# Fitting large-scale Log-Gaussian Cox Process model with an efficient sampling algorithm

## and its application to malaria incidence mapping.

**Elizaveta Semenova & Penelope Vounatsou**

Swiss Tropical and Public Health Institute

elizaveta.semenova@swisstph.ch

Swiss TPH

Swiss Tropical and Public Health Institute
Schweizerisches Tropen- und Public Health-Institut
Institut Tropical et de Santé Publique Suisse

Associated Institute of the University of Basel

## Summary

We use hierarchical Bayesian modeling of the log-Gaussian Cox Process for high resolution spatio-temporal malaria incidence mapping. We propose an optimized version of sampling from the Multivariate Normal Distribution in the case when covariance matrix can be represented as a Kronecker product, established through tensor (Kronecker) calculus. This allows to reduce computation time and storage requirements. The technique is applied to malaria incidence data.

## Introduction

Malaria data is correlated in space and time which is related to the correlation structure of the exposure. Ignoring geographical dependence may lead to incorrect estimates of the effects of predictors and of the estimated predictions. Spatial models introduce an extra term for each observed location or area, the so-called spatial random effect. Log-Gaussian Cox Process (LGCP), is a rigorous method to analyze spatially random **point patterns** data.

## Main Objectives

1. Perform Bayesian point patterns analyses of malaria incidence data.
2. Assess the contribution of environmental risk factors to the geographical distribution of malaria at the elimination stage.
3. Exploit efficient algorithm of sampling from the Multivariate Normal Distribution with Kronecker-structured covariance matrix.
4. Produce high-resolution monthly risk maps of malaria in Swaziland.

## Data

### Incidence Data and Importation

- Malaria case data was obtained from the National Malaria Control Program of Swaziland. The data contains dates and geo-coded locations of case households and travel history allowing to distinguish between local and imported cases.

### Natural and man-made environmental and population data

- *Climatic and natural environmental data*: elevation, normalized difference vegetation index, enhanced vegetation index , normalized difference water index, topographic wetness index, day and night land surface temperature, Land Cover, rainfall, presence of water bodies (obtained from MODIS and downloaded through the Google Earth Engine [1] interface.)
- *Man-made environment*: urban extent, presence of primary, secondary, tertiary roads and railways, travel time to Mozambique, intensity of night light.

## Analysis of point patterns

Consider a random realization of a point process $S = \{s_1, \ldots, s_n\}$ - a *point pattern* in a study region $D$, characterized by the *intensity function* $\lambda(s)$. Conditional on $\lambda_t(s)$ the data is distributed as a Poisson process:

$$S \mid \lambda_t(s) \sim \mathrm{PP}(\lambda_t(s)).$$

The *Log-Gaussian Cox Process (LGCP)* suggests the intensity function of the form

$$\lambda_t(s) = \lambda_0 \exp\{\mathbf{x}^T(s,t)\boldsymbol{\beta} + f(s,t)\} = \exp\{GP(s,t)\},$$

Here $GP(s)$ denotes a Gaussian random field, which is completely specified by its first and second moments. Finite realizations of the stationary Gaussian Process compose a vector distributed as

$$f_{\text{finite}} \sim \mathrm{MVN}(m, \Sigma).$$

Here $\Sigma$ is a spatially structured covariance matrix with elements

$$\Sigma_{ij} = \sigma^2 R_{ij}.$$

Our choice of the correlation matrix is the squared-exponential functional form:

$$R_{ij} = exp\left(-\frac{1}{2l^2}\sum_{d=1}^{D}(x_i^d - x_j^d)^2\right)$$

resulting in the Kronecker structure of the covariance matrix. The observed number of cases within a space-time grid cell is modeled as

$$y_{ijt} \mid f_{ijt} \sim \mathrm{Pois}(|s_{ij}|\lambda_{ijt}).$$

## Sampling within the fitting procedure

To sample from the **the array-normal distribution**

$$f_{n,p,t} \sim \mathrm{AN}_{n,p,t}(\mu_{n,p,t}, \Sigma_n, \Sigma_p, \Sigma_T),$$

defined as $\mathrm{vec}(f_{n,p,t}) \sim \mathrm{MVN}(\mathrm{vec}(\mu_{n,p,t}), \Sigma_T \otimes \Sigma_p \otimes \Sigma_n)$, we use an efficient algorithm for matrix-vector multiplication, where the matrix can be represented as a Kronecker product

