Malaria, river-blindness and plague: three case studies in spatial modelling of tropical diseases

Emanuele Giorgi and Peter J. Diggle

Lancaster Medical School, Lancaster University, Lancaster, UK

LSHTM, 24 April 2015

Three issues of disease mapping in low-resource settings

1. **Combining data from multiple surveys of different quality**: e.g. randomised and non-randomised surveys.

Three issues of disease mapping in low-resource settings

1. **Combining data from multiple surveys of different quality**: e.g. randomised and non-randomised surveys.
2. **Spatially structured zero-inflation**: chance finding or disease-free community?

Three issues of disease mapping in low-resource settings

1. **Combining data from multiple surveys of different quality**: e.g. randomised and non-randomised surveys.
2. **Spatially structured zero-inflation**: chance finding or disease-free community?
3. **Analysis of spatio-temporal data at multiple spatial-scales**: how to integrate district-level disease case reports with high resolution environmental data?

Overview

1. **Malaria**: combining data from a community and a (biased) school-based surveys in Nyanza Province, Kenya.

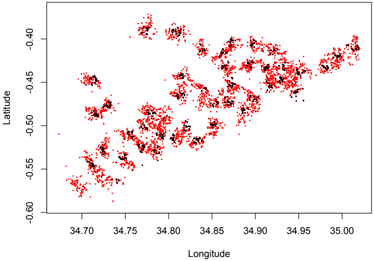
Overview

1. **Malaria**: combining data from a community and a (biased) school-based surveys in Nyanza Province, Kenya.
2. **River-blindness**: geostatistical modelling of spatially-structured zero inflation in prevalence data from Mozambique, Malawi and Tanzania.

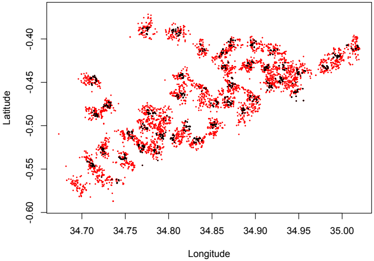
Overview

1. **Malaria**: combining data from a community and a (biased) school-based surveys in Nyanza Province, Kenya.
2. **River-blindness**: geostatistical modelling of spatially-structured zero inflation in prevalence data from Mozambique, Malawi and Tanzania.
3. **Plague**: spatio-temporal analysis of monthly plague incidence from 2000 to 2007 in Madagascar.

Malaria mapping in Nyanza Province



Malaria mapping in Nyanza Province



Statistical issues

- Different unmeasured risk factors for malaria.
- Joint model or gold-standard data only?

Accounting for residual spatial bias

- $r_d(x)$ = ``odds for the disease''
- $r_s(x)$ = ``odds within the survey''
- $r_c(x)$ = ``relative odds between participants and non-participants to the survey''

Accounting for residual spatial bias

- $r_d(x)$ = ``odds for the disease''
- $r_s(x)$ = ``odds within the survey''
- $r_c(x)$ = ``relative odds between participants and non-participants to the survey''

$$r_s(x) = r_d(x) \times r_c(x)$$

Accounting for residual spatial bias

- $r_d(x)$ = ``odds for the disease''
- $r_s(x)$ = ``odds within the survey''
- $r_c(x)$ = ``relative odds between participants and non-participants to the survey''

$$r_s(x) = r_d(x) \times r_c(x) \iff \log\{r_s(x)\} = S(x) + B(x) \quad (1)$$

Accounting for residual spatial bias

- $r_d(x)$ = ``odds for the disease''
- $r_s(x)$ = ``odds within the survey''
- $r_c(x)$ = ``relative odds between participants and non-participants to the survey''

$$r_s(x) = r_d(x) \times r_c(x) \Longleftrightarrow \log\{r_s(x)\} = S(x) + B(x) \quad (1)$$

- $Y_{ijk} | S(x_{ij}), B(x_{ij}) \sim$ Bernoulli(p_{ijk}) (i=``survey", $j$=``household", k=``'individual'').

