

A joint spatial model for dengue and severe dengue in Medellin, Colombia.
Supplementary Material

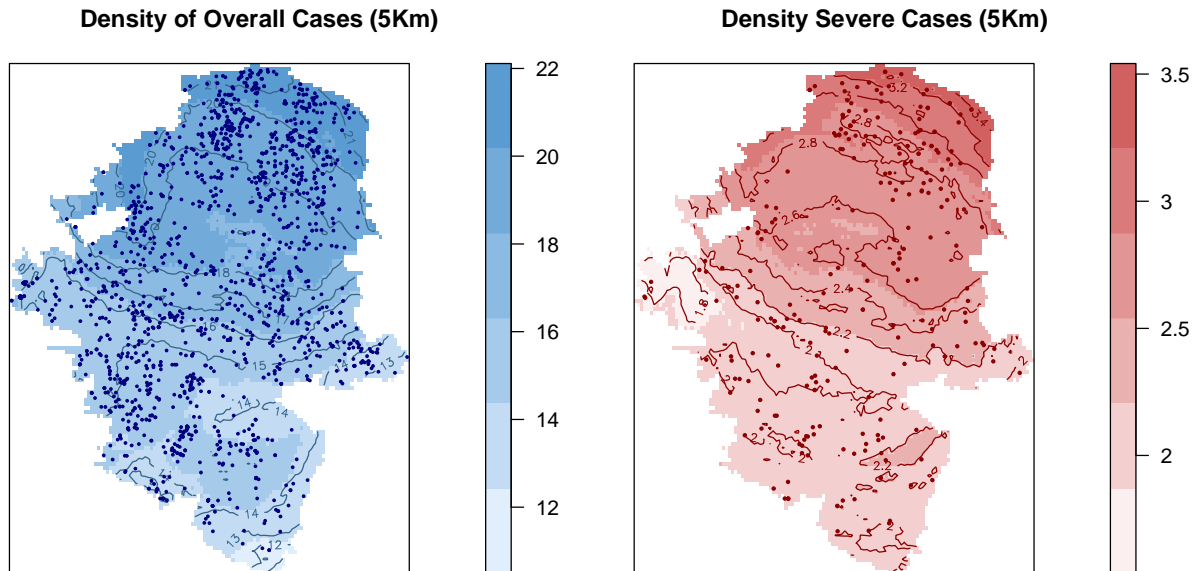


Figure S1. Distribution of dengue cases notified in Medellin in 2013. Kernel density of overall and severe dengue cases using a 5 Km bandwidth. Label indicates the number of overall and severe cases per 5 Km².

1. Spatial point process model:

A log-Gaussian Cox point process model assesses the distribution of the individual location of the outcome points (dengue cases) in a spatial structure, and it is used in order to consider both, observed and unobserved variation in the assessment of such distribution¹. It estimates the spatial distribution of the individual cases as a function of a continuous latent Gaussian random field, assuming conditional independence of the points presented on the field. This indicates that conditional on the latent field, the distribution of the point pattern (dengue cases), follows a Poisson process¹⁻³. This analysis uses individual level information and allows covariates at the spatial level to vary according to the random field. Thus, providing information about the presence and degree of clustering within the spatial structure, while considering simultaneously the spatial autocorrelation between and within spatial units⁴.

A Marked log-Gaussian Cox point process model uses an individual characteristic of the point, the ‘mark’, to assess the individual distribution of an event given such specific characteristic. In our case, each point represents an individual case, and the “mark” is the presence/absence of severity for each dengue case. Thus, the information about the “mark” is used to estimate two aspects simultaneously: 1) the overall distribution of the point pattern (i.e: all points/cases) and 2) the distribution of the point pattern given the mark (i.e: severe cases). The model could also be considered a “labeling” of the Poisson process, where the ‘marks’ work also as a response variable². This approach is intended to identify whether there is an underlying mechanism leading to a differential distribution of severe cases, with the added advantage of modelling simultaneously an individual feature of the point (the severity) and the spatial distribution, while accounting for its dependence^{2,3}.

To conduct a point process analysis, it is necessary to consider each case-location as $x_i: i = 1, \dots, n$, where, x indicates the location and i indicates the dengue case identifier that in theory could have occurred in any location inside a given spatial region $A \subset \mathfrak{R}^2$ ¹. Likewise, it is important to accept two main structural assumptions: i) that the function of the Cox-Process is a stochastic non-negative process $\Lambda = \{ \Lambda(x): x \in \mathfrak{R}^2 \}$, and ii) that conditional on the realization $\Lambda(x) = \lambda(x): x \in \mathfrak{R}^2$, the point process is an inhomogeneous Poisson process with intensity $\lambda(x)$. Also, it is necessary to consider the distribution of dengue cases as a phenomenon $S(x)$ that is incompletely observed and spatially continuous, given that $S = \{S(x) : x \in \mathfrak{R}^2\}$ is a Gaussian stochastic process and that S determines $\lambda(x)$, which is the intensity of the distribution; $\lambda(x) = \exp\{S(x)\}$ ^{1,3}. To analyze this intensity, it is necessary to approximate the spatially continuous random field to a constructed grid^{1,3,5}. Then, considering $\{y_i\}$ the observed number of points in the neighborhood, we assume that the number of points (cases) in a grid-cell/neighborhood i follows a Poisson distribution conditional on a first latent field, $\eta_i^{(1)}$:

$$y_i | \eta_i^{(1)} \sim Po(E_i \eta_i^{(1)}), \text{ Equation (S.1)}$$

The offset of the pattern E_i is specified as the expected count of cases in each neighborhood and obtained via indirect standardization⁶. To model the marked point process, we add to equation 1, the analysis of the marks (severe or not severe). For that we let m_i be the number of patients with severe dengue in each spatial unit i . Then, conditional on the value of a second latent field $\eta_i^{(2)}$ in the same neighborhood, m_i follows a binomial distribution:

$$m_i | \eta_{ij}^{(2)} \sim \text{Bin}(y_i, p_{ij}), \text{ Equation (S.2)}$$

where $p_{ij} = \exp(\eta_{ij}^{(2)}) / (1 + \exp(\eta_{ij}^{(2)}))$ is the probability of individual j being a severe case in a given neighborhood i while y_i is the total number of cases of dengue in neighborhood i . This, constituting a matrix outcome of two links (i.e: Poisson for overall point patterns, and Binomial for severity), each one on a separate latent field η , that are jointly analyzed in relation to a vector of sociodemographic covariates². We constructed the final model for each latent field $\eta_i^{(1)}$ and $\eta_{ij}^{(2)}$, including empirical covariates for neighborhood level and individual level as fixed effects, and spatial structures as random effects as follows:

$$\log(\eta_i^{(1)}) = \beta_0^{(1)} + \beta_1^{(1)}IB(s_i) + \beta_2^{(1)}UNDER20(s_i) + \beta_3^{(1)}P.FEMALE(s_i) + \beta_4^{(1)}SES(s_i) + f_s^j(s_i) + u(s_i), \text{ Equation (S.3)}$$

$$\text{logit}(\eta_{ij}^{(2)}) = \beta_0^{(2)} + \beta_1^{(2)}AGE(s_{ij}) + \beta_2^{(2)}SEX(s_{ij}) + \beta_3^{(2)}INSURANCE(s_{ij}) + \beta_4^{(2)}DISTANCEKm(s_{ij}) + g_s^j(s_i) + v(s_i), \text{ Equation (S.4)}$$

$\beta_0^{(1)}$ and $\beta_0^{(2)}$ are the pattern and marks intercepts. $\beta_1^{(1)} \dots \beta_4^{(1)}$ are the coefficients associated to the empirical covariates for the distribution of cases at the neighborhood level; and $\beta_1^{(2)} \dots \beta_4^{(2)}$ are the coefficients associated to the empirical covariates for severity at the individual level, as described in the main text. $DISTANCEKm(s_{ij})$ is the standardized minimum nearest-neighbor Euclidean distance between overall and severe cases, parameterized as a continuous variable. The

components $f_s^j(s_{ij})$ and $g_s^j(s_{ij})$ are the Gaussian Markov Random Field (GMRF), reflecting separately the spatial autocorrelation in the latent field, working as spatially structured effects for the pattern and the marks, respectively. $u(s_i)$ is the spatially unstructured random effect for the pattern and $v(s_i)$ is the spatially unstructured random effect for the marks. To express the dependence between the pattern and the marks, we used a single (common) random field replacing equation (4) as follows:

$$\text{logit}(\eta_{ij}^{(2)}) = \beta_0^{(2)} + \beta_1^{(2)} \text{AGE}(s_{ij}) + \beta_2^{(2)} \text{SEX}(s_{ij}) + \beta_3^{(2)} \text{INSURANCE}(s_{ij}) + \beta_4^{(2)} \text{DISTANCEKm}(s_{ij}) + \beta_s f_s^i(s_i) + v(s_i), \text{Equation (S.5)}$$

which makes the spatial effect for the severity proportional to the spatial effect of the pattern of case distribution². The spatially structured and unstructured effects were modeled following the Besag-York-Mollie (BYM) specification⁷; where after adjusting for the fixed effects, the structured component is modeled using an intrinsic conditional autoregressive structure (iCAR) and the unstructured effect is modeled using an independent prior^{6,8,9}.

