Early warning signals anticipate emergence in empirically-based simulations of measles dyanmics

Andrew T. Tredennick[1,2*], Pejman Rohani[1,2], Eamon O'Dea[1,2], Tobias Brett[1,2], & John M. Drake[1,2]

[1] *Odum School of Ecology, University of Georgia, Athens, GA, USA*
[2] *Center for the Ecology of Infectious Diseases, University of Georgia, Athens, GA, USA*

Abstract

Forecasting the trajectory of infectious disease outbreaks over time is a fundamental challenge facing society.

Introduction

Theory shows that epidemic transitions can be anticipated by trends in the statistical properties of disease time series (AERO papers). The existence of statistical trends in the data that precede critical transitions, so-called 'early warning signals' (EWS), imply that we may be able to anticipate disease emergence and outbreaks. The end goal is a model-independent detection system, where statistical properties of disease surveilence data can trigger warnings of impending outbreaks without the need to fit mechanistic models of disease transmission (Han and Drake 2017).

However, there is currently a gap between the theoretical work, which has relied on knowing the underlying disease dynamics, and the eventual goal of applying EWS in real-world situations where the underlying disease dynamics may be unknown. Theoretical development of EWS has focused on anticipating when the population becomes supercritical, when $\mathscr{R}_0 > 1$, after which an outbreak is inevitable, perhaps with some bifurcation delay (Dibble et al. 2016). Knowing the value of $\mathscr{R}_0$ through time makes it possible to test the accuracy of EWS that are estimated from state variables alone. Empirical application of EWS does not require knowing the value of $\mathscr{R}_0$ through time, meaning that "tests" require making assumptions about when critical transitions occur. Whether EWS track and/or anticipate underlying dynamics of real disease time series remains unknown, and is a critical knowledge gap that must be filled before EWS can confidently be deployed.

To fill this gap we will fit a mechanistic model to incidence data of measles in Niger to estimate the temporal epidemiology of the disease, yielding the very same parameters that are known in data-free modeling studies. In particular, we are interested in the correlation between EWS and the time-varying repreoductive ratio, known as the effective reprodutive ratio ($\mathscr{R}_E$). If EWS and $\mathscr{R}_E$ are significantly and positively correlated, then we have empirical evidence that EWS are applicable in real-world settings. If EWS and $\mathscr{R}_E$ are negatively correlated or not significantly posivitely correlated, then we have evidence that EWS may not be applicable in certain settings.

*Corresponding author: atredenn@gmail.com

Materials and Methods

Data

We used weekly measles case report data from four Nigerien cities, collected over an 11 year period (1995-2005) (Figure 1). These data are ideal for stress testing EWS because each city has different population sizes, has different dynamics in terms of size of outbreaks and length of inter-epidemic periods, and each time series has different amounts of demographic stochasticity due to differences in population size. The data come from [*somewhere/someone*], and used here with permission from [*somewhere/someone*].