Accounting for residual spatial bias

- $r_d(x)$ = ``odds for the disease''
- $r_s(x)$ = ``odds within the survey''
- $r_c(x)$ = ``relative odds between participants and non-participants to the survey''

$$r_s(x) = r_d(x) \times r_c(x) \Longleftrightarrow \log\{r_s(x)\} = S(x) + B(x) \tag{1}$$

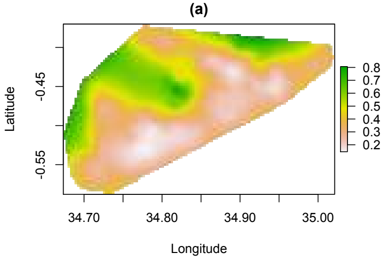
- $Y_{ijk}|S(x_{ij}), B(x_{ij}) \sim$ Bernoulli(p_{ijk}) (i=``survey", $j$=``household", k=```individual'').
- $\log\{p_{ijk}/(1+p_{ijk})\} = d_{ijk}^\top \beta + S(x_{ij}) + B(x_{ij})$;
  $B(x) = 0, \forall x \in \mathbb{R}^2$ if i is ``community''.

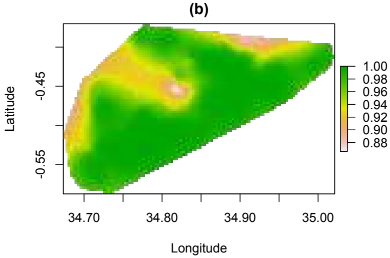
Results (1)

Term	Estimate	95% CI
Intercept	-1.412	(-2.303, -0.521)
Age	-0.141	(-0.174, -0.109)
Rachuonyo District	2.006	(1.228, 2.785)
Socio-Economic-Status	-0.121	(-0.169, -0.072)
School survey	-0.761	(-1.354, -0.167)
Age (bias)	0.094	(0.046, 0.142)

Results (2)

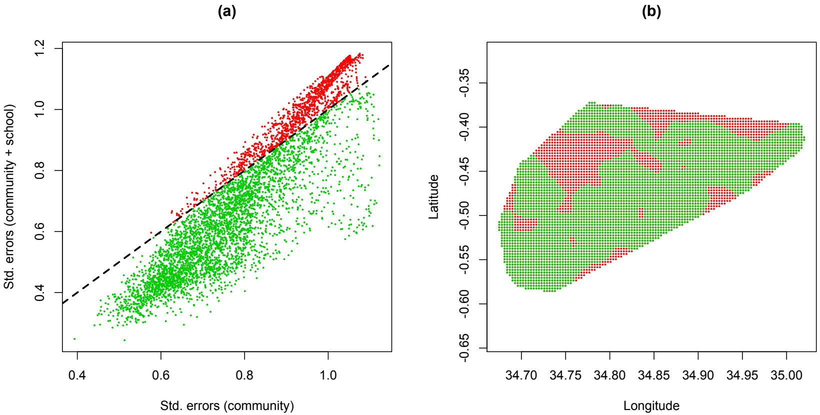
Mapping of (a) $r_c(x)$ and (b) $1 - P(0.9 < r_c(x) < 1.1|y)$.



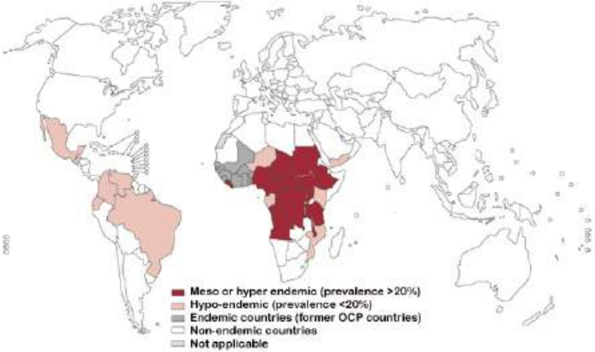


Results (3)

Gain/loss in the accuracy for the estimates of $S(x)$.

