Prevalence mapping

Peter Diggle & Emanuele Giorgi

Lancaster Medical School, Lancaster University, Lancaster, UK

Model-based geostatistics: geospatial statistical methods for public health applications, 5-9 October 2015

Overview

1. Prevalence mapping using linear models and binomial models.
2. Plug-in and Bayesian predictions.
3. Predictive summaries: prevalence, odds and exceedance probabilities.
4. Combining data from multiple surveys (continue).

Geostatistical prediction

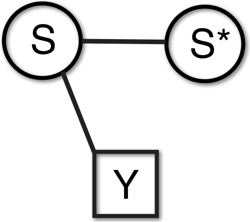
- $\mathcal{S} = \{S(x) : x \in \mathbb{A} \subset \mathbb{R}^2\}$
- $X = \{x_1, \ldots, x_n\}$ (sampled locations).
- $X^* = \{x_1^*, \ldots, x_q^*\}$ (prediction locations).

Geostatistical prediction

- $\mathcal{S} = \{S(x) : x \in \mathbb{A} \subset \mathbb{R}^2\}$
- $X = \{x_1, \ldots, x_n\}$ (sampled locations).
- $X^* = \{x_1^*, \ldots, x_q^*\}$ (prediction locations).
- $S = (S(x_1), \ldots, S(x_n))$.
- $S^* = (S(x_1^*), \ldots, S(x_q^*))$.

Geostatistical prediction

- $\mathcal{S} = \{S(x) : x \in \mathbb{A} \subset \mathbb{R}^2\}$
- $X = \{x_1, \ldots, x_n\}$ (sampled locations).
- $X^* = \{x_1^*, \ldots, x_q^*\}$ (prediction locations).
- $S = (S(x_1), \ldots, S(x_n))$.
- $S^* = (S(x_1^*), \ldots, S(x_q^*))$.



$$[S, S^*, Y] = [Y][S|Y][S^*|S, Y] = [Y][S|Y][S^*|S]$$

- **Linear model based on logit transformed prevalence.**

$$\begin{aligned} Z_i &= \log\{(Y_i + 0.5)/(m_i - Y_i + 0.5)\} \\ &= d(x_i)^\top \beta + S(x_i) + U_i \end{aligned}$$

- **Linear model based on logit transformed prevalence.**

$$
\begin{aligned}
Z_i &= \log\{(Y_i + 0.5)/(m_i - Y_i + 0.5)\} \\
&= d(x_i)^\top \beta + S(x_i) + U_i \\
[T, T^*, Z] &= [Z][T|Z][T^*|T]
\end{aligned}
$$

Prevalence mapping: a general framework

- **Linear model based on logit transformed prevalence.**

$$
\begin{aligned}
Z_i &= \log\{(Y_i + 0.5)/(m_i - Y_i + 0.5)\} \\
&= d(x_i)^\top \beta + S(x_i) + U_i \\
[T, T^*, Z] &= [Z][T|Z][T^*|T]
\end{aligned}
$$

- $T = (d(x_1)^\top \beta + S(x_1), \ldots, d(x_n)^\top \beta + S(x_n))$

- **Linear model based on logit transformed prevalence.**

$$\begin{aligned} Z_i &= \log\{(Y_i + 0.5)/(m_i - Y_i + 0.5)\} \\ &= d(x_i)^\top \beta + S(x_i) + U_i \\ [T, T^*, Z] &= [Z][T|Z][T^*|T] \end{aligned}$$

- $T = (d(x_1)^\top \beta + S(x_1), \ldots, d(x_n)^\top \beta + S(x_n))$
- $T^* = (d(x_1^*)^\top \beta + S(x_1^*), \ldots, d(x_n^*)^\top \beta + S(x_q^*))$

- **Linear model based on logit transformed prevalence.**

$$\begin{aligned} Z_i &= \log\{(Y_i + 0.5)/(m_i - Y_i + 0.5)\} \\ &= d(x_i)^\top \beta + S(x_i) + U_i \\ [T, T^*, Z] &= [Z][T|Z][T^*|T] \end{aligned}$$