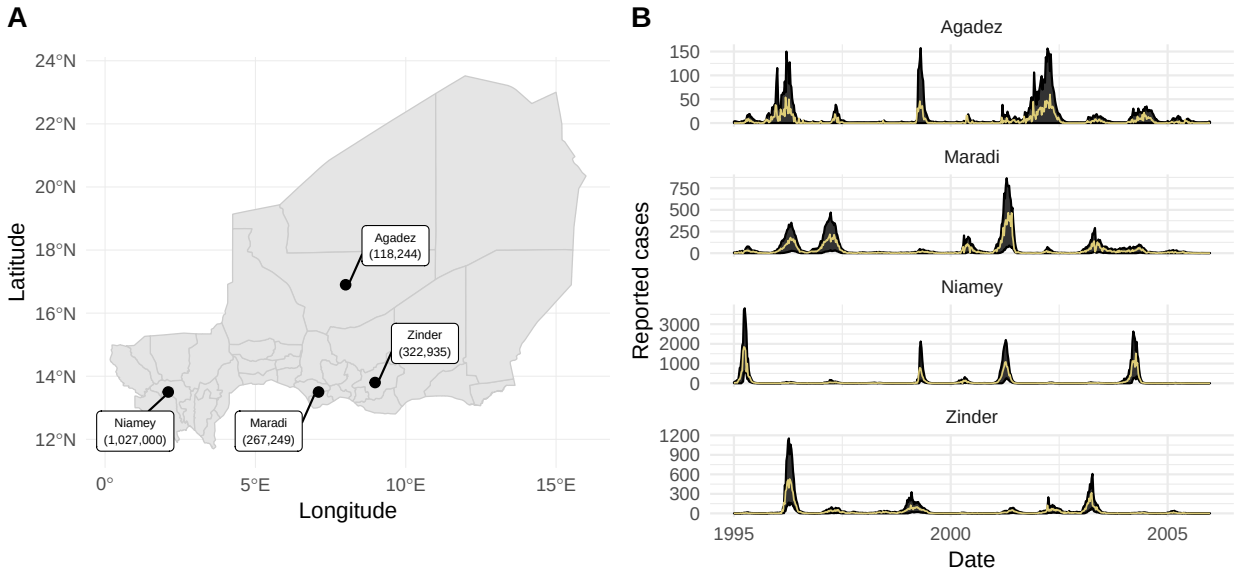

Figure 1: (A) Locations and population sizes (in parantheses) of our four focal cities in Niger. (B) Time series of reported cases (yellow solid lines) and the 95% prediction intervals for one-step-ahead forecasts from our fitted SEIR models for each city.

Stochastic *SEIR* model

The model is a discrete-time approximation of a continuous-time SEIR model with limited demography, specified as a set of difference equations,

$$S_{t+dt} = n_{S,t} - n_{E,t} \tag{1}$$
$$E_{t+dt} = n_{E,t} - n_{I,t} \tag{2}$$
$$I_{t+dt} = n_{I,t} + n_{O,t} - n_{R,t}, \tag{3}$$

where $\mathbf{n}_t$ are random variables representing the number of individuals transitioning into or out of each class at each timestep $t \to t+dt$. n_S is the number of births, n_E is the number of newly infected individuals that have the disease but are not infectious, n_I is the number of newly infectious indiviuals, n_O is the number of imported infections, and n_R is the number of newly recovered individuals who are no longer infectious and have life-long immunity. The stochastic random variables are specified as follows:

$$n_{S,t} \sim \text{Poisson}(\mu_t N_t \times dt) \tag{4}$$
$$n_{E,t} \sim \text{Binomial}(\lambda_{E,t}, S_t) \tag{5}$$
$$n_{I,t} \sim \text{Binomial}(\lambda_{I,t}, E_t) \tag{6}$$
$$n_{O,t} \sim \text{Poisson}(\psi \times dt) \tag{7}$$
$$n_{R,t} \sim \text{Binomial}(\lambda_{R,t}, I_t), \tag{8}$$

where μ_t is the birth rate at time t, ψ is the rate of imported infections, and λ_E, λ_I, and λ_R are the probabilities of exposure, becoming infectious, and recovery, respectively. These probabilities reflect the processes of tranmission, transition from the latent period to the infectious period, and recovery, which we model as:

$$\lambda_{E,t} = 1 - e^{-\frac{\beta_t I_t dt}{N_t}} \tag{9}$$
$$\lambda_{I,t} = 1 - e^{-\eta E_t dt} \tag{10}$$
$$\lambda_{R,t} = 1 - e^{-\gamma I_t dt}, \tag{11}$$

where β_t is time-varying rate of transmission, η is time-invariant rate from the exposed class to the infectious class, and γ is time-invariant recovery rate. We model rate of transmission as:

$$\beta_t = \beta \left(1 + \sum_{i=1}^{6} q_i \xi_{i_t} \right) \Gamma_t. \tag{12}$$

β is the mean transmission rate, ψ accounts for measles infections from external sources that are not part of the local dynamics, and the term $\sum_{i=1}^{6} q_i \xi_{i_t}$ is a B-spline to model seasonality in transmission. The B-spline bases (ξ_{i_t}) are periodic with a 1 year period. The transmission rate (β_t) is also subject to stochastic process noise at each time step, Γ_t, which we model as a gamma-distributed white (temporally uncorrelated) noise with mean 1 and variance σ^2 (Bretó and Ionides 2011).