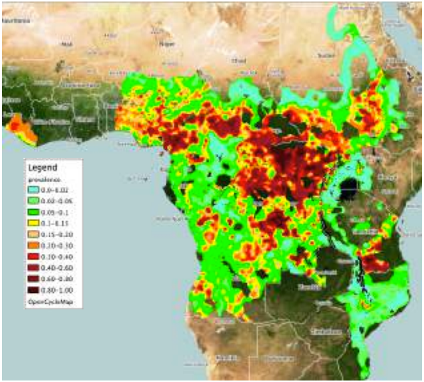


River blindness

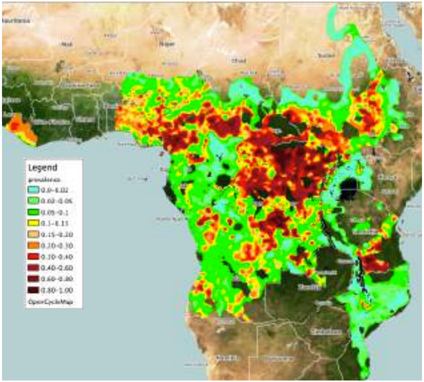


Rapid epidemiological mapping of onchocerciasis

Prevalence estimates in the 20 APOC countries



Prevalence estimates in the 20 APOC countries



Statistical issue

1. $3,320$ villages with no case and $11,154$ villages with at least one case.
2. Accounting for zero-inflation: how to identify disease-free areas?

Extending the standard geostatistical model

- **Standard model**: $Y|S(x) \sim \text{Bin}(\cdot; n, p(x))$ where

$$\log\{p(x)/(1-p(x))\} = \mu + S(x).$$

Extending the standard geostatistical model

- **Standard model**: $Y|S(x) \sim \text{Bin}(\cdot; n, p(x))$ where

$$\log\{p(x)/(1-p(x))\} = \mu + S(x).$$

- **Allowing for zero-inflation**:

$$P(Y = y|S_1(x), S_2(x), D(x)) = \begin{cases} \pi(x) + (1-\pi(x))\text{Bin}(0; n, p(x)) & \text{if } y = 0 \\ (1-\pi(x))\text{Bin}(y; n, p(x)) & \text{if } y > 0 \end{cases}$$

Extending the standard geostatistical model

- **Standard model**: $Y|S(x) \sim \text{Bin}(\cdot; n, p(x))$ where

$$\log\{p(x)/(1-p(x))\} = \mu + S(x).$$

- **Allowing for zero-inflation**:

$$P(Y=y|S_1(x), S_2(x), D(x)) = \begin{cases} \pi(x) + (1-\pi(x))\text{Bin}(0; n, p(x)) & \text{if } y = 0 \\ (1-\pi(x))\text{Bin}(y; n, p(x)) & \text{if } y > 0 \end{cases}$$

where

$$\begin{aligned} \text{logit}(\pi(x)) &= \mu_1 + S_1(x) + D(x), \\ \text{logit}(p(x)) &= \mu_2 + S_2(x) + D(x). \end{aligned}$$

Extending the standard geostatistical model

- **Standard model**: $Y|S(x) \sim \text{Bin}(\cdot; n, p(x))$ where

$$\log\{p(x)/(1-p(x))\} = \mu + S(x).$$

- **Allowing for zero-inflation**:

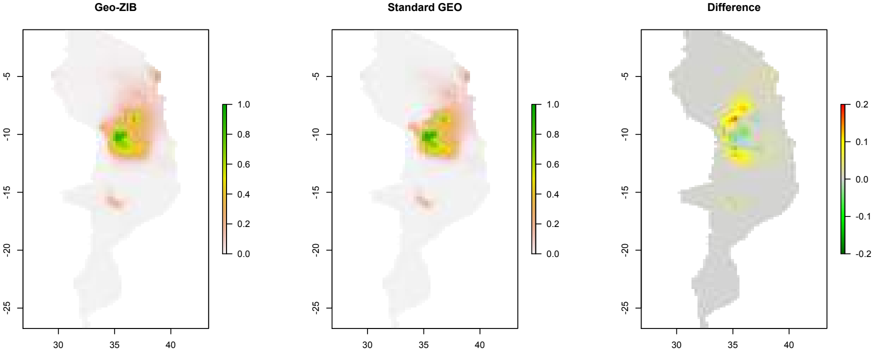
$$P(Y = y|S_1(x), S_2(x), D(x)) = \begin{cases} \pi(x) + (1-\pi(x))\text{Bin}(0; n, p(x)) & \text{if } y = 0 \\ (1-\pi(x))\text{Bin}(y; n, p(x)) & \text{if } y > 0 \end{cases}$$

where

$$\begin{aligned} \text{logit}(\pi(x)) &= \mu_1 + S_1(x) + D(x), \\ \text{logit}(p(x)) &= \mu_2 + S_2(x) + D(x). \end{aligned}$$