- $T = (d(x_1)^\top \beta + S(x_1), \ldots, d(x_n)^\top \beta + S(x_n))$
- $T^* = (d(x_1^*)^\top \beta + S(x_1^*), \ldots, d(x_n^*)^\top \beta + S(x_q^*))$
- $T^*|T \sim$ MVN(μ^*, Σ^*).

$$\begin{aligned} \mu^* &= D^* \beta + C\Sigma_T^{-1}(T - D\beta) \\ \Sigma^* &= \Sigma_{T^*} - C\Sigma_T^{-1}C^\top \end{aligned}$$

- **Linear model based on logit transformed prevalence.**

$$\begin{aligned} Z_i &= \log\{(Y_i + 0.5)/(m_i - Y_i + 0.5)\} \\ &= d(x_i)^\top \beta + S(x_i) + U_i \\ [T, T^*, Z] &= [Z][T|Z][T^*|T] \end{aligned}$$

- $T = (d(x_1)^\top \beta + S(x_1), \ldots, d(x_n)^\top \beta + S(x_n))$
- $T^* = (d(x_1^*)^\top \beta + S(x_1^*), \ldots, d(x_n^*)^\top \beta + S(x_q^*))$
- $T^*|T \sim \text{MVN}(\mu^*, \Sigma^*)$.

$$\begin{aligned} \mu^* &= D^* \beta + C \Sigma_T^{-1} (T - D\beta) \\ \Sigma^* &= \Sigma_{T^*} - C \Sigma_T^{-1} C^\top \end{aligned}$$

How to estimate prevalence?

Prevalence mapping: plug-in and Bayesian prediction

- $(T^*_{(1)}, \ldots, T^*_{(B)}) = B$ simulated samples from $[T|T^*]$

Prevalence mapping: plug-in and Bayesian prediction

- $(T^*_{(1)}, \ldots, T^*_{(B)}) = B$ simulated samples from $[T|T^*]$

$$\hat{P}(x^*_j) = \frac{1}{B} \sum_{i=1}^{B} \frac{\exp\{T_{(i)}(x^*_j)\}}{1 + \exp\{T_{(i)}(x^*_j)\}}$$

- $\theta = (\beta, \sigma^2, \phi, \tau^2)$

Prevalence mapping: plug-in and Bayesian prediction

- $(T^*_{(1)}, \ldots, T^*_{(B)}) = B$ simulated samples from $[T|T^*]$

$$\hat{P}(x^*_j) = \frac{1}{B} \sum_{i=1}^{B} \frac{\exp\{T_{(i)}(x^*_j)\}}{1 + \exp\{T_{(i)}(x^*_j)\}}$$

- $\theta = (\beta, \sigma^2, \phi, \tau^2)$
- **Plug-in prediction:**

$$[T^*|T] = [T^*|T; \hat{\theta}].$$

Prevalence mapping: plug-in and Bayesian prediction

- $(T^*_{(1)}, \ldots, T^*_{(B)}) = B$ simulated samples from $[T|T^*]$

$$\hat{P}(x_j^*) = \frac{1}{B}\sum_{i=1}^{B} \frac{\exp\{T_{(i)}(x_j^*)\}}{1+\exp\{T_{(i)}(x_j^*)\}}$$

- $\theta = (\beta, \sigma^2, \phi, \tau^2)$
- **Plug-in prediction:**

$$[T^*|T] = [T^*|T; \hat{\theta}].$$

- **Bayesian prediction:**

$$[T^*|T] = \int [T^*|T, \theta][\theta]\, d\theta.$$

Predictive summaries

- **Prevalence.** $P(x) = \exp\{T(x)\}/(1 + \exp\{T(x)\})$.

Predictive summaries

- **Prevalence.** $P(x) = \exp\{T(x)\}/(1 + \exp\{T(x)\})$.
- **Odds.** $O(x) = \exp\{T(x)\}$.