1.1 Fitting separated models for each random field

Table S1a. Posterior mean of the **Incidence Rate Ratio** and 95% credible intervals for covariates (fixed effects) on the single-separated model for **overall dengue** cases in Medellin, 2013

Characteristic	IRR	2.5%	97.5%
(Intercept)	0.78	0.60	0.42
Proportion of Female cases	1.07	1.01	1.12
Proportion of cases <20 years old	1.06	1.01	1.12
Medium Breteau Index	1.10	0.84	1.44
Low SES	0.43	0.33	0.56
High SES	0.77	0.55	1.07

Proportion of female cases and proportion of cases under 20 years old indicate 10% increase in the proportion of cases per neighborhood. SES reference group = Medium SES; Breteau Index reference group = Low.

Table S1b. Posterior mean and 95% credible intervals of Hyperparameters:

Parameter	Mean	SD	2.5%	97.5%	mode
Precision for $f_s^j(s_{ij}) + u_{(s_{ij})}$	1.69	0.26	1.23	2.28	1.63
Phi for $u_{(s_{ij})}$	0.38	0.14	0.13	0.68	0.34

Table S2a. Posterior mean of the Odds Ratio and 95% credible intervals for covariates (fixed effects) on the single-separated model for **severe dengue** cases in Medellin, 2013

Characteristic	OR	2.5%	97.5%
Age			
15-34 Years	1.09	0.78	1.55
35-54 Years	1.08	0.74	1.59
>55 Years	1.49	0.98	2.27
Sex (Female)	0.92	0.72	1.19
Contributory Insurance	0.86	0.66	1.12
Distance to Severe cases (Km)	0.50	0.27	0.89

Age group reference= <15 years of age; Sex reference= Male; Health Insurance reference is Subsidized scheme.

Table S2b. Posterior mean and 95% credible intervals of Hyperparameter:

Parameter	Mean	SD	2.5%	97.5%	mode
Precision $g_s^j(s_{ij}) + v_{(s_{ij})}$	514.20	3018.98	12.39	3360.07	26.67
Phi for $v_{(s_{ij})}$	0.31	0.26	0.01	0.90	0.02

1.2 Sensitivity Analyses: Joint models

In addition to the full model, we fitted other joint models including a model without the SES covariate and the full joint model using a different spatial effect for severity. Models were examined using Deviance Information Criteria (DIC) and are summarized in Table S3 and results are presented in Table S4 and S5 respectively.

Table S3. Summary of Deviance Information Criterion (DIC) values and specification for joint models of overall (pattern) and severe (marks) dengue cases.

Model	Model Components	DIC
Full without the SES covariate.	$\log(\eta_i^{(1)}) = \beta_0^{(1)} + \beta_1^{(1)}IB(s_i) + \beta_2^{(1)}UNDER20(s_i) + \beta_3^{(1)}P.FEMALE(s_i) + f_s^j(s_i) + u(s_i);$ $\text{logit}(\eta_i^{(2)}) = \beta_0^{(2)} + \beta_1^{(2)}AGE(s_i) + \beta_2^{(2)}SEX(s_i) + \beta_3^{(2)}INSURANCE(s_i) + \beta_4^{(2)}DISTANCEKm(s_i) + \beta_s f_s^i(s_i) + v(s_i),$	2849.2
Full model with two separate spatial structures.	$\log(\eta_i^{(1)}) = \beta_0^{(1)} + \beta_1^{(1)}IB(s_i) + \beta_2^{(1)}UNDER20(s_i) + \beta_3^{(1)}P.FEMALE(s_i) + \beta_4^{(1)}SES(s_i) + f_s^j(s_i) + u(s_i);$ $\text{logit}(\eta_i^{(2)}) = \beta_0^{(2)} + \beta_1^{(2)}AGE(s_i) + \beta_2^{(2)}SEX(s_i) + \beta_3^{(2)}INSURANCE(s_i) + \beta_4^{(2)}DISTANCEKm(s_i) + g_s^{(2)}(s_i) + v(s_i)$	2868.1

Table S4a. Incidence Rate Ratio (IRR), Odds Ratios (OR) and 95% credible intervals of covariates (fixed effects) for the model without the SES covariates.

Fixed Effects	Estimates and 95% Credible intervals		
Overall Cases (Pattern)			
	IRR	2.5%	97.5%
Intercept (Pattern)	0.60	0.46	0.78
Proportion of Female Cases	1.04	1.00	1.09
Proportion of <20 years old	1.04	0.99	1.09
Breteau Index (Medium)	1.05	0.83	1.33
Severity (Marks)			
	OR	2.5%	97.5%
Sex (Female)	0.88	0.68	1.14
Age			
15-34 Years	0.97	0.69	1.38
35-54 Years	1.07	0.73	1.58
>55 Years	1.52	0.99	2.33
Contributory Insurance	1.00	0.77	1.31

Proportion of woman and cases under 20 years old indicate 10% increase. Breteau Index reference group = Low; Age group reference= <15 years of age; Sex reference= Male; Health Insurance reference is Subsidized scheme.

