Bayesian inference for spatiotemporal individual-level epidemic models

L.A.C. Chapman[1,2], S.E.F. Spencer[3], T.M. Pollington[3], C.P. Jewell[2], T.D. Hollingsworth[4], C. Bern[5], G.F. Medley[1]

[1]London School of Hygiene and Tropical Medicine, UK
[2]Lancaster University, UK
[3]University of Warwick, UK
[4]University of Oxford, UK
[5]University of California San Francisco, USA

Introduction

- Bayesian inference methods for individual-level spatiotemporal epidemic data not developed for endemic or highly asymptomatic diseases
- Challenging due to lots of missing data and unknown initial population immune status
- Visceral leishmaniasis (VL) is a parasitic disease with these characteristics
- We developed MCMC data augmentation methods to address this gap and reconstruct transmission trees for person-to-person disease spread

Data

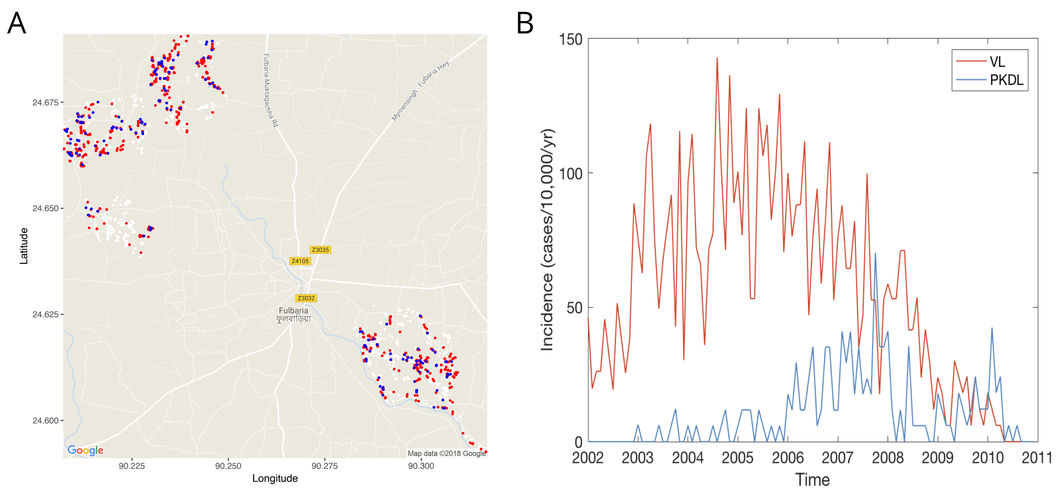
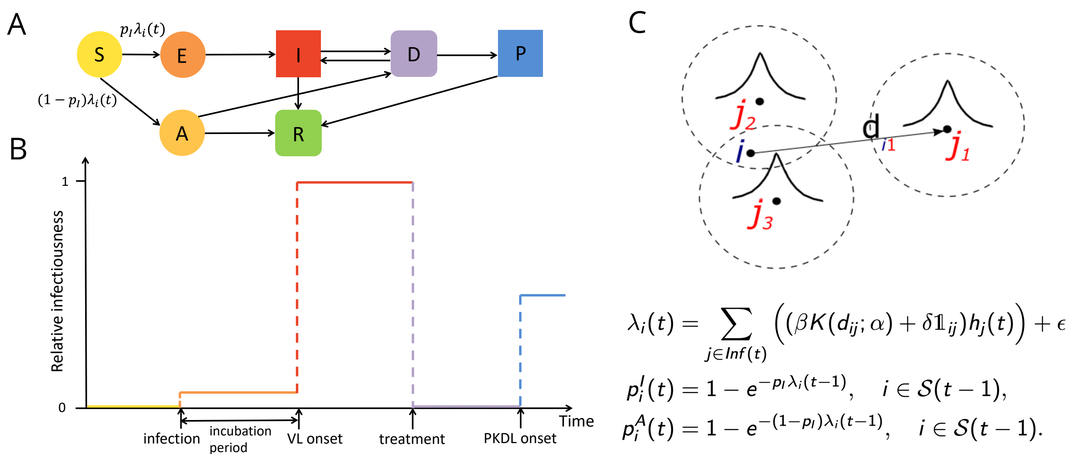


Fig. 1: (A) Locations of 1018 cases of VL (red), and 190 cases of its infectious sequela PKDL (blue) among 24,781 individuals in 5118 households (white dots) in the study area in Bangladesh. (B) Incidence of VL (red) and PKDL (blue) in the study area, 2002-2010.

Model



$$\lambda_i(t) = \sum_{j \in Inf(t)} \Big((\beta K(d_{ij}; \alpha) + \delta \mathbb{1}_{ij}) h_j(t) \Big) + \epsilon$$

$$p_i^I(t) = 1 - e^{-p_I \lambda_i(t-1)}, \quad i \in \mathcal{S}(t-1),$$

$$p_i^A(t) = 1 - e^{-(1-p_I)\lambda_i(t-1)}, \quad i \in \mathcal{S}(t-1).$$

Fig. 2. Individual-level discrete-time spatiotemporal transmission model. (A) Flow diagram. S = susceptible, A = asymptomatic, E = presymptomatic, I = VL, D = dormant infection, P = PKDL, R = recovered. (B) Infectiousness of different infection states in the top pathway in (A). (C) Infection pressure on susceptible individual *i* at time *t*, $\lambda_i(t)$, depends on how many infectious individuals there are around them (*Inf(t)* = {j_1, j_2, j_3}), how far away they are (d_{ij}) as shown by curves ($K(d_{ij}, \alpha)$), and how infectious they are ($h_j(t)$).