Predictive summaries

- **Prevalence.** $P(x) = \exp\{T(x)\}/(1 + \exp\{T(x)\})$.
- **Odds.** $O(x) = \exp\{T(x)\}$.
- **Excedance probabilities.**

$$\text{Prob}(P(x) > p|y)$$

Predictive summaries

- **Prevalence.** $P(x) = \exp\{T(x)\}/(1 + \exp\{T(x)\})$.
- **Odds.** $O(x) = \exp\{T(x)\}$.
- **Excedance probabilities.**

$$\text{Prob}(P(x) > p|y) = \text{Prob}(O(x) > p/(1-p)|y)$$

Predictive summaries

- **Prevalence.** $P(x) = \exp\{T(x)\}/(1 + \exp\{T(x)\})$.
- **Odds.** $O(x) = \exp\{T(x)\}$.
- **Excedance probabilities.**

$$\text{Prob}(P(x) > p|y) = \text{Prob}(O(x) > p/(1-p)|y)$$

- **Decomposition of explanatory variables and spatial residuals effects.**

$$O(x) \quad = \quad \exp\{T(x)\}$$

Predictive summaries

- **Prevalence.** $P(x) = \exp\{T(x)\}/(1 + \exp\{T(x)\})$.
- **Odds.** $O(x) = \exp\{T(x)\}$.
- **Excedance probabilities.**

$$\text{Prob}(P(x) > p|y) = \text{Prob}(O(x) > p/(1-p)|y)$$

- **Decomposition of explanatory variables and spatial residuals effects.**

$$\begin{aligned} O(x) &= \exp\{T(x)\} \\ &= \exp\{d(x)^\top \beta + S(x)\} \end{aligned}$$

Predictive summaries

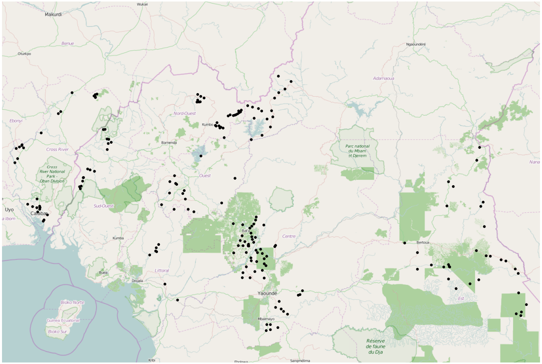
- **Prevalence.** $P(x) = \exp\{T(x)\}/(1 + \exp\{T(x)\})$.
- **Odds.** $O(x) = \exp\{T(x)\}$.
- **Excedance probabilities.**

$$\text{Prob}(P(x) > p|y) = \text{Prob}(O(x) > p/(1-p)|y)$$

- **Decomposition of explanatory variables and spatial residuals effects.**

$$\begin{aligned} O(x) &= \exp\{T(x)\} \\ &= \exp\{d(x)^\top\beta + S(x)\} \\ &= \exp\{d(x)^\top\beta\}\exp\{S(x)\} \end{aligned}$$

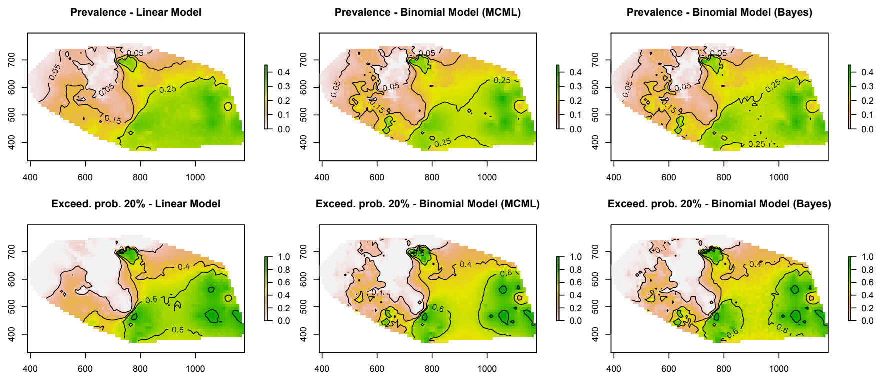
Loa loa prevalence mapping (1)



