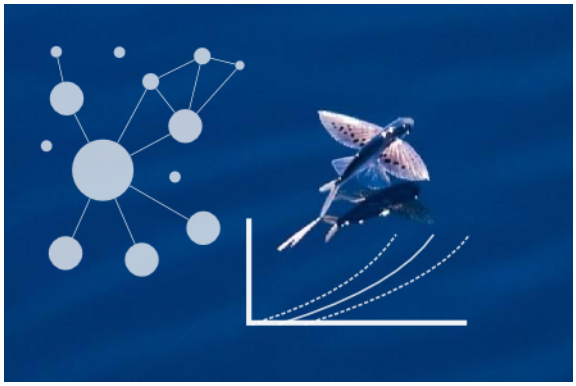
A spatial Poisson transmission model for Ebola (and other diseases)

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

September 18, 2014



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

- I_t^i: incidence in patch i at time t (data)
- I_t: vector of incidence of all patches $(I_t^1, \ldots, I_t^N)$
- T: the last date of the data
- D_{ij}: distance between i and j
- $w(.)$: the known probability mass distribution of the generation time / serial interval
- N: number of patches in the model
- n_t^i: the number of infected individuals in patch i at time t
- $d_{j \to i}$: intensity of dispersion from j to i
- δ: general dispersal parameter
- R: the effective reproduction number
- t_k: the date of infection of individual k
- $f_{\mathcal{P}}(a, b)$: the Poisson density for a observations and a rate b
- $k(c, d)$: a spatial kernel for a distance c and a parameter d

2 Model

The incidence I_t^i is assumed to follow a Poisson distribution of parameter λ_t^i. 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^i the force of infection coming from infected individuals in patch i at time t, defined by the renewal equation:

$$\beta_t^i = R \sum_{k=1}^{n_t^i} w(t - t_k) \tag{1}$$

Where R is the reproduction number. 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. This can get more complicated quickly.

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

$$\lambda_t^i = \sum_{j=1}^{N} d_{j \to i} \beta_t^i \tag{2}$$

where $d_{j \to i}$ is the diffusion from j to i defined by a kernel k:

$$d_{j \to i} = k(D_{ij}, \delta) \tag{3}$$

where D_{ij} is the known distance between i and j (which needs not be Euclidean or symmetric).

The likelihood of I_t, the incidence vector at time t, is defined as:

$$p(I_t | R, \delta) = \prod_{i=1}^{N} f_{\mathcal{P}}(I_t^i, \lambda_t^i) \tag{4}$$

where $f_{\mathcal{P}}$ is the pmf of a Poisson distribution.

By extension, the likelihood for the entire data is:

$$p(I_1, \ldots, I_t | R, \delta) = \prod_{t=1}^{T} \prod_{i=1}^{N} f_{\mathcal{P}}(I_t^i, \lambda_t^i) \tag{5}$$

3 Unobserved reservoir

So far the model assumes that all patches are known, and the connectivity between the patches is known as well – up to a dispersal parameter δ. An unobserved reservoir (e.g. zoonotic cases) can be added using a few assumptions. To model the reservoir, we need to estimate the force of infection coming from the reservoir to the different patches:

$$\phi_t^i = d_{res \to i} \beta_t^{res} \tag{6}$$

We can make different assumptions on the two components, depending on whether we are interested in modeling spatial or temporal heterogeneity. The simplest approach is to assume a constant force of infection over time and equal across patches, in which case only one additional parameter needs to be estimated. A more flexible approach would be to use a distribution of ϕ_t^i to account for heterogeneity between patches. This is still a parsimonious approach in terms of parameters (typically 1-2 would be needed).

4 Complexity

In its simplest form, the model has 2 parameters without reservoir (R,δ), and 3 with reservoir (R,δ,ϕ). Likelihood computation will be fast, as we need only compute Poisson densities, and no data augmentation. It can become much more complex with time-varying R and time-varying, spatially heterogeneous infection from the reservoir ϕ. There will probably also be issues trying to estimate time-varying R and ϕ, as the two effects would be largely confounded.