Mapping Malaria Relative Risk in Colombia A Bayesian Approach Using Zero-Inflated Models and Intrinsic CAR Prior Specification

Jose E. Gomez[1], **Marlon R. Ruiz**[2] and **Carlos E. Melo**[3]
Universidad Distrital Francisco José de Caldas
[1] jegomezg@correo.udistrital.edu.co
[2] mrruizf@correo.udistrital.edu.co
[3] cmelo@udistrital.edu.co
UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

CEB 2019 EIB VALÈNCIA

Introduction

Although the mortality rate is not high Colombia was one of the first countries where resistance to chloroquine-based treatment was reported, which justified all kinds of studies related to this disease, such as characteristics of the parasite, as well as social phenomena related to malaria transmission. In this paper, we try to describe the overall scene of the environmental factors that affect the behavior of the disease taking into account the zero-inflated effect in the spatial modeling for a comprehensive assessment of the relative risk factors associated with the physical phenomena of each geographical area of the country. For this purpose, we explore the general additive models in the Bayesian hierarchical framework and implement the integrated nested Laplace approximation as an alternative to the classic simulation methods.

The Data

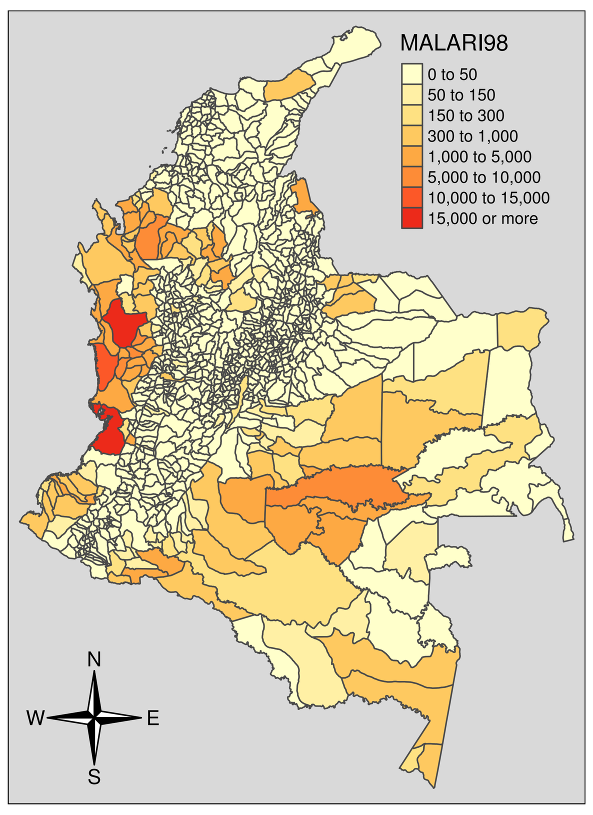


Fig. 1: Registered malaria cases in Colombia for 1998.

For the study area of Colombia, we analyze municipalities-level data of malaria cases for the year of 1998 registered by the Colombian National Institute of Health (INS, in Spanish). Even though the country presents a historical high of cases in this period of time with 185.490 cases registered, more than 83% of the records in the data are zero, the maximum of the registered cases in one municipality is up to 22.489.

To analyze the global impact of environmental factors three co-variables are included in the model: precipitation (millimeter per month), temperature (Celsius degrees) which are calculated as multi-annual averages by the Institute of Hydrology, Meteorology and Environmental Studies of Colombia (IDEAM, in Spanish); and the normalized difference vegetation index (NDVI) extracted from LANDSAT images.

