield, Chair

M. Elizabeth Halloran

James Hughes

Program Authorized to Offer Degree:
Biostatistics

University of Washington

**Abstract**

Bayesian Modeling of Partially Observed Epidemic Count Data

Jonathan Fintzi

Co-Chairs of the Supervisory Committee:
Co-chair Vladimir Minin

Co-chair Jon Wakefield

An incredible abstract with all the best words will appear here.

# TABLE OF CONTENTS

Page

List of Figures . . . . . iii

List of Tables . . . . . iv

Glossary . . . . . v

Chapter 1: Introduction and data setting . . . . . 1
1.1 Motivating examples . . . . . 1
1.1.1 Influenza in a British boarding school . . . . . 1
1.1.2 Ebola in West Africa . . . . . 1
1.1.3 Pandemic A(H1N1) influenza in Finland . . . . . 1
1.2 Organization of this dissertation . . . . . 1

Chapter 2: Background . . . . . 2
2.1 Models for the spread of infectious disease . . . . . 2
2.1.1 Deterministic representations . . . . . 2
2.1.2 Stochastic representations . . . . . 2
2.1.3 Large–population approximations . . . . . 2
2.2 Computational approaches to fitting stochastic epidemic models . . . . . 2
2.3 Bayesian computation . . . . . 2
2.3.1 Markov chain Monte Carlo . . . . . 2
2.3.2 Bayesian data augmentation . . . . . 2

Chapter 3: Agent–Based Data Augmentation for Fitting Stochastic Epidemic Models to Prevalence Data . . . . . 3
3.1 Overview . . . . . 3
3.2 The data augmentation algorithm for an SIR model . . . . . 3
3.3 Generalizing the algorithm to other models . . . . . 3

3.3.1 Data augmentation for SEIR dynamics . . . 3
3.3.2 Data augmentation for SIRS dynamics . . . 3
3.3.3 Data augmentation for arbitrary dynamics . . . 3
3.4 Simulation results . . . 3
3.5 Example: Influenza in a British boarding school . . . 3
3.6 Discussion . . . 3

Chapter 4: Approximate Inference for Stochastic Epidemic Models of Outbreaks in Large Populations . . . 4
4.1 Overview . . . 4
4.2 Fitting . . . 5

Chapter 5: Dynamic Transmission Modeling of Pandemic A(H1N1) Influenza in Finland . . . 6

Chapter 6: Discussion and Future Work . . . 7

Bibliography . . . 8

Appendix A: Appendix A . . . 9

# LIST OF FIGURES

Figure Number Page

# LIST OF TABLES

Table Number Page

# GLOSSARY

AFSS: Automated factor slice sampler.

CTMC: Continuous–time Markov chain.

DA: Data augmentation.

ELIPTSS: Elliptical slice sampling.

ESS: Effective sample size.

LNA: Linear noise approximation.

MJP: Markov jump process.

SEM: Stochastic epidemic model.

# ACKNOWLEDGMENTS

Very grateful to many people.

# DEDICATION

Dedication to important people.

