A Bayesian geospatial modelling framework for the synthesis of point prevalence and health facility catchment data

Katherine Battle, Ewan Cameron, Peter Gething, Justin Millar, Andre Python, and Tasmin Symons
Nuffield Department of Medicine, Big Data Institute, University of Oxford, United Kingdom

Introduction

Sources of clinical incidence data on malaria:
- (A) malaria attributable fever: MAF
- (B) non-malaria febrile illness with asymptomatic *P. falciparum* (*Pf*) infection: NMFI

Incidence data alone cannot distinguish (A) from (B)

PfPR-to-incidence relationship based on PR survey data alone can only estimate (A)

Aims

Predict pixel-month and pixel-annual malaria incidence along with:
- uncertainty quantification on predicted incidence
- predicted prevalence and background fever surfaces
- estimated proportion of MAF and NMFI cases
- estimated probability to visit a specific health facility (catchment model)

Model overview

THREE GAUSSIAN PROCESSES jointly fit

(1) MALARIA PREVALENCE PfPR classical geostatistics approach with logit-transformed prevalence fit as a linear function of spatio-temporal covariates and a spatio-temporal random effect given by a second Gaussian process $GP(lat,lon,t)$ which captures the variability unexplained by the covariates;

(2) PfPR-TO-INCIDENCE relationship Gaussian process (GP$s$) allows for a smooth but complex, non-linear, PfPR-to-incidence relationship to be learned statistically;

(3) BACKGROUND FEVER PREVALENCE BFPR$_{F+}$ modelled very similarly to the PfPR surface (1) with a Gaussian process $GP_{F+}$ which is used for $BFPR_{F+}$, but are selected independently.

Future results

MAPS
- pixel-month and pixel-year
- incidence rate and case counts
- jointly-fit surfaces of PfPR and BFPR$_{F+}$

UNCERTAINTY
- quantification of the model uncertainty based on non-spatial bootstrapping method as an alternative to computationally expensive posterior sampling methods.

DISENTANGLING MAF AND NMFI:
- estimates of the proportion of cases which are malaria-attributable (MAF) vs non-malarial febrile illnesses co-infection with a *P. falciparum* infection (NMFI) complementary to [2].

Model structure

- MODELLING FRAMEWORK: Bayesian hierarchical geospatial model
- DATA: prevalence survey data $y^{inc}_{i,t}$ and health facility incidence count data $y^{inc}_{i,t} = MAF + NMFI_{Pr}$

INCIDENCE DATA MODEL: $y^{inc}_{i,t} \sim \text{Poisson} \left( \sum_{j} C_{i,j,t} \times \text{pop}_j \times TS_j \right)$ composed of:

- population at risk estimated as population $\text{pop}_j$ multiplied by probability of treatment seeking $TS_j$
- catchment model for each health facility $j$; $C_{i,j,t} \sim \text{Uni}\left( M / \text{pop}_j \right)$ based on a modified gravity model
- incidence rate $inc_{i,j,t}$ given as a transformation of the prevalence fields [1]:
  $$inc_{i,j,t} = \alpha PfPR_{i,t} \exp(f(PFPR_{i,t}) + \beta BFPR_{F+*i,t})$$
- PfPR-to-incidence modelled as a function of a Gaussian process: $f \sim GP_{\phi}$
- BFPR$_{F+*i,t}$ as background fever prevalence modelled similar to the PfPR surface with a Gaussian process $GP_{F+}$

DATA MODEL for PfPR: $y^{PR}_{i,t} \sim \text{Binomial}(PfPR_{i,t}, N_{i,t})$, with $N_{i,t}$: tested cases

LINEAR PREDICTOR for PfPR and BFPR: Gaussian process $GP(lat, lon, t)$ + covariates $X$:

- linear fit of PfPR$_{i,t}$ as $GP$ model
- linear fit of BFPR$_{F+*i,t}$ as $GP$ model

PRIORS set for: prevalence-incidence $\alpha, \gamma$, Gaussian processes $\phi, \psi$, prevalence surface slopes $\beta, \beta'$ and intercepts $c, c'\sim\pi$

Summary

- Modelling framework that distinguishes incidence cases coming from malaria (MAF) and non-malarial attributed fever (NMFI$_{Pr}$)
- Joint model that fits together incidence rate and prevalence in space and time
- Joint fit of several data sources: incidence & PfPR & background fever prevalence

References