$$L_N = \otimes_{d=1}^{D} L_d, \quad N = \prod_{d=1}^{D} d.$$

Within each MCMC step the sampling is done via

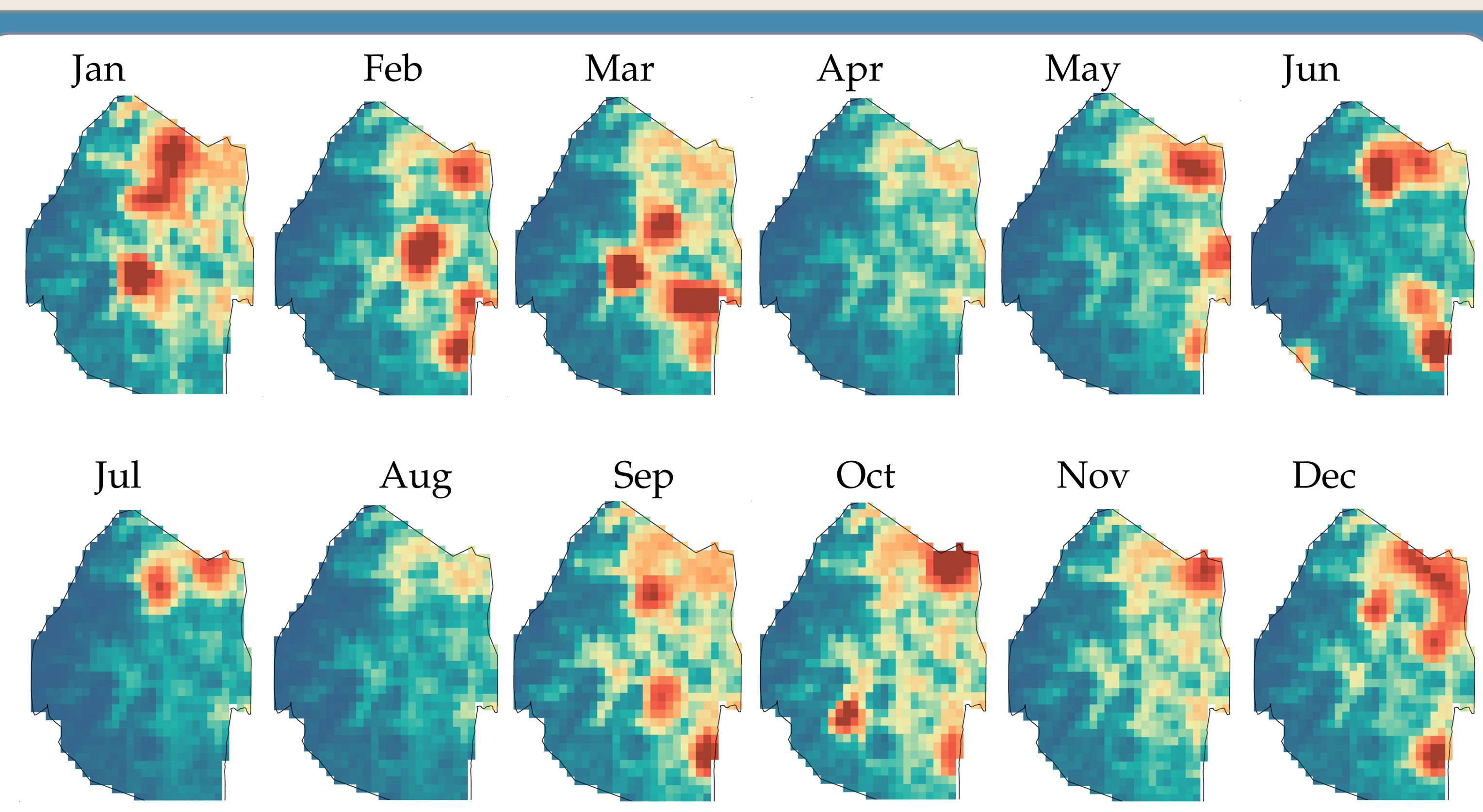
$$f = (\otimes L_d)\, z \sim \mathrm{MVN}(0, \Sigma_N), \text{ where } z \sim \mathrm{MVN}(0, \mathrm{I}_N).$$

## Achieved gains:

- **Storage:** $n^2 + p^2 + t^2$ instead of $n^2p^2t^2$ elements.
- **Complexity:** $O(n^3 + p^3 + t^3 + n^2pt + np^2t + npt^2)$ instead of $O(n^3p^3t^3)$ operations.

## Results

Our analyses showed that natural environment is only partially responsible for the local transmission with the man-made factors playing a game changing role in the dynamics.




Monthly maps of the predicted number of local cases per 5km$^2$ on the log-scale

| Covariate | Mean | SD | 2.5% | 50% | 97.5% |
|---|---|---|---|---|---|
| Intercept | -10.25 | 0.54 | -11.33 | -10.25 | -9.25 |
| Urban | -0.06 | 0.30 | -0.67 | -0.05 | 0.52 |
| Elevation | -1.22 | 0.20 | -1.63 | -1.22 | -0.83 |
| Distance to motor ways | 0.26 | 0.11 | 0.05 | 0.26 | 0.48 |
| Distance to primary and secondary roads | -0.10 | 0.14 | -0.38 | -0.10 | 0.18 |
| Distance to tertiary roads | -1.15 | 0.25 | -1.65 | -1.15 | -0.66 |
| Travel time to Mozambique | -1.52 | 0.16 | -1.84 | -1.52 | -1.22 |
| Distance to water bodies | -0.51 | 0.11 | -0.73 | -0.51 | -0.31 |
| NDVI | 0.06 | 0.13 | -0.20 | 0.06 | 0.32 |
| LSTD | 0.98 | 0.18 | 0.64 | 0.98 | 1.32 |
| Rainfall | 0.07 | 0.21 | -0.34 | 0.08 | 0.46 |
| Distance to the nearest imported case | -0.31 | 0.16 | -0.62 | -0.30 | -0.01 |

Estimates and credible intervals of the effects of the covariates on the spatio-temporal distribution of locally acquired cases

## References

[1] Google Earth Engine Team. Google earth engine: A planetary-scale geo-spatial analysis platform, 12 2015.

[2] Jesper Møller, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen. Log gaussian cox processes. *Scandinavian journal of statistics*, 25(3):451–482, 1998.

**Swiss TPH**, Socinstrasse 57, P.O. Box, 4002 Basel, Switzerland
T +41 61 284 81 11, F +41 61 284 81 01, www.swisstph.ch

University of Basel