# Thesis

Sang Woo Park

October 29, 2018

## 1 Introduction

Modeling ecological data is challenging.

TSIR.

$$\begin{aligned} S_{t+1} &= S_t + B_t - I_{t+1} \\ I_{t+1} &= \beta_t I_t^\alpha S_t / N_t \end{aligned} \tag{1}$$

## 2 Susceptible reconstruction

Recall that the susceptible population dynamics can be written as follows:

$$S_{t+1} = S_t + B_t - I_{t+1}. \tag{2}$$

Rewriting incidence as product of observed cases, $C_{t+1}$, divided by the reporting rate, $\rho_{t+1}$, we obtain the following:

$$S_{t+1} = S_t + B_t - \frac{C_{t+1}}{\rho_{t+1}}. \tag{3}$$

Iteratively solving the equation yields the following relation:

$$S_{t+1} = S_1 + \sum_{i=1}^t B_i - \sum_{i=1}^t \frac{C_{i+1}}{\rho_{i+1}}. \tag{4}$$

Representing susceptible dynamics as a deviation from the mean number of susceptibles $S_t = \bar{S} + Z_t$, Finkenstadt and Grenfell (2000) showed dynamics of $Z_t$ and $\rho_t$ can be recovered from the data by fitting a smooth curve to cumulative birth as a function of cumulative cumulative cases. Dalziel et al. (2016) then claimed that the result can be generalized by representing susceptible dynamics as a deviation from moving average, $S_t = \sigma N_t + W_t$, where $\sigma$ is proportion susceptible:

$$\sum_{i=1}^t B_i = \sum_{i=1}^t \frac{C_{i+1}}{\rho_{i+1}} + W_{t+1} - W_1 + \sigma(N_{t+1} - N_1). \tag{5}$$

However, they did not provide sufficient theoretical justification; here, we analyze the claimed generalization.

## Constant reporting rate

First, consider the case when the reporting is contant ($\rho_i = \rho$ for all $\rho \in \mathbb{N}$), equation 5 simplifies to

$$\sum_{i=1}^{t} B_i = \frac{1}{\rho} \sum_{i=1}^{t} C_{i+1} + W_{t+1} - W_1 + \sigma(N_{t+1} - N_1). \tag{6}$$

Writing $X_{t+1} = \sum_{i=1}^{t} C_{i+1}$ and $Y_{t+1} = \sum_{i=1}^{t} B_i$, we obtain a familiar form:

$$Y_{t+1} = \frac{1}{\rho} X_{t+1} + W_{t+1} - W_1 + \sigma(N_{t+1} - N_1).,\tag{7}$$

When population size stays constant ($N_t = N_1$ for all $t \in \mathbb{N}$), the equation simplifies to

$$Y_{t+1} = \frac{1}{\rho} X_{t+1} + W_{t+1} - W_1,\tag{8}$$

which suggests that linear regression can be used to estimate the reporting rate and susceptible dynamics (Finkenstadt and Grenfell, 2000). In general, one can write

$$N_{t+1} = N_t + B_t - D_t,$$

where $D_t$ is the number of individuals that died between time $t$ and $t+1$. Then, we have

$$Y_{t+1} = \frac{1}{\rho} X_{t+1} + W_{t+1} - W_1 + \sigma\left(\sum_{i=1}^{t} B_i - \sum_{i=1}^{t} D_i\right).\tag{9}$$

Writing $M_t = \sum_{i=1}^{t} D_i$, we get

$$(1-\sigma) Y_{t+1} = \frac{1}{\rho} X_{t+1} + W_{t+1} - W_1 + \sigma M_t,\tag{10}$$

which suggests that a linear regression is likely to overestimate susceptible dynamics and underestimate reporting rate.