Inference

- Estimate joint posterior of parameters $\boldsymbol{\theta} = (\beta, \alpha, \delta, \epsilon)$ and missing data $\boldsymbol{X}$:

$$P(\boldsymbol{\theta}, \boldsymbol{X}|\boldsymbol{Y}) = \frac{P(\boldsymbol{Y}, \boldsymbol{X}|\boldsymbol{\theta})}{P(\boldsymbol{Y})} \propto L(\boldsymbol{\theta}; (\boldsymbol{Y}, \boldsymbol{X}))P(\boldsymbol{\theta})$$

  where $\boldsymbol{Y}$ is observed data and $L(\boldsymbol{\theta}; (\boldsymbol{Y}, \boldsymbol{X}))$ is complete data likelihood by iteratively sampling from $P(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X})$ and $P(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{Y})$

- Update unobserved asymptomatic infection periods efficiently by:
  - assigning each non-symptomatic individual an asymptomatic infection and recovery time pair, and including pairs representing infection and recovery before/after the study
  - proposing new asymptomatic infection times according to running estimate of infection pressure

Results

- Reconstructed transmission trees (Fig. 4) post-hoc by drawing an infector for each infectee (including asymptomatics) from posterior distribution of possible infectors 1000 times
- Inferred transmission trees showed multiple super-spreading events around VL and PKDL cases (Fig. 4), strongly correlated with symptom duration
- Estimates of key epidemiological quantities, e.g.:
  - Decrease in contribution of VL to new VL cases and increase in contribution of PKDL from 2002-2010 (Fig. 3B)
  - Average distance to VL infectees of <300m for most infectors
  - Average time from VL onset to VL-infectee infection of <4 months for most cases

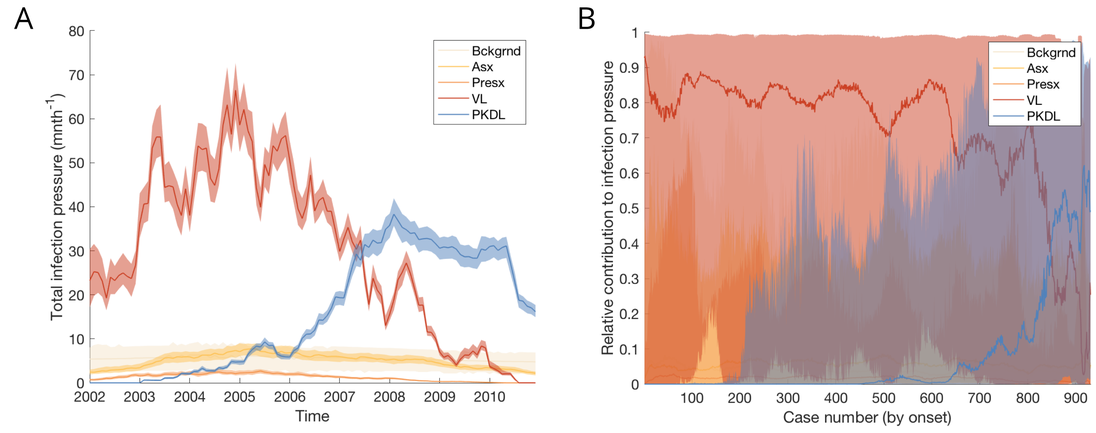


Fig. 3. (A) Total infection pressure on susceptible individuals from each infection state and background transmission. (B) Relative contribution of each infection state to infection pressure on each VL case at their infection time. N.B. Time is increasing non-linearly on horizontal axis since cases are ordered by their onset time.

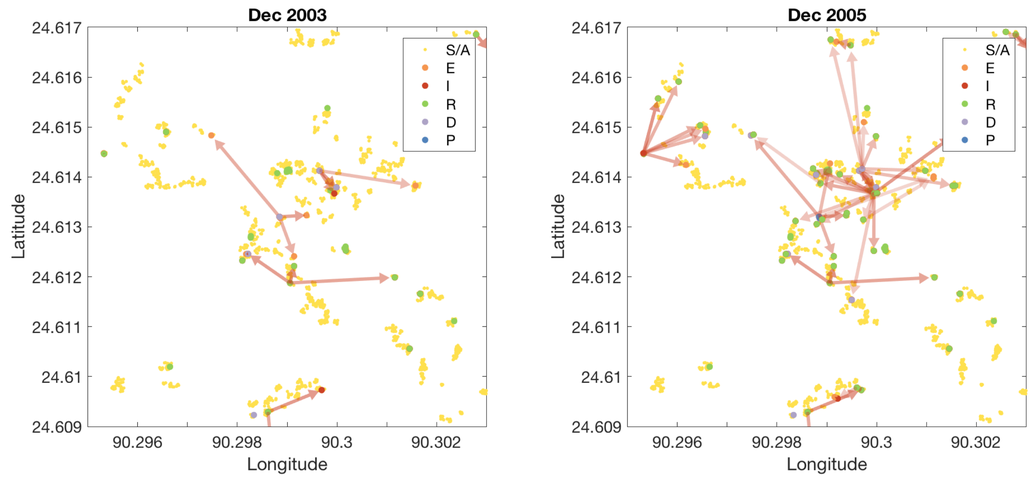


Fig. 4. Inferred transmission tree in part of the south-east village cluster at different stages of the epidemic. Arrows show most likely source of infection for each case over all sampled transmission trees, coloured by type of infection source and shaded according to proportion of trees in which that individual was most likely infector (darker shading = higher proportion). Dot colour shows infection state of individual (see key).

Conclusions

- Data augmentation methods can be used to:
  - infer spatiotemporal transmission trees for endemic diseases in which asymptomatic infection plays a hidden role even for reasonable population sizes (~25,000) and relatively high proportions of asymptomatic infection (~80%)
  - provide estimates of parameters key to disease control.

BILL & MELINDA GATES foundation