We do not include a death process in the model because the rate of infection is much faster than the rate of death. Excluding deaths means we can avoid making further assumptions about demographic rates – we are already making assumptions about birth rates (e.g., the rate is the same across cities, but with city-specific population size). We model demographic stochasticity in births and imported infections by drawing time-specific values from Poisson distributions. Transitions in the model are shown in Table 1. In this model, the effective reproductive ratio at time t is: $\mathscr{R}_E(t) = \frac{\beta_t}{\gamma} \frac{S_t}{N_t}$.

We assume observed case reports ($\mathbf{y}$) are drawn from a Negative Binomial distribution subject to a constant reporting fraction (ρ) and dispersion parameter τ,

$$y_t \sim \text{Negative Binomial}\left(\rho I_t, \tau\right). \tag{13}$$

Table 1: Transitions in the SEIR model. We show the determinstic transmission rate for clarity, but our model uses the stochastic tranmission rate.

Transition	$(\Delta S, \Delta E, \Delta I)$	Propensity
birth	$(1,0,0)$	$N_t \mu_t$
transmission (deterministic)	$(-1,1,0)$	$SI\beta_t/N_t$
transmission (stochastic)	$(-k,k,0)$	$\frac{S}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j+1}\tau_{\mathrm{f}}^{-1}\ln(1+(\beta_t I/N_t))\tau_{\mathrm{f}}(S-j)$
symptomatic (infectious)	$(0,-1,1)$	$E\eta$
imported infections	$(0,0,1)$	ψ_t
recovery	$(0,0,-1)$	$I\gamma$

Model fitting and inference

We fit the SEIR model to time series of case reports from each of our focal cities using Maximization by Iterated particle Filtering (MIF) in the R package `pomp`. We estimated 14 parameters for each city: six seasonal transmission parameters (q_i), mean transmission rate (β), three initial conditions $\left(S_{(t=0)}, E_{(t=0)}, I_{(t=0)}\right)$, the number of imported infections (ψ), reporting rate (ρ), one parameter accounting for process noise, and one parameter accounting for measurement noise. To ensure identifiability, and to make the model easier to fit, we assumed the infectious period was fixed at $1/\eta = 8$ days and the recovery period was fixed at $1/\gamma = 5$ days.

Model simulations

Early warning signals

We considered ten candidate early warning signals (EWS; Table 3). EWS were calculated using the function `spaero::get_stats()` (O'Dea 201x) in R (R Core Team 201x).

Table 2: List of candidate early warning signals and their estimating equations. Note that b denotes the bandwidth. See Brett et al. (2018) for details.

EWS	Estimator	Theoretical Correlation with $\mathscr{R}_E(t)$
Mean	$\mu_t = \sum_{s=t-(b-1)\delta}^{t+(b-1)\delta} \frac{X_s}{2b-1}$	Positive
Variance	$\sigma_t^2 = \sum_{s=t-(b-1)\delta}^{t+(b-1)\delta} \frac{(X_s-\mu_s)^2}{2b-1}$	Positive
Coefficient of variation	$CV_t = \frac{\sigma_t}{\mu_t}$	Null
Index of dispersion	$ID_t = \frac{\sigma_t^2}{\mu_t}$	Positive
Skewness	$S_t = \frac{1}{\sigma_t^3}\sum_{s=t-(b-1)\delta}^{t+(b-1)\delta} \frac{(X_s-\mu_s)^3}{2b-1}$	Positive
Kurtosis	$K_t = \frac{1}{\sigma_t^4}\sum_{s=t-(b-1)\delta}^{t+(b-1)\delta} \frac{(X_s-\mu_s)^4}{2b-1}$	Positive
Autocovariance	$\mathrm{ACov}_t = \sum_{s=t-(b-1)\delta}^{t+(b-1)\delta} \frac{(X_s-\mu_s)(X_{s-\delta}-\mu_{s-\delta})}{2b-1}$	Positive

EWS	Estimator	Theoretical Correlation with $\mathscr{R}_E(t)$
Autocorrelation	$\mathrm{AC}_t = \frac{\mathrm{ACov}_t}{\sigma_t \sigma_{t-\delta}}$	Positive
Decay time	$\overline{\tau}_t = -\delta / \ln[\mathrm{AC}_t(\delta)]$	Positive
First differenced variance	$\Delta\sigma_t^2 = \sigma_t^2 - \sigma_{t-\delta}^2$	Positive

Results

The fitted models were able to adequately reproduce observed dynamics (Figure 1), with R^2s ranging from 0.54 to 0.89 (Figure 2).