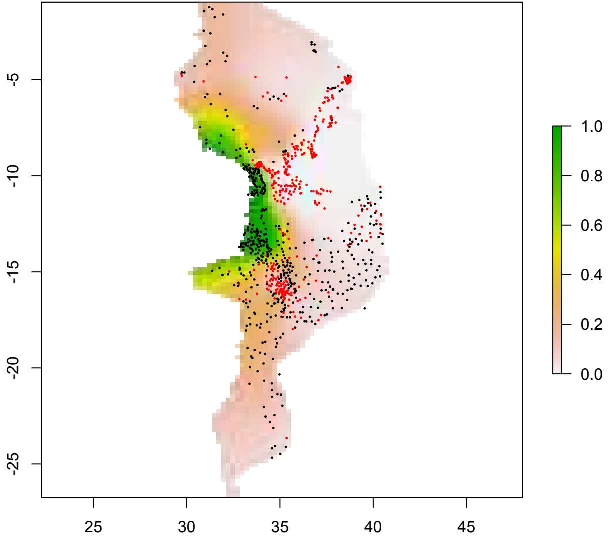
- In the next example $D(x) = 0, \forall x$.

Mapping in Mozambique - Malawi - Tanzania (1)

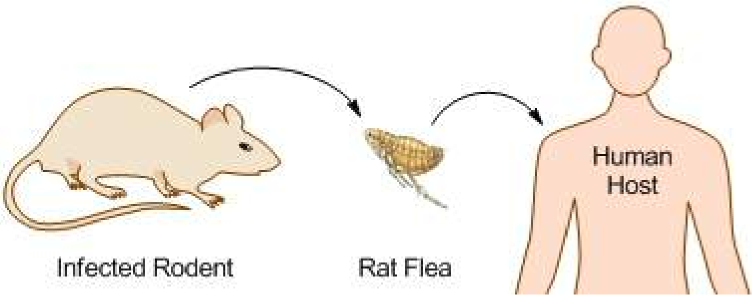


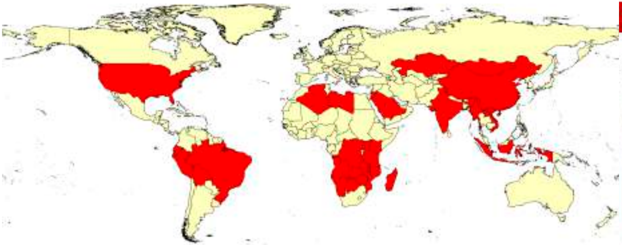
Mapping in Mozambique - Malawi - Tanzania (2)

Mapping of $\pi(x)$.



