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

Andre Python[1], Ewan Cameron[1], Katherine Battle[1], Peter Gething[2], Justin Millar[1], and Tasmin Symons[1]

[1]University of Oxford
[2]Curtin University

November 21, 2022

Abstract

We present a Bayesian geospatial modeling framework developed for the synthesis of point prevalence and health facility catchment data of mixed types: Plasmodium parasite prevalence and malaria febrile incidence. Since the clinical case definition for health facility record keeping is less strict than that used in cohort studies used to construct previous parasite prevalence to clinical incidence relationships (the latter usually incorporating a parasite density threshold to remove background fevers accompanied by coincidence asymptomatic parasite infections) our model learns a smooth prevalence-to-incidence conversion during posterior sampling. Also jointly fitted are a catchment model based on the relative travel times between each pixel location and its nearby health facilities, as well as a distribution regression-based covariance structure for explaining the residual errors at health facility level based on the similarity between the 'bags' of covariate values in their respective catchments.

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: **$NMFI_{Pf+}$**

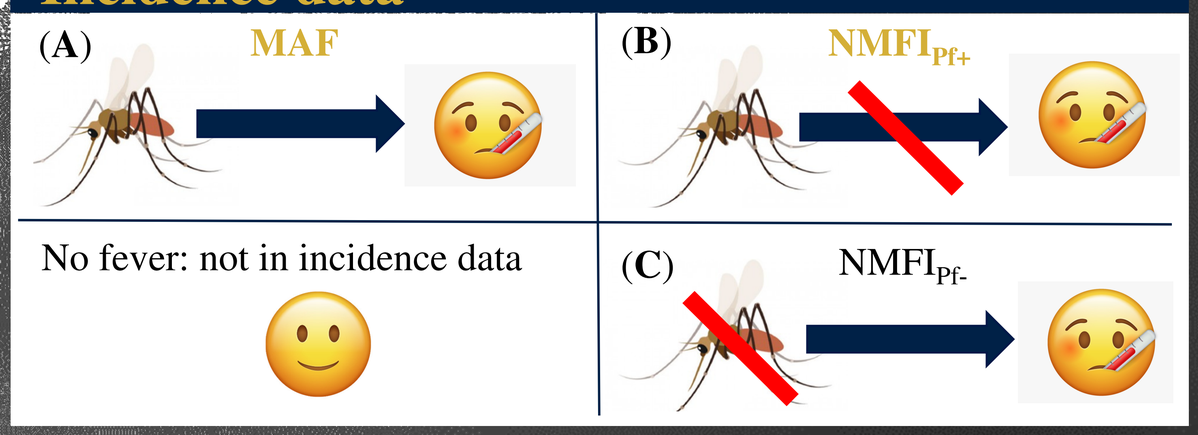
- Incidence data alone cannot distinguish (**A**) from (**B**)
- *Pf*PR-to-incidence relationship based on PR survey data alone can only estimate (**A**)

Aim

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 **$NMFI_{Pf+}$** cases
- estimated probability to visit a specific health facility (catchment model)

Incidence data



Model overview

THREE GAUSSIAN PROCESSES jointly fit

(1) MALARIA PREVALENCE *Pf*PR

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 $GP(\text{lat}, \text{lon}, \text{t})_{\psi}$ which captures the variability unexplained by the covariates;

(2) *Pf*PR-TO-INCIDENCE relationship

Gaussian process (GP_{ϕ}) allows for a smooth but complex, non-linear, *Pf*PR-to-incidence relationship to be learned statistically;

(3) BACKGROUND FEVER PREVALENCE $BFPR_{Pf+}$

modelled very similarly to the *Pf*PR surface (1) with a Gaussian process $GP(\text{lat}, \text{lon}, \text{t})_{\rho}$. The covariates for $BFPR_{Pf+,i,t}$ are chosen from the same pool as for *Pf*PR, but are selected independently.

Model structure

- MODELLING FRAMEWORK: Bayesian hierarchical geospatial model
- DATA: prevalence survey data $y_{i,t}^{PR}$ and health facility incidence count data $y_{i,t}^{inc}$ = **MAF** + **$NMFI_{Pf+}$**
- INCIDENCE DATA MODEL: $y_{i,t}^{\text{inc}} \sim \text{Poisson}\left(\sum_j C_{j\to i} \times \text{inc}_{j,t} \times \text{pop}_j \times TS_j\right)$ composed of:
  - population at risk estimated as population pop_j multiplied by probability of treatment seeking TS_j
  - catchment model for each health facility j: $C_{\text{j}\to\text{i}} \propto \frac{M_j}{d_{j\to i}^2}$ based on a modified gravity model
  - incidence rate $inc_{i,t}$ given as a transformation of the prevalence fields [1]:

$$inc_{i,t} = \alpha PfPR_{i,t} \exp f\left(PfPR_{i,t}\right) + \gamma BFPR_{Pf+,i,t}$$

  - *Pf*PR-to-incidence modelled as a function of a Gaussian process: $f \sim GP_{\phi}$
  - $BFPR_{Pf+,i,t}$ as background fever prevalence ($BFPR$) times malaria prevalence ($PfPR$):

$$BFPR_{Pf+,i,t} = BFPR_{i,t} \times PfPR_{i,t}$$

- DATA MODEL for *Pf* PR: $y_{i,t}^{PR} \sim Binomial(PfPR_{i,t}, N_{i,t})$, with $N_{i,t}$: tested cases
- LINEAR PREDICTOR for $PfPR$ and $BFPR$: Gaussian process $GP(lat, lon, t)_{i,t}$ + covariates $\boldsymbol{X}$:

$$logit\left(PfPR_{i,t}\right) = GP(lat, lon, t)_{i,t,\psi}\boldsymbol{X}\beta + c$$
$$logit\left(BFPR_{i,t}\right) = GP(lat, lon, t)_{i,t,\rho}\boldsymbol{X}\beta' + c'$$

- PRIORS set for: prevalence-incidence α, γ, Gaussian processes ϕ, ψ, ρ prevalence surface slopes β, β' and intercepts c, c'

$$\alpha, \gamma, \phi, \psi, \rho, \beta, \beta' c, c' \sim \pi$$

Future results

MAPS
- pixel-month and pixel-year
- incidence rate and case counts
- jointly-fit surfaces of *Pf*PR and $BFPR_{Pf+,i,t}$

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

DISENTANGLE **MAF** AND **$NMFI_{Pf+}$**
- estimates of the proportion of cases which are malaria-attributable (**MAF**) vs non-malarial febrile illnesses co-incidence with a *P. falciparum* infection (**$NMFI_{Pf+}$**) complementary to [2].

Summary

- Modelling framework that distinguishes incidence cases coming from malaria (**MAF**) and non-malaria attributed fever (**$NMFI_{Pf+}$**)
- Joint model that fits together incidence rate and prevalence in space and time
- Joint fit of several data sources: incidence & *Pf*PR & background fever prevalence

References

[1] Cameron et al., 2015. Defining the relationship between infection prevalence and clinical incidence of Plasmodium falciparum malaria. *Nature Com.* 6, 8170.

[2] Dalrymple, U. et al., 2019. The Contribution of Non-Malarial Febrile Illness Co-Infections to Plasmodium Falciparum Case Counts in Health Facilities in Sub-Saharan Africa. *Malaria J. 18(195).*