

PROTOCOL FOR:

Linking early warning signals to the temporal epidemiology of measles in Nigerian cities

Andrew Tredennick (atredenn@gmail.com)

23 August, 2018

Authors

- Andrew Tredennick
- Eamon O'Dea
- Pejman Rohani
- John Drake

Background

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-free detection system, where statistical properties of disease surveillance 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. To bridge 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. We are particularly interested in the correlation between EWS and the time-varying reproductive ratio, known as the effective reproductive ratio (R_E).

Study design

I. Fit a mechanistic SIR model to measles incidence data from 4 cities in Niger

The goal is fit a model to estimate:

- The latent SI states
- Time-varying rate of transmission
- Effective reproductive ratio (R_E)

The model is a continuous time SIR model, specified as a set of stochastic differential equations (SDEs). We will fit the model to data via iterated filtering to maximize the likelihood, as implemented in the R package `pomp`. Model fitting will proceed in three steps:

1. Generate initial parameter guesses by maximizing a synthetic likelihood defined by (1) the number of zero-case weeks in the time series, (2) the maximum number of cases observed in a week, (3) the cumulative number of cases, and (4) the lag-1 autocorrelation. This is implemented via probe matching in `pomp`:

```
probe_zeroes <- function(y){  
  # number of zeros in the data  
  xy <- y["cases", ]  
  as.numeric(length(which(xy == 0)))  
}
```

```

}

probe_max <- function(y){
  # max incidence in the data
  max(y["cases", ], na.rm = TRUE)
}

probe_cumsum <- function(y){
  # total number of cases in the data
  cases <- y["cases", ]
  max(cumsum(cases))
}

plist <- list(
  probe_zeroes,
  probe_max,
  probe_cumsum,
  probe.acf("cases", lags = 1, transform = sqrt, type = "correlation")
)

pm <- probe.match(
  measles_pomp,
  probes = plist,
  est = param_names,
  nsim = 500,
  transform = TRUE,
  start = coef(measles_pomp),
  method = "Nelder-Mead",
  maxit = 10000
)

```

2. Conduct a global parameter search using a Latin hypercube search space of 50000 parameter sets, with upper and lower values for each parameter specified as the best parameters from probe matching (`pm_params`) multiplied by a factor of ± 5 . The parameter sets within the Latin hypercube will serve as starting conditions for iterated filtering using the function `pomp::mif2()`.
3. Conduct a local parameter search using the the 5000 parameter sets from the global search with the highest likelihoods. We will consider the parameter set with the highest likelihood at this stage to be the MLEs ($\hat{\theta}$).
4. To complete our inference from the model, we will perform particle MCMC using $\hat{\theta}$ as the initial conditions. The MCMC analysis allows us to estimate the uncertainty around all parameters and to generate estimates of the latent states and the R_E for the observed data. That is, we will end up with a time series of R_E that corresponds to the observation time series.

II. Calculate a suite of EWS (from the R package `spaero`) from the same data as used in I

This will be done with the function `spaero::get_stats`.

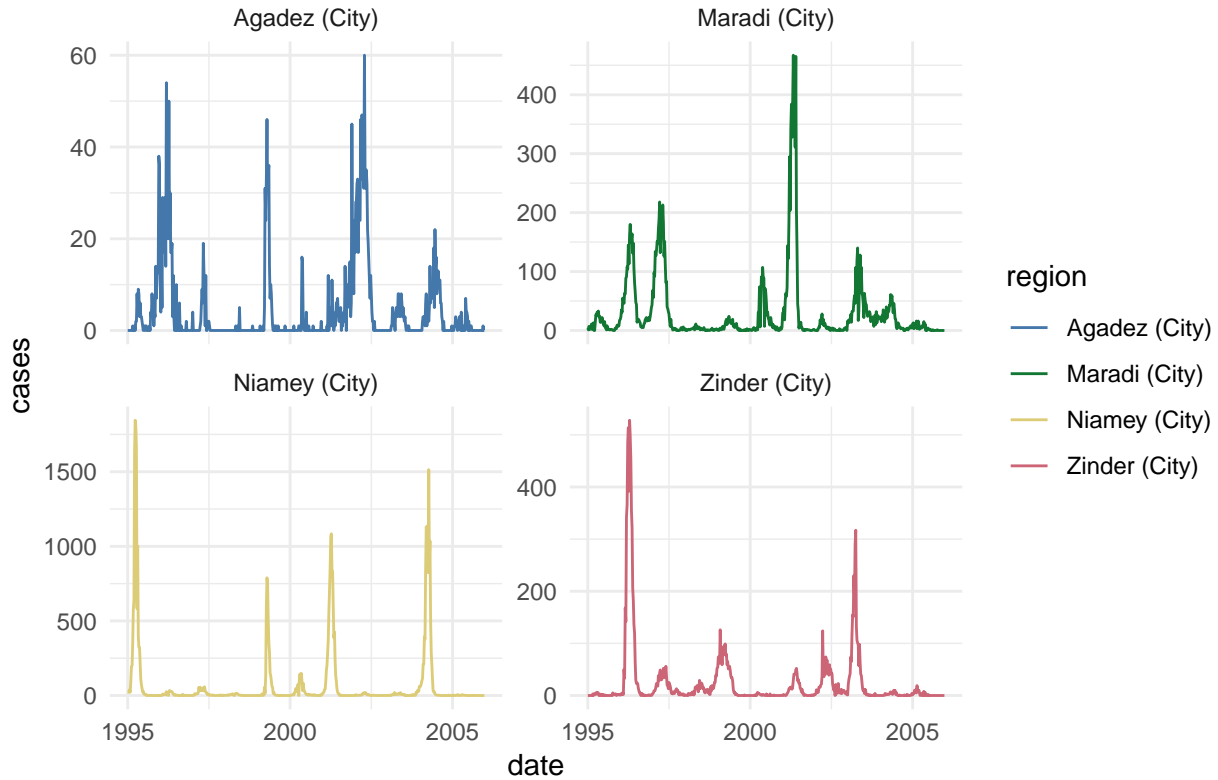
III. Calculate the correlation between EWS and (R_E) and the significance of the correlation