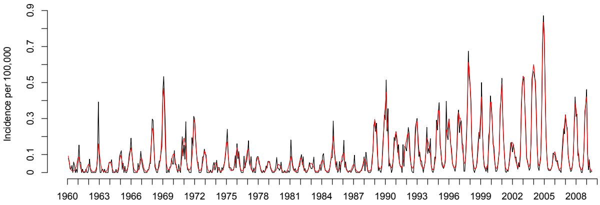
Plague

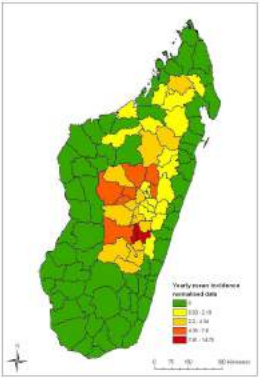

Infected Rodent
Rat Flea
Human
Host



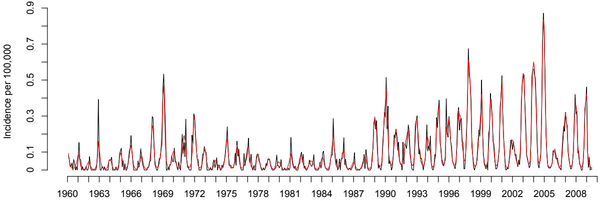

BBC
NEWS
Madagascar plague outbreak kills 40, says WHO

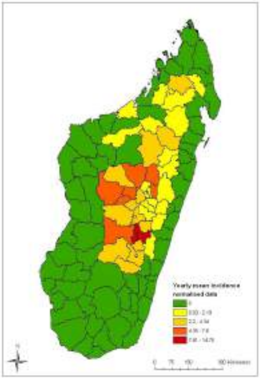
Plague in Madagascar





Plague in Madagascar





Statistical issue

1. Ecological bias.
2. How to integrate spatially continuous data with district-level case reports?

Modelling framework

- $N_{i,t}$ = ``number of cases in the i-th district at time t''

Modelling framework

- $N_{i,t}$ = ``number of cases in the i-th district at time t''
- $\lambda(x,t)$ = ``intensity of an inhomogeneous Poisson process at location x and time t''

Modelling framework

- $N_{i,t}$ = ``number of cases in the i-th district at time t''
- $\lambda(x,t)$ = ``intensity of an inhomogeneous Poisson process at location x and time t''
- $N_{i,t} \sim$ Poisson $\left(\int_t^{t+1} \int_{A_i} \lambda(u,v)\, du\, dv\right)$.

Modelling framework

- $N_{i,t}$ = ``number of cases in the i-th district at time t''
- $\lambda(x,t)$ = ``intensity of an inhomogeneous Poisson process at location x and time t''
- $N_{i,t} \sim$ Poisson $\left(\int_t^{t+1} \int_{A_i} \lambda(u,v)\, du\, dv\right)$.
- $\int_t^{t+1} \int_{A_i} \lambda(u,v)\, du\, dv \approx \Delta \sum_{j=1}^{n_{i,t}} \lambda(x_j, t)$.

Modelling framework

- $N_{i,t}$ = ``number of cases in the i-th district at time t''
- $\lambda(x, t)$ = ``intensity of an inhomogeneous Poisson process at location x and time t''
- $N_{i,t} \sim$ Poisson $\left(\int_t^{t+1} \int_{A_i} \lambda(u, v)\, du\, dv\right)$.
- $\int_t^{t+1} \int_{A_i} \lambda(u, v)\, du\, dv \approx \Delta \sum_{j=1}^{n_{i,t}} \lambda(x_j, t)$.
- $\mathcal{X}_{i,t} = \{X_{j,t} \in A_i\} | N_{i,t} = n_{i,t}$ has multinomial distribution with probabilities

$$p_k = \frac{\lambda(x_k, t)}{\sum_{j=1}^{n_{i,t}} \lambda(x_j, t)}, \text{ for } k = 1, \ldots, n_{i,t}.$$

Intensity of the point process

- $m(x,t)$ = ``number of susceptibles at location x and time t''

Intensity of the point process

- $m(x, t)$ = ``number of susceptibles at location x and time t''
- $\lambda(x, t) = m(x, t) \exp\{d(x, t)^\top \beta + S(x) + V(t)\}$,

Intensity of the point process

- $m(x,t)$ = ``number of susceptibles at location x and time t''
- $\lambda(x,t) = m(x,t)\exp\{d(x,t)^\top\beta + S(x) + V(t)\},$

$$d(x,t)^\top\beta \quad = \quad \beta_1 + \beta_2\text{temp}(x,t) + \beta_3\text{rain}(x,t) +$$

Intensity of the point process

- $m(x,t)$ = ``number of susceptibles at location x and time t''
- $\lambda(x,t) = m(x,t)\exp\{d(x,t)^\top\beta + S(x) + V(t)\}$,

$$
\begin{aligned}
d(x,t)^\top\beta \quad = \quad & \beta_1 + \beta_2\text{temp}(x,t) + \beta_3\text{rain}(x,t) + \\
& \beta_4 I(\text{elev}(x) > 800m) + \beta_5\text{elev}(x) + \\
& \beta_6\text{elev}(x) \times I(\text{elev}(x) > 800m) +
\end{aligned}
$$