Table S4b. Posterior mean and 95% credible intervals of hyperparameters for the model without the SES covariates.

Parameter	mean	SD	2.5%	97.5%	mode
Precision for $f_s^j(s_i) + u(s_i)$	1.39	0.21	1.03	1.84	1.35
Phi for $u(s_i)$	0.41	0.14	0.17	0.69	0.39
β Coefficient for Severity for $g_s^j(s_{ij})$	-0.31	0.11	-0.53	-0.09	-0.32

Table S5a. Incidence Rate Ratio (IRR), Odds Ratios (OR) and 95% credible intervals of covariates (fixed effects) for the model with two separate spatial structures.

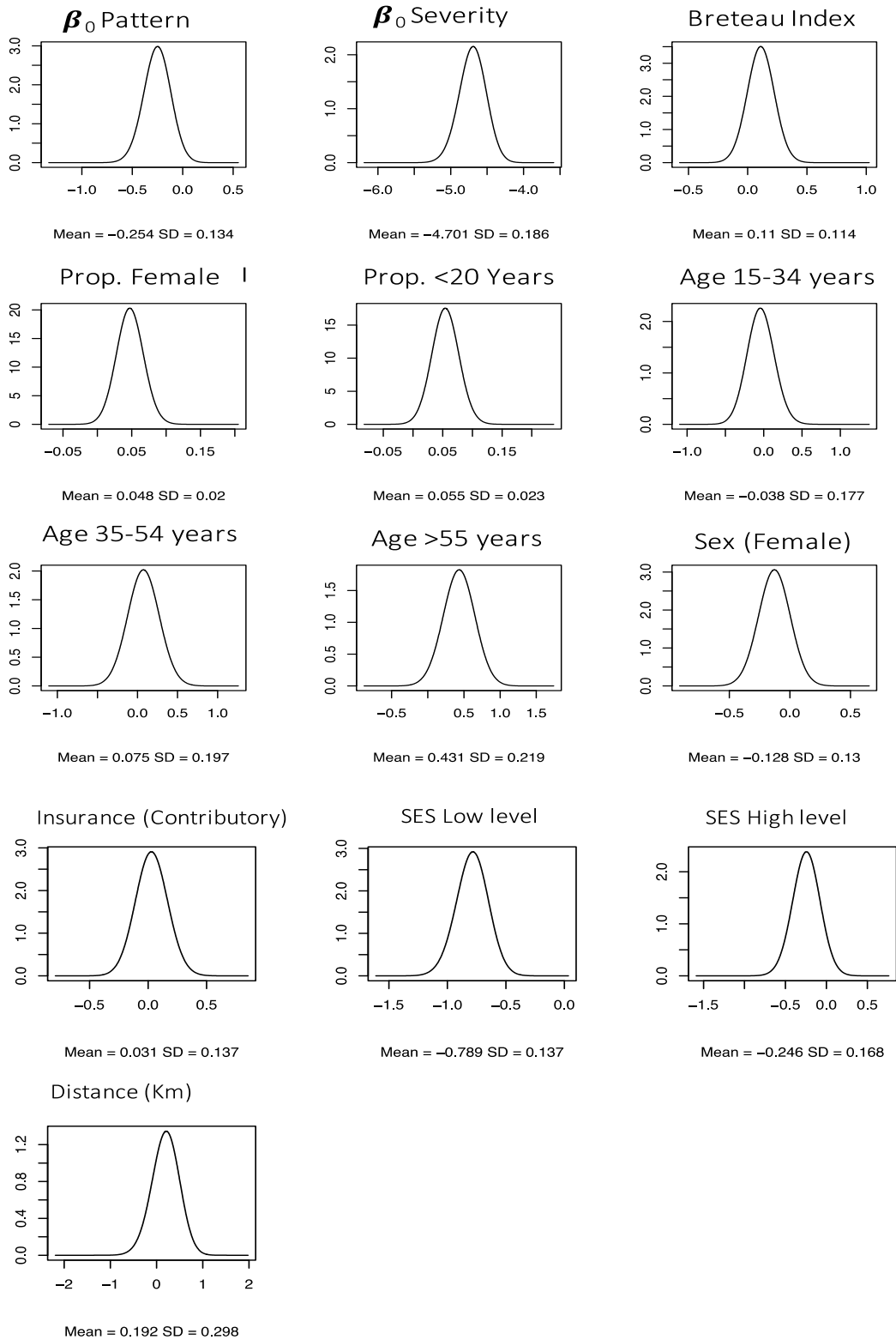
Fixed Effects	Posterior mean and 95% Credible intervals		
Overall Cases (Pattern)			
	IRR	2.5%	97.5%
Intercept (Pattern)	0.