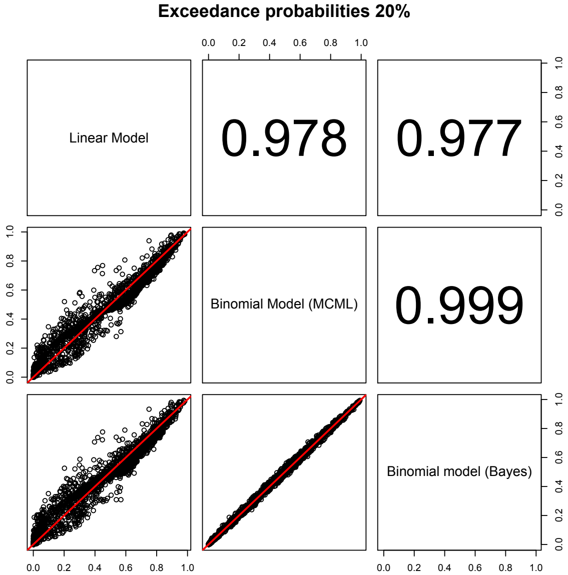
Epidemiological and geostistical questions

- What are the main risk factors of Loa loa?
- How do we identify Loa-loa hotspots?

Loa loa prevalence mapping (2)



Loa loa prevalence mapping (3)

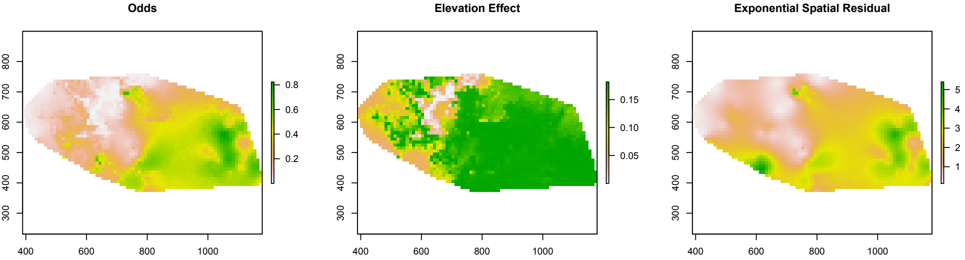


Effects decomposition

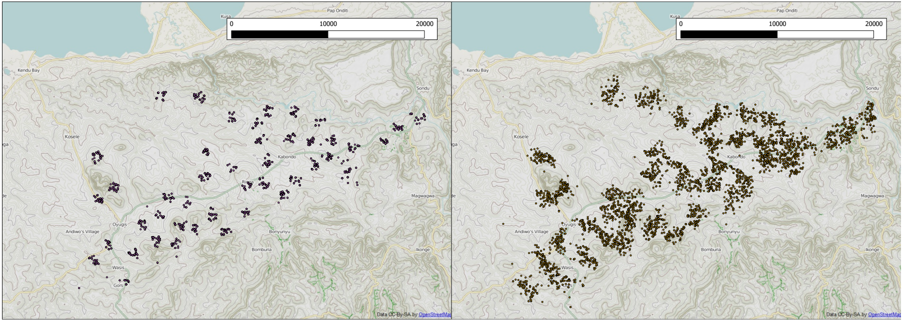
- $d(x)^\top \beta = \beta_0 + \beta_1 \mathsf{elev}(x) + \beta_2 \mathsf{elev}^2(x)$

Effects decomposition

- $d(x)^\top \beta = \beta_0 + \beta_1 \text{elev}(x) + \beta_2 \text{elev}^2(x)$



Combining information from multiple surveys



- **Community survey:** 1430 individuals; 740 compounds.
- **School survey:** 4852 pupils (46 schools); 3791 compounds.

Model formulation

- **Community survey.**

$$\log\{p_{ij}/(1-p_{ij})\} = d_{ij}^{\top}\beta + S(x_i) + Z_i.$$

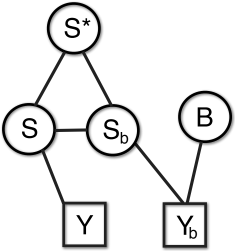
Model formulation

- **Community survey.**

$$\log\{p_{ij}/(1-p_{ij})\} = d_{ij}^\top \beta + S(x_i) + Z_i.$$

- **School survey.**

$$\log\{p_{ij}/(1-p_{ij})\} = d_{ij}^\top (\beta + \delta) + S(x_i) + B(x_i) + Z_i.$$



What is the gain in accuracy in estimating S^*?