Intensity of the point process

- $m(x,t)$ = ``number of susceptibles at location x and time t''
- $\lambda(x,t) = m(x,t)\exp\{d(x,t)^\top\beta + S(x) + V(t)\}$,

$$\begin{aligned} d(x,t)^\top\beta &= \beta_1 + \beta_2\text{temp}(x,t) + \beta_3\text{rain}(x,t) + \\ &\quad \beta_4 I(\text{elev}(x) > 800m) + \beta_5\text{elev}(x) + \\ &\quad \beta_6\text{elev}(x) \times I(\text{elev}(x) > 800m) + \\ &\quad \beta_7 \log\{\text{pop}(x)\} + \end{aligned}$$

Intensity of the point process

- $m(x,t)$ = ``number of susceptibles at location x and time t''
- $\lambda(x,t) = m(x,t)\exp\{d(x,t)^\top\beta + S(x) + V(t)\}$,

$$\begin{aligned} d(x,t)^\top\beta &= \beta_1 + \beta_2\text{temp}(x,t) + \beta_3\text{rain}(x,t) + \\ &\quad \beta_4 I(\text{elev}(x) > 800m) + \beta_5\text{elev}(x) + \\ &\quad \beta_6\text{elev}(x) \times I(\text{elev}(x) > 800m) + \\ &\quad \beta_7 \log\{\text{pop}(x)\} + \\ &\quad [\beta_8 + \beta_9\text{temp}(x,t) + \beta_{10}\text{rain}(x,t)]\sin(2\pi t/12) + \\ &\quad [\beta_{11} + \beta_{12}\text{temp}(x,t) + \beta_{13}\text{rain}(x,t)]\cos(2\pi t/12) \end{aligned}$$

Intensity of the point process

- $m(x,t)$ = ``number of susceptibles at location x and time t''
- $\lambda(x,t) = m(x,t)\exp\{d(x,t)^\top\beta + S(x) + V(t)\}$,

$$
\begin{aligned}
d(x,t)^\top\beta &= \beta_1 + \beta_2\textsf{temp}(x,t) + \beta_3\textsf{rain}(x,t) + \\
&\quad \beta_4 I(\textsf{elev}(x) > 800m) + \beta_5\textsf{elev}(x) + \\
&\quad \beta_6\textsf{elev}(x) \times I(\textsf{elev}(x) > 800m) + \\
&\quad \beta_7 \log\{\textsf{pop}(x)\} + \\
&\quad [\beta_8 + \beta_9\textsf{temp}(x,t) + \beta_{10}\textsf{rain}(x,t)]\sin(2\pi t/12) + \\
&\quad [\beta_{11} + \beta_{12}\textsf{temp}(x,t) + \beta_{13}\textsf{rain}(x,t)]\cos(2\pi t/12) \\
S(x) &= S_i, \text{ if } x \in \textsf{district}_i
\end{aligned}
$$

Intensity of the point process

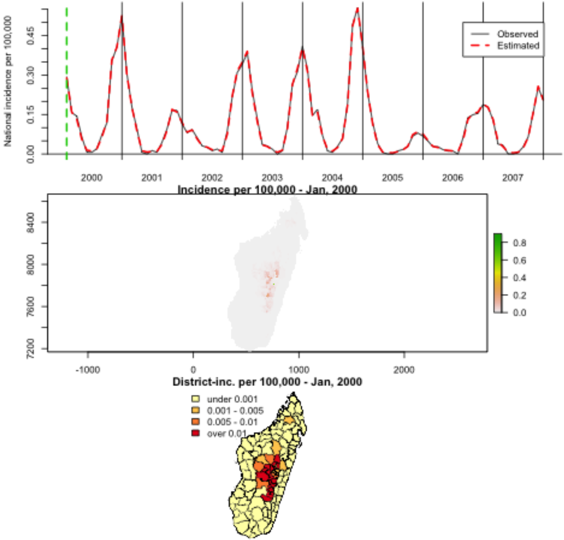
- $m(x,t)$ = ``number of susceptibles at location x and time t''
- $\lambda(x,t) = m(x,t)\exp\{d(x,t)^\top\beta + S(x) + V(t)\}$,