This will be done with the functions `cor` and `cor.test`.

Data sources

- Niger measles data from project AERO
- Permissions?

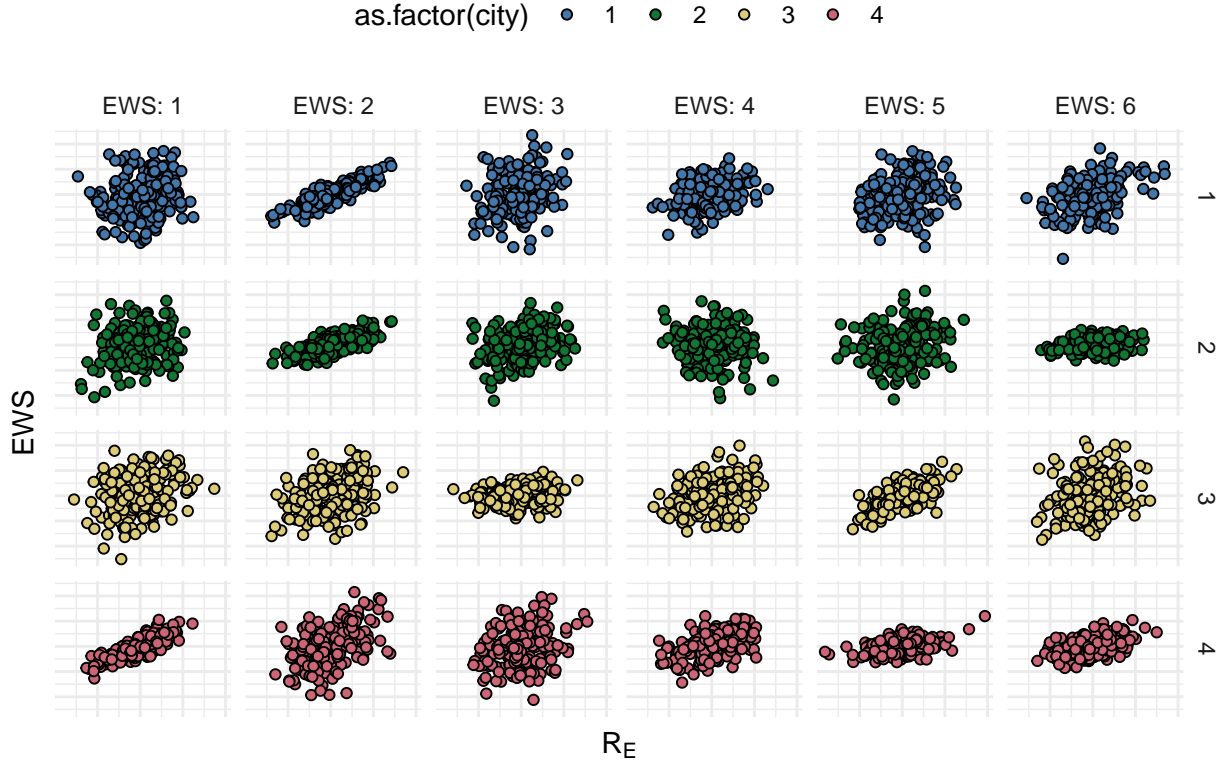
Time series of data from 4 cities in Niger



Analysis

1. Plot time-indexed R_E and each EWS as scatterplots to visualize their correlations (ρ). This might look something like:

Random data showing example plot for the paper



2. For each region, calculate temporal correlations between each EWS and R_E , resulting in a table (or heatmap figure) like this:

Region	EWS	ρ	P -value
Agadez	Variance	0.8	0.001
Agadez	Mean	0.4	0.09
\vdots	\vdots	\vdots	\vdots
Zinder	Kurtosis	0.7	0.006

Checklist

- Create cleaned dataset of just the four focal cities from Niger
- Write R code to define the `pomp` model object (SDE)
- Simulate time series from the `pomp` object with parameters that give us irregular dynamics like the observed data
- Do test analysis of steps II-III using the simulated data
- Perform iterated filter parameter search for MLEs
 - Get set up on UGA cluster
- Calculate EWS on the observed data
 - Work with Eamon on bandwidth, etc.
- Calculate correlations with real data and empirically-derived R_E
- Write the paper!