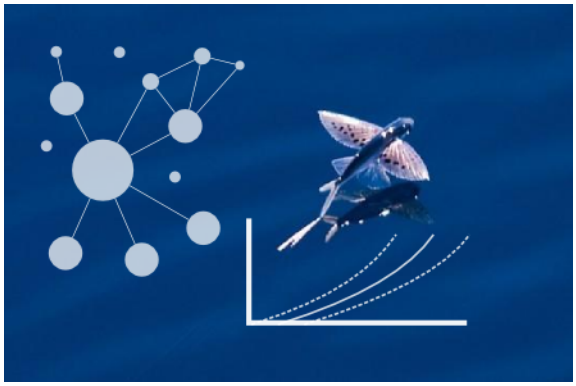
A spatial Poisson transmission model for Ebola (and other diseases)

Thibaut, Pierre, Anne, Simon, Nick G. and the Ebola team

February 1, 2015



Abstract

The model is a meta-population model using a known (spatial) connectivity matrix between patches and a simple kernel to model dispersal. The model is based on incidence data, with only infected individuals being known. Optionally, it can include infection from an unsampled reservoir. Unlike *outbreaker*, we are not trying to model individual ancestries, but merely dynamics within and between patches. In its simplest form, the model has only two parameters and its likelihood can be computed very fast. More complex extensions can allow for time-varying reproduction number, and/or time/spatially varying infection from the reservoir.

1 Notations

- T: number of time steps in the data
- P: number of patches
- I_t^i: observed incidence in patch i at time t (data)
- N_t^i: true, unobserved incidence in patch i at time t (augmented data)
- I_t, N_t: vectors of (observed, true) incidence of all patches ($I_t = (I_t^1, \ldots, I_t^P)$; $N_t = (N_t^1, \ldots, N_t^P)$)
- I, N: matrix of observed / true incidence of all patches at all time steps ($I = (I_1, \ldots, I_T)$; $N = (N_1, \ldots, N_T)$
- $I_{1\to t}, N_{1\to t}$: matrices of observed / true incidence from time 1 to time t ($I_{1\to t} = (I_1, \ldots, I_t)$; $N_{1\to t} = (N_1, \ldots, N_t)$
- D_{ij}: distance between i and j
- $w(.)$: the known probability mass distribution of the generation time / serial interval
- $d_{j\to i}$: intensity of dispersion from j to i
- δ: general dispersal parameter
- π: proportion of cases reported
- τ: vector of temporal trends in under-reporting ($\tau \in \mathbb{R}_+^T$)
- γ: vector of spatial trends in under-reporting ($\gamma \in \mathbb{R}_+^P$)
- R: the effective reproduction number
- $k(a, b)$: a spatial kernel for a distance a and a parameter b
- θ_δ: (fixed) parameter for the prior of δ
- θ_π: (fixed) parameter for the prior of π

2 Computing the number of actual cases

The matrix of actual number of cases is computed as:

$$N = \frac{1}{\pi}\tau \cdot \gamma' \circ I \tag{1}$$

where γ' is the transposed of γ, and $\circ$ is the Hadamard (term-wise) product. The term $\tau \cdot \gamma'$ is a $T \times P$ matrix whose terms are under-ascertainment correction factors valued in $\mathbb{R}_+$ for each patch and each time step; values higher than 1 indicate that under-ascertainment is stronger than the average, while values less than 1 indicate that fewer cases have been missed. $\tau \cdot \gamma'$ should be scaled so that the average value across all its entries is 1.

3 Model

We want to sample from the posterior distribution proportional to:

$$p(I, N, R, \rho, \delta, \pi) = p(I, N|R, \rho, \delta, \pi)p(R, \rho, \delta, \pi) \tag{2}$$

which can be rewritten

$$\underbrace{p(I|N, \pi)p(N|R, \delta)p(R|\rho)}_{likelihood}\underbrace{p(\rho)p(\delta)p(\pi)}_{priors} \tag{3}$$

The term $p(I|N, \pi)$ is the probability of the observed incidence given the true incidence N and the reporting probability π. It is computed as:

$$p(I|N, \pi) = \prod_i \prod_t p(I_t^i|N_t^i, \pi) \tag{4}$$

with:

$$p(I_t^i|N_t^i, \pi) = f_{\mathcal{B}}(I_t^i, N_t^i, \pi) \tag{5}$$

where $f_{\mathcal{B}}(x, a, b)$ is the Binomial p.m.f. for x successes, a draws and a probability b.

The term $p(N|R, \delta)$ is the probability of the true incidence given the infectivity in the system and the spatial processes at play. It is computed as:

$$\begin{aligned} p(N|R, \delta) &= p(N_1, \ldots, N_t|R, \delta) && (6)\\ &= \underbrace{(N_1|R, \delta)}_{constant} p(N_2|N_1, R, \delta) \ldots p(N_t|N_1, \ldots, N_{t-1}, R, \delta) && (7)\\ &\propto \prod_{t=2}^{T} p(N_t|N_{1\to t-1}, R, \delta) && (8) \end{aligned}$$

$$= \prod_{t=2}^{T}\prod_{i=1}^{P} p(N_t^i|N_{1\to t-1}, R, \delta) \tag{9}$$

$$\tag{10}$$

with:

$$p(N_t^i|N_{1\to t-1}, R, \delta) = f_{\mathcal{P}}(N_t^i, \lambda_t^i) \tag{11}$$

where $f_{\mathcal{P}}$ if the p.m.f of a Poisson distribution and λ_t^i the force of infection towards patch i and time t. λ_t^i is a sum over of the forces of infection of all patches towards i (including $i \to i$). We note β_t^j the global infectiousness coming from infected individuals in patch j at time t, defined by the renewal equation:

$$\beta_t^j = \sum_{s=1}^{t-1} N_s^j \; w(t-s) \tag{12}$$

The force of infection experienced by patch i at time t is then:

$$\lambda_t^i = R \sum_{j=1}^{P} \beta_t^j \; k(d_{j\to i}, \delta) \tag{13}$$

where R is the reproduction number. The likelihood of the augmented data is thus given by:

$$p(N|R, \delta) \propto \prod_{t=2}^{T}\prod_{i=1}^{P} f_{\mathcal{P}}(N_t^i, R \sum_{j=1}^{P}\sum_{s=1}^{t-1} N_s^j \; w(t-s) \; k(d_{j\to i}, \delta)) \tag{14}$$

Note that we can easily turn it into a time-dependent term R_t, in which case we will have to assume i) a functional form or ii) R_t constant between break-points. Similarly, it can be turned into a patch-specific R_i, in which case heterogeneity between patches can be modeled using a given distribution. For the time being, heterogeneity in R across space and time will be modelled using a gamma distribution:

$$p(R|\rho) = f_{\gamma}(R, \rho_1, \rho_2) \tag{15}$$

where ρ is a vector containing the mean ρ_1 and variance ρ_2 of a gamma distribution. The corresponding shape parameter α is:

$$\alpha = \frac{\rho_1^2}{\rho_2} \tag{16}$$

and the corresponding rate parameter β is:

$$\beta = \frac{\rho_1}{\rho_2} \tag{17}$$

The prior of ρ_1 and ρ_2 are normally distributed; by default we use:

$$\rho_1 \sim \mathcal{N}(3, 0.5) \text{ and } \rho_2 \sim \mathcal{N}(6, 0.5) \tag{18}$$

The prior of δ is an exponential distribution of parameter θ_δ:

$$p(\delta) = f_{exp}(\delta, \theta_\delta) \tag{19}$$

The prior of π is a beta distribution of parameter θ_π:

$$p(\pi) = f_{beta}(\delta, \theta_\pi) \tag{20}$$

4 Data augmentation: Gibbs sampler

We seek a Gibbs sampler for the augmented incidence at time t, N_t. For this, we need to know (up to an additive or multiplicative constant) the distribution of N_t conditional on everything else:

$$p(N_t|N_{1\to t-1}, I_{1\to t-1}, R, \rho, \delta, \pi) \propto p(N_{1\to t}, I_{1\to t}, R, \rho, \delta, \pi) \tag{21}$$

$$\propto p(I_{1\to t}|N_{1\to t}, \pi)p(N_t|N_{1\to t-1}, R, \delta) \tag{22}$$

$$= (\prod_{s=1}^{t} p(I_s|N_s, \pi))\ p(N_t|N_{1\to t-1}, R, \delta)) \tag{23}$$

$$\propto p(I_t|N_t, \pi)\ p(N_t|N_{1\to t-1}, R, \delta) \tag{24}$$

$$= (\prod_{j=1}^{P} f_{\mathcal{B}}(I_t^j, N_t^j, \pi))(\prod_{j=1}^{P} f_{\mathcal{P}}(N_t^j, \lambda_t^j)) \tag{25}$$

$$= \prod_{j=1}^{P}[\binom{N_t^j}{I_t^j}\pi^{I_t^j}(1-\pi)^{N_t^j - I_t^j}\ \frac{{\lambda_t^j}^{N_t^j} e^{-\lambda_t^j}}{N_t^j!}] \tag{26}$$

$$\propto \prod_{j=1}^{P}[\frac{N_t^j!}{I_t^j!(N_t^j - I_t^j)!}(1-\pi)^{N_t^j - I_t^j}\ \frac{{\lambda_t^j}^{N_t^j}}{N_t^j!}] \tag{27}$$

$$\propto \prod_{j=1}^{P}[\frac{1}{(N_t^j - I_t^j)!}(1-\pi)^{N_t^j - I_t^j}\ {\lambda_t^j}^{N_t^j - I_t^j}{\lambda_t^j}^{I_t^j}] \tag{28}$$

$$\propto \prod_{j=1}^{P}\frac{1}{(N_t^j - I_t^j)!}[(1-\pi)\lambda_t^j]^{N_t^j - I_t^j} \tag{29}$$

$$\propto \prod_{j=1}^{P} f_{\mathcal{P}}(N_t^j - I_t^j, (1-\pi)\lambda_t^j) \tag{30}$$

The conditional distribution of N_t is therefore proportional to the probability of the number of unobserved cases $N_t^j - I_t^j$, given by a Poisson distribution of rate $(1-\pi)\lambda_t^j$. This is an intuitive result: the actual incidence is larger for smaller reporting rates (π) and higher forces of infection (λ_t^j). In practice, we can derive N_t by sampling directly the number of unobserved cases $N_t^j - I_t^j$ from a Poisson distribution of parameter $(1-\pi)\lambda_t^j$.