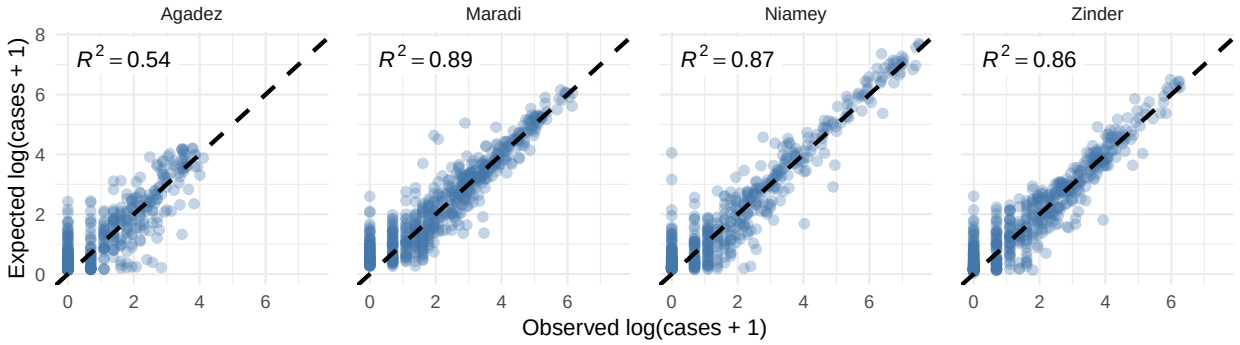

Figure 2: Comparison of in-sample model predictions and observations for each city. Expected cases are one-step-ahead predictions from the fitted models. The dashed line shows 1:1.

Stochastic simulations from sub-critical ($\mathscr{R}_{E(t)} \ll 0$) to near-critical dynamics ($\mathscr{R}_{E(t)} \approx 1$) differed among the four cities (Figure 3).

Acknowledgments

This research was funded by the National Institute of General Medical Sciences of the National Institutes of Health (Award Number U01GM110744). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This work was done on the Olympus High Performance Compute Cluster located at the Pittsburgh Supercomputing Center at Carnegie Mellon University, which is supported by National Institute of General Medical Sciences Modeling Infectious Disease Agent Study (MIDAS) Informatics Services Group grant 1U24GM110707.