# Chapter 1

# INTRODUCTION AND DATA SETTING

## *1.1 Motivating examples*

### *1.1.1 Influenza in a British boarding school*

### *1.1.2 Ebola in West Africa*

### *1.1.3 Pandemic A(H1N1) influenza in Finland*

## *1.2 Organization of this dissertation*

# Chapter 2

# BACKGROUND

## *2.1 Models for the spread of infectious disease*

### *2.1.1 Deterministic representations*

### *2.1.2 Stochastic representations*

*Agent–based models*

*Population–level models*

### *2.1.3 Large–population approximations*

*Diffusion approximations of Markov jump processes*

*Linear noise approximation*

## *2.2 Computational approaches to fitting stochastic epidemic models*

## *2.3 Bayesian computation*

### *2.3.1 Markov chain Monte Carlo*

### *2.3.2 Bayesian data augmentation*

# Chapter 3

# AGENT–BASED DATA AUGMENTATION FOR FITTING STOCHASTIC EPIDEMIC MODELS TO PREVALENCE DATA

***3.1 Overview***

***3.2 The data augmentation algorithm for an SIR model***

***3.3 Generalizing the algorithm to other models***

*3.3.1 Data augmentation for SEIR dynamics*

*3.3.2 Data augmentation for SIRS dynamics*

*3.3.3 Data augmentation for arbitrary dynamics*

***3.4 Simulation results***

***3.5 Example: Influenza in a British boarding school***

***3.6 Discussion***

# Chapter 4

# APPROXIMATE INFERENCE FOR STOCHASTIC EPIDEMIC MODELS OF OUTBREAKS IN LARGE POPULATIONS

## *4.1 Overview*

Surveillance and outbreak response systems often report incidence counts of new cases detected in each inter–observation time interval. Analyzing this type of time series data is challenging since we must overcome many of the same challenges that we face in modeling the transmission dynamics of infectious diseases in small population settings with prevalence data — discrete snapshots of a continuously evolving epidemic process, detecting a fraction of the new cases, and often directly observing only one aspect of the disease process. Furthermore, our task is made more difficult by the additional computational burden that results from repeated evaluation of CTMC likelihoods; the products of exponential waiting time distributions consist of polynomially increasing numbers of terms, and agent–based data augmentation MCMC algorithms become unwieldy as the numbers of subject–path proposals required to meaningfully perturb the CTMC likelihood get large [3].

In this chapter, we show how the LNA of Section 2.1.3 can be adapted to obtain approximate inference for SEMs fit to epidemic count data in large populations. Our contributions are threefold: First, we demonstrate how the SEM dynamics should be reparameterized so that the LNA can be used to approximate transition densities of the counting processes for disease state transition events. Second, we fold the LNA into a Bayesian data augmentation framework in which latent LNA paths are sampled using the elliptical slice sampling (EliptSS) algorithm of [6]. This provides us with general machinery for jointly updating the latent paths while absolving us of the *de facto* modeling choice that the data be Gaussian in order to efficiently perform inference as in [5, 2], or the need to use particle filter methods

for non–Gaussian emission distributions as in [4]. Finally, we introduce a non–centered parameterization for the latent LNA process that massively improves the efficiency of our DA MCMC framework and makes it tractable for fitting complex models.

Chapter 5

# DYNAMIC TRANSMISSION MODELING OF PANDEMIC A(H1N1) INFLUENZA IN FINLAND

# Chapter 6

# DISCUSSION AND FUTURE WORK

# BIBLIOGRAPHY

[1] L.J.S. Allen. An introduction to stochastic epidemic models. In *Mathematical Epidemiology*, pages 81–130. Springer, New York, 2008.

[2] P. Fearnhead, V. Giagos, and C. Sherlock. Inference for reaction networks using the linear noise approximation. *Biometrics*, 70:457–466, 2014.

[3] J. Fintzi, X. Cui, J. Wakefield, and V.N. Minin. Efficient data augmentation for fitting stochastic epidemic models to prevalence data. *Journal of Computational and Graphical Statistics*, 26:918–929, 2017.

[4] A. Golightly, D.A. Henderson, and C. Sherlock. Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Statistics and Computing*, 25(5):1039–1055, 2015.

[5] M. Komorowski, B. Finkenstädt, C.V. Harper, and D.A. Rand. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics*, 10:343, 2009.

[6] I. Murray, R.P. Adams, and D.J.C. MacKay. Elliptical slice sampling. *JMLR: W&CP*, 9:541–548, 2010.

[7] P. Neal and G.O. Roberts. A case study in non-centering for data augmentation: stochastic epidemics. *Statistics and Computing*, 15:315–327, 2005.

[8] P.D. O'Neill. Introduction and snapshot review: relating infectious disease transmission models to data. *Statistics in Medicine*, 29:2069–2077, 2010.

Appendix A

# APPENDIX A