$$
\begin{aligned}
d(x,t)^\top\beta &= \beta_1 + \beta_2\text{temp}(x,t) + \beta_3\text{rain}(x,t) + \\
&\quad \beta_4 I(\text{elev}(x) > 800m) + \beta_5\text{elev}(x) + \\
&\quad \beta_6\text{elev}(x) \times I(\text{elev}(x) > 800m) + \\
&\quad \beta_7 \log\{\text{pop}(x)\} + \\
&\quad [\beta_8 + \beta_9\text{temp}(x,t) + \beta_{10}\text{rain}(x,t)]\sin(2\pi t/12) + \\
&\quad [\beta_{11} + \beta_{12}\text{temp}(x,t) + \beta_{13}\text{rain}(x,t)]\cos(2\pi t/12) \\
S(x) &= S_i, \text{ if } x \in \text{district}_i \\
V(t) &\mid V(t-1), \ldots, V(0) = V(t) \mid V(t-1)
\end{aligned}
$$

Results (1)

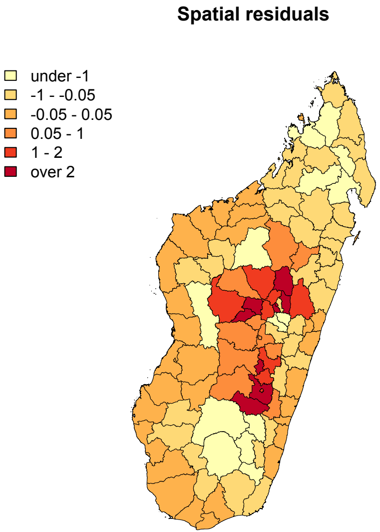
Term	Estimate	95% CI
β_1	-41.301	(-54.246, -30.104)
β_2	0.055	(-0.008, 0.125)
β_3	0.028	(-0.009, 0.061)
β_4	22.241	(10.566, 35.322)
β_5	0.027	(0.011, 0.043)
β_6	-0.024	(-0.041, -0.008)
β_7	0.069	(0.014, 0.140)
β_8	0.735	(-0.103, 1.444)
β_9	-0.293	(-0.314, -0.269)
β_{10}	0.025	(-0.003, 0.053)
β_{11}	5.017	(4.513, 5.498)
β_{12}	0.040	(0.007, 0.078)
β_{13}	-0.026	(-0.063, 0.013)

Results (2)



National incidence per 100,000
0.00
0.15
0.30
0.45
Observed
Estimated
2000
2001
2002
2003
2004
2005
2006
2007
Incidence per 100,000 - Jan, 2000
8400
8000
7600
7200
-1000
0
1000
2000
0.8
0.6
0.4
0.2
0.0
District-inc. per 100,000 - Jan, 2000
under 0.001
0.001 - 0.005
0.005 - 0.01
over 0.01

Results (3)



Discussion

Discussion

- How to account for spatio-temporally structured bias?

Discussion

- How to account for spatio-temporally structured bias?
- Alternative modelling approaches to zero-inflation.
  1. Admitting discontinuities in prevalence.
  2. Constraining prevalence to smoothly approach zero.

Discussion

- How to account for spatio-temporally structured bias?

- Alternative modelling approaches to zero-inflation.
  1. Admitting discontinuities in prevalence.
  2. Constraining prevalence to smoothly approach zero.

- Comparison of the mechanistic approach with empirical models.

Discussion

- How to account for spatio-temporally structured bias?
- Alternative modelling approaches to zero-inflation.
  1. Admitting discontinuities in prevalence.
  2. Constraining prevalence to smoothly approach zero.
- Comparison of the mechanistic approach with empirical models.
- Acknowledgements: Gillian Stresman (LSHTM), Jennifer Stevenson (LSHTM), Hans Remme (Ornex), Cyril Caminade (Liverpool), Matthew Baylis (Liverpool).

Discussion

- How to account for spatio-temporally structured bias?
- Alternative modelling approaches to zero-inflation.
  1. Admitting discontinuities in prevalence.
  2. Constraining prevalence to smoothly approach zero.
- Comparison of the mechanistic approach with empirical models.
- Acknowledgements: Gillian Stresman (LSHTM), Jennifer Stevenson (LSHTM), Hans Remme (Ornex), Cyril Caminade (Liverpool), Matthew Baylis (Liverpool).

Thanks for your attention.