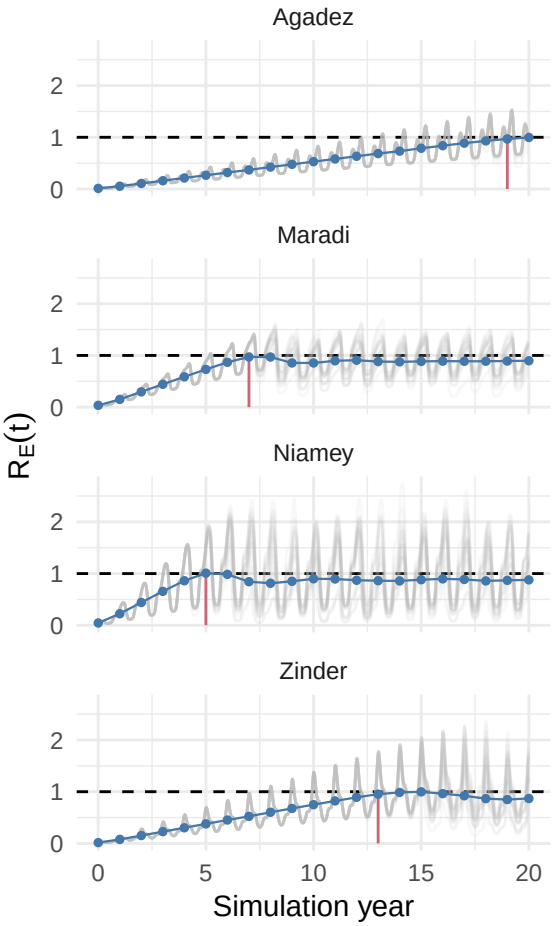

Figure 3: Yearly average $\mathscr{R}_{E(t)}$ (blue lines) across 500 models simulations at the MLE parameters and 20 representations simulations (grey lines) for each city. The horizontal dashed line shows where $\mathscr{R}_{E(t)} = 1$. The vertical solid red lines show the time point at which yearly average $\mathscr{R}_{E(t)} \geq 1$. The time periods before the red line for each city were used for testing early warning signals.

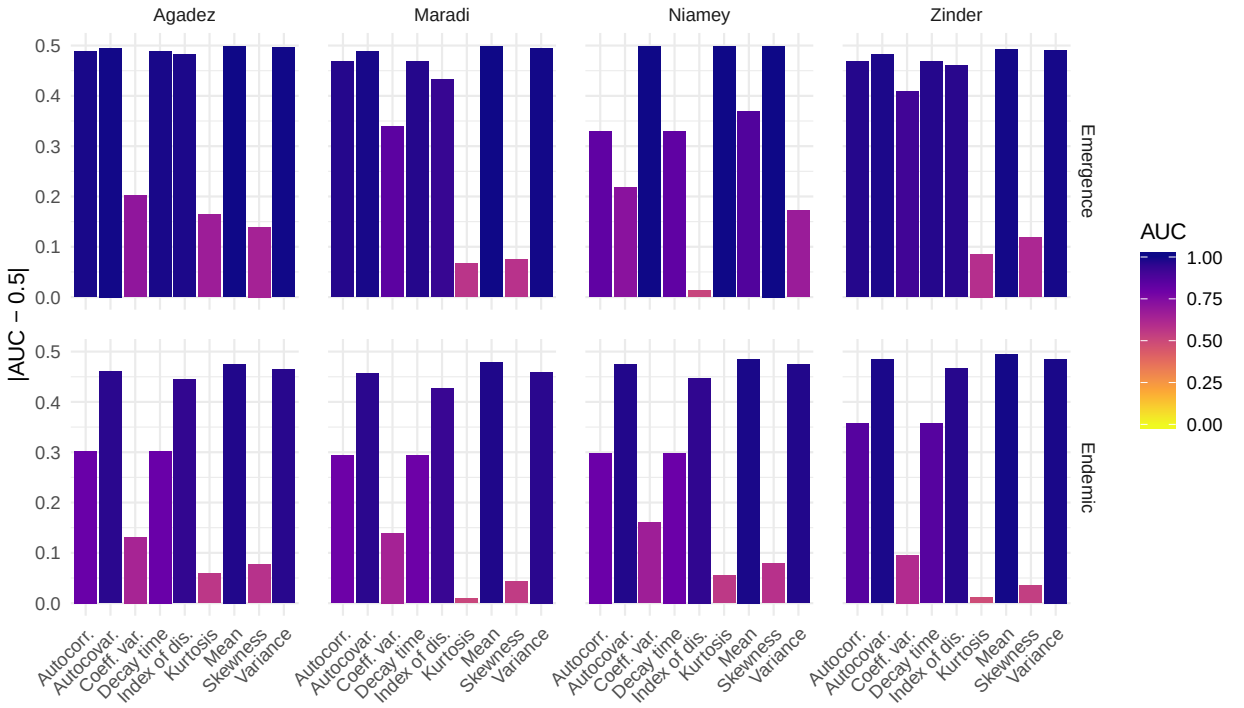

Figure 4: Performance of EWS calculated over two windows (far from and near $\mathscr{R}_{E(t)} = 1$) from the time series of 500 simulated dynamics. The two windows were defined as equally-sized windows over the course of the time series up to $\mathscr{R}_{E(t)} = 1$ (red lines in Figure 3)