## Non-constant reporting rate

Consider the following equation:

$$\sum_{i=1}^{t} B_i = \sum_{i=1}^{t} \frac{C_{i+1}}{\rho_{i+1}} + W_{t+1} - W_1 + \sigma(N_{t+1} - N_1).\tag{11}$$

Define $\hat{\rho}$ such that $1/\hat{\rho} = E[1/\rho_t]$. Then,

$$Y_{t+1} = \frac{1}{\hat{\rho}} X_{t+1} + \sum_{i=1}^{t} \left(\frac{1}{\rho_{i+1}} - \frac{1}{\hat{\rho}}\right) C_{i+1} + W_{t+1} - W_1 + \sigma(N_{t+1} - N_1).\tag{12}$$

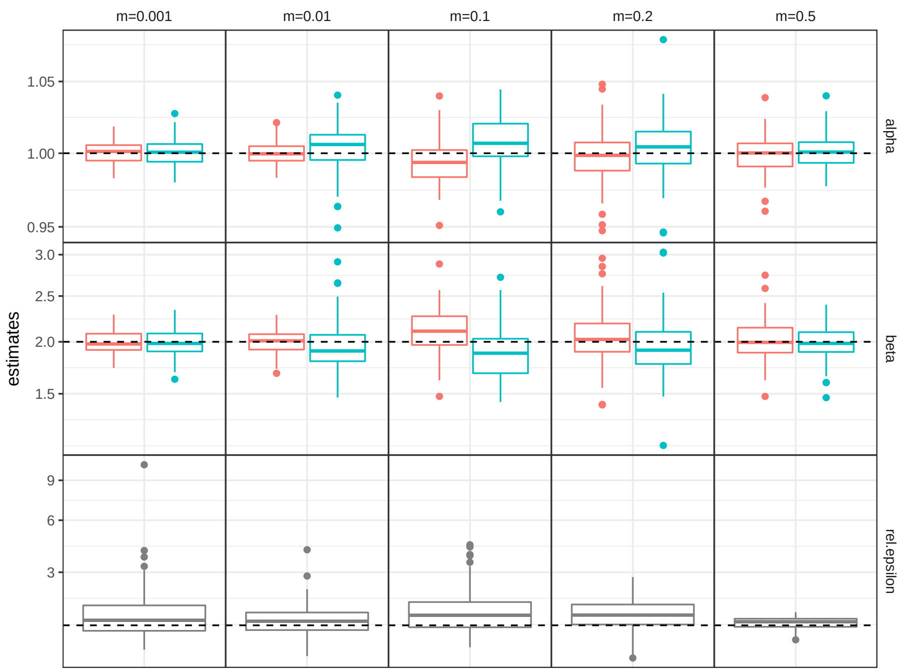


Figure 1: **SIR model** TODO: caption

Let

$$R_{t+1} = \sum_{i=1}^{t} \left( \frac{1}{\rho_{i+1}} - \frac{1}{\hat{\rho}} \right) C_{i+1}.$$

Then, we obtain an expression that generalizes the result derived by Finkenstadt and Grenfell (2000) by accounting for changes in population sizes.

$$\begin{aligned} Y_{t+1} &= \frac{1}{\hat{\rho}} X_{t+1} + \left( \frac{1}{\rho_{t+1}} - \frac{1}{\hat{\rho}} \right) C_{t+1} + R_t + W_{t+1} - W_1 + \sigma(N_{t+1} - N_1) \\ &= \frac{1}{\rho_{t+1}} X_{t+1} + \left( \frac{1}{\hat{\rho}} - \frac{1}{\rho_{t+1}} \right) X_t + R_t + W_{t+1} - W_1 + \sigma(N_{t+1} - N_1) \end{aligned} \tag{13}$$

# 3 Spatial coupling

# 4 Case study: measles

# References

Dalziel, B. D., O. N. Bjørnstad, W. G. van Panhuis, D. S. Burke, C. J. E. Metcalf, and B. T. Grenfell (2016). Persistent chaos of measles epidemics in the prevaccination united states caused by a small change in seasonal transmission patterns. *PLoS computational biology 12*(2), e1004655.

Finkenstadt, B. F. and B. T. Grenfell (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 49*(2), 187–205.