Prevalence mapping using convenience samples

$$[S, S_b, S^*, B|Y, Y_b] = [B|Y_b][S|Y][S_b|Y_b][S^*|S, S_b]$$

Prevalence mapping using convenience samples

$$[S, S_b, S^*, B|Y, Y_b] = [B|Y_b][S|Y][S_b|Y_b][S^*|S, S_b]$$

- **Odds within the convenience samples.**

$$O_j(x) \quad = \quad \exp\{d_j^\top \beta + S(x)\} \exp\{d_j^\top \delta\} \exp\{B(x)\}$$

Prevalence mapping using convenience samples

$$[S, S_b, S^*, B|Y, Y_b] = [B|Y_b][S|Y][S_b|Y_b][S^*|S, S_b]$$

- **Odds within the convenience samples.**

$$O_j(x) = \exp\{d_j^\top \beta + S(x)\} \exp\{d_j^\top \delta\} \exp\{B(x)\}$$

(a) Odd spatial bias $= \exp\{B(x)\}$

Prevalence mapping using convenience samples

$$[S, S_b, S^*, B|Y, Y_b] = [B|Y_b][S|Y][S_b|Y_b][S^*|S, S_b]$$

- **Odds within the convenience samples.**

$$O_j(x) \;=\; \exp\{d_j^\top \beta + S(x)\} \exp\{d_j^\top \delta\} \exp\{B(x)\}$$

(a) Odd spatial bias $= \exp\{B(x)\}$

(b) $1 - P[0.9 < \exp\{B(x)\} < 1.1|Y, Y_b]$

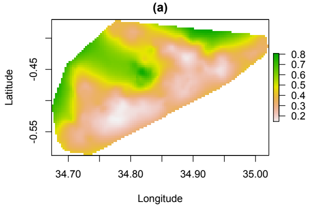
Prevalence mapping using convenience samples

$$[S, S_b, S^*, B|Y, Y_b] = [B|Y_b][S|Y][S_b|Y_b][S^*|S, S_b]$$

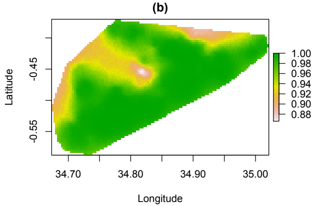
- **Odds within the convenience samples.**

$$O_j(x) \quad = \quad \exp\{d_j^\top \beta + S(x)\} \exp\{d_j^\top \delta\} \exp\{B(x)\}$$

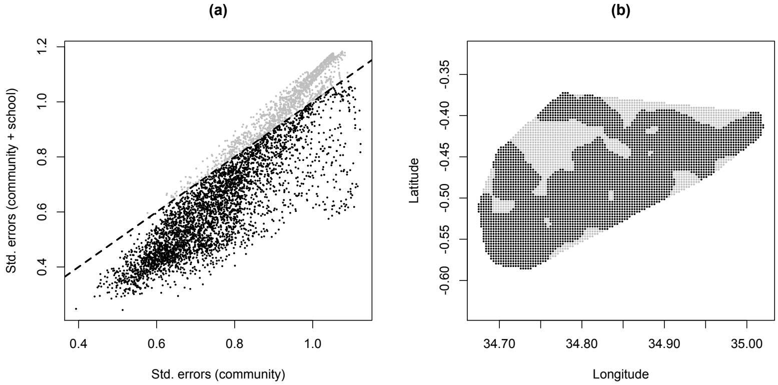
(a) Odd spatial bias $= \exp\{B(x)\}$



(b) $1 - P[0.9 < \exp\{B(x)\} < 1.1|Y, Y_b]$



Gain in efficiency



Bibliography

- Diggle, P. J., Giorgi, E. (2015). *Model-based geostatistics for prevalence mapping in low-resource settings.* Under review. Available at `http://arxiv.org/abs/1505.06891`
- Giorgi, E., Diggle, P. J. (2015). *PrevMap: and R package for prevalence mapping.* Under review.
- Giorgi, E., Sanie, S. S. S., Terloouw, D. J., Diggle, P. J. (2015). *Combining data from multiple spatially referenced prevalence surveys using generalized linear geostatistical models.* JRSS A, 178:445-464.