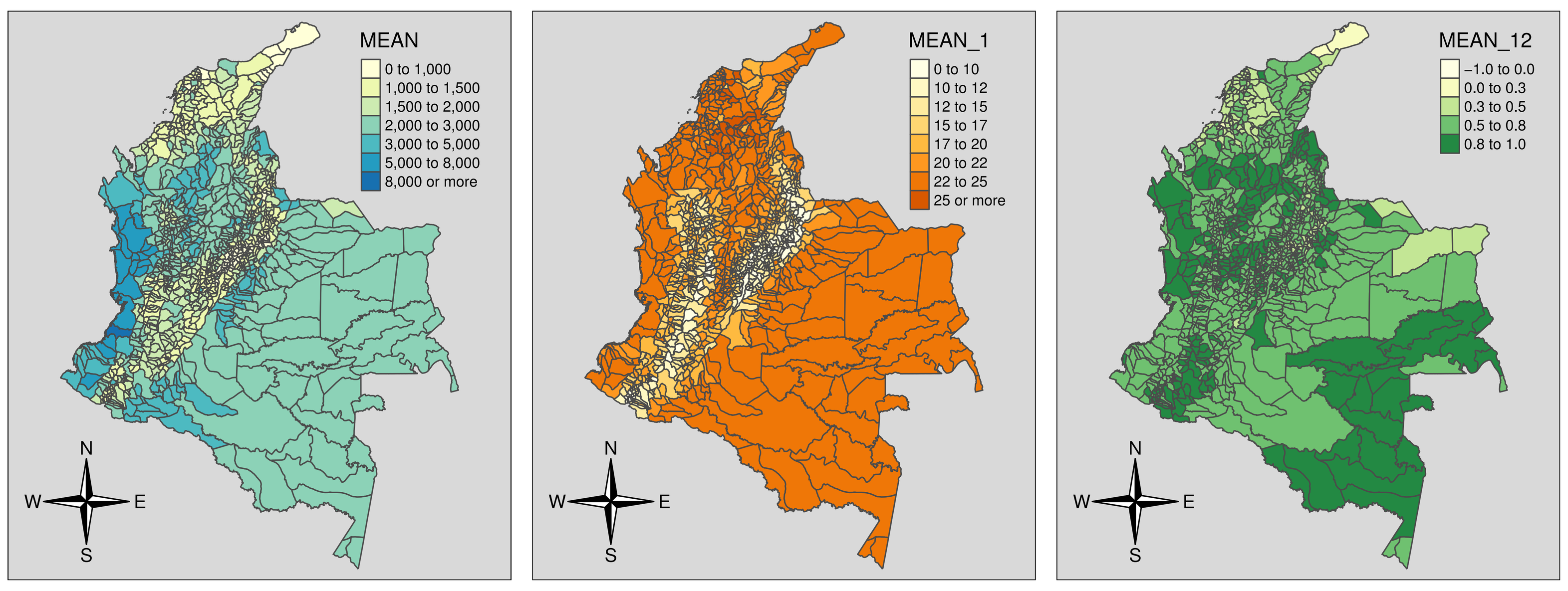


Fig. 2: Co-variable for the model, from the left: (1) precipitation, (2) temperature and (3) NDVI.

Bayesian Disease Mapping

If we consider a set of n contiguous areas with Y_k representing the number of cases, E_k the expected cases in each one and R_k the relative risk, based on the general form of hierarchical Bayesian models for risk estimate we have:

$$\begin{aligned} Y_k|E_kR_k &\sim F(E_kR_k) \\ Ln(R_k) &= \zeta + \beta X_k + \Phi_k \end{aligned} \tag{1}$$

Where the risk is explained through an intercept ζ that can be understood as the general risk above the territory, a set of co-variables X_k and its respective coefficient β_k which stands for the relative risk factors associated with each co-variable and a structured random effect Φ_k that captures the spatial correlation presented in the data.

The Model

In order to define the model, it is necessary to select an appropriated functional form ($F(E_kR_k)$ in (1)) as well as the structured random effect prior specification that will describe the spatial correlation in the data (Φ_k in (1)). For this study, given the data conditions, namely zero inflation and over-dispersion, the functional form implemented to overcome these terms is the Zero-Inflated Negative Binomial (ZINB) model, given by:

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1-\pi_i)f(0) & y_i = 0 \\ (1-\pi_i)f(Y_i) & y_i > 0 \end{cases} \tag{2}$$

$$f(y_i) = \frac{\Gamma(y_i+\alpha)}{y_i!\Gamma(\alpha)}\left(\frac{\alpha}{\alpha+R_i}\right)^{\alpha}\left(\frac{R_i}{\alpha+R_i}\right)^{y_i}$$

On the other hand, as we explore the empirical Bayesian method of integrated nested Laplace approximation (INLA) for the model estimation, we assumed a simple structured random effect prior that is the Besag prior or intrinsic conditional auto-regressive (iCAR) model:

$$\Phi_k|\Phi_{-k}, Q, \tau_l^2 \sim N\left(\frac{1}{n_k}\sum_{j\sim k}\Phi_j, \frac{\tau_l^2}{n_k}\right) \tag{3}$$

Results

We use a well-known risk estimator, namely, the Standardized Incidence Ratio (SIR), as a point of reference to compare our results. We found decent coherence between the model and the classic risk estimator, nonetheless it is clear that the ZINB model furnish a smoother result allowing to discriminate big areas with similar probability to present a certain factor risk unlike SIR that that seems to present some degree of bias due to the individual areas with high counts, showing high contrasts even among neighboring municipalities with almost identical characteristics that should represent similar risk throughout the territory.

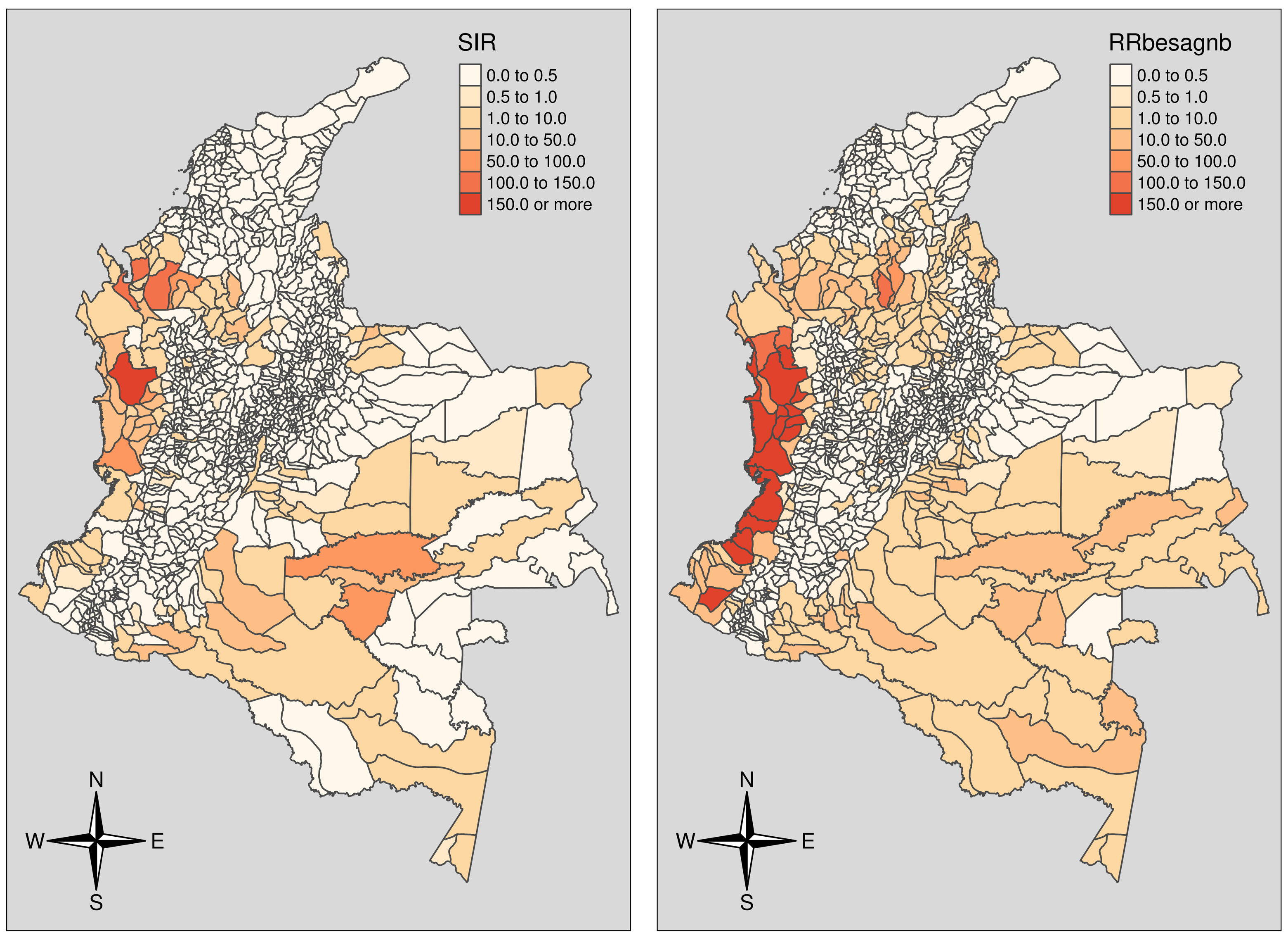
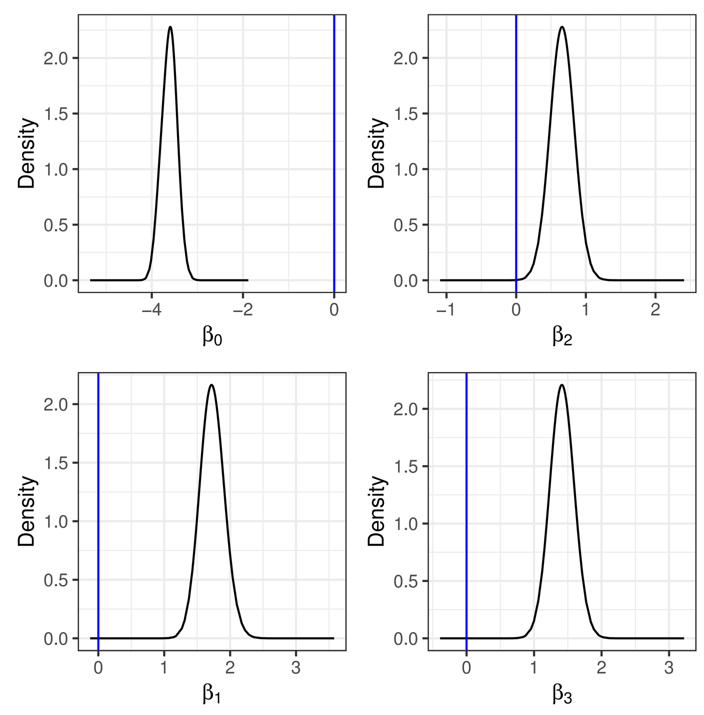


Fig. 3: Risk map estimated with ZINB model.

The risk map obtained enables to clearly identify risk-based subregions in the country, for example, the Pacific region, including the Choco department and the coastal regions of Nariño, Cauca and Valle del Cauca departments, which present by far the higher relative risk of malaria infection in the country, it is the thin zone between the Pacific ocean and the mountain complex where most of the cases where registered and the environmental factors propitiate the presence of malaria.

There is also the eastern plains zone comprehended by the Meta, Caqueta, Guaviare, Amazonas and Casanare departments, we find a very extensive geographical area with equal medium risk across the territory. Finally, the is the mountainous region that crosses the center of the country where the risk is minimum or non-existent, it corresponds with the zone with a large number of zero infection counts.

The model also allows to determinate the relative risk factors associated with each co-variable, which show to be, each one of them, significant within the model, we find a positive relation with temperature and precipitation, conditions that favor the presence of the disease, that is hot and wet areas such as the coast and low height plains. In the case of NDVI, the model suggests that areas with a high presence of vegetation are not associated with an increase in the risk of contracting the disease, this can be explained for the Andean region which has almost none registered case despite presenting the same high NDVI values found in endemic zones.



Variable	Coefficient	0.025	0.975
Intercept	-3.617	-3.960	-3.285
Precipitation	1.728	1.374	2.101
NDVI	0.659	0.316	1.003
Temperature	1.415	1.063	1.773

Fig. 4: Relative risk factors associated with the co-variables.

Conclusions

- Negative Binomial Zero-inflated models provide a powerful approach to analyze the overall behavior of a disease that changes its intensity over extensive geographical areas where the environmental characteristics may oscillate sharply and quantify their impact on the risk of infection.
- The risk map obtained by using the ZINB model shows to be a smoother risk estimation compared to the classic methods such as Standardized Infection Rate giving a more comprehensive understanding of the high-risk zones, which allows formulating adequate public health policies.
- Colombia present high risk of malaria infection due to its tropical characteristics which is enhanced in the coastal zones and reduced almost to the nullity in the mountainous area generating high contrasts in the epidemiological behavior of the disease.

Bibliography

- Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology.*
- Lee, D. (2011). *A comparison of conditional autoregressive models used in Bayesian disease mapping.*
- Moraga, P. (2018) *Small area disease risk estimation and visualization using r*
- Olano, V. A., Brochero, H. L., Sáenz, R., Quiñones, M. L., y Molina, J. A. (2001). *Mapas preliminares de la distribución de especies de anopheles vectores de malaria en Colombia*
- Rue, H., Martino, S., y Chopin, N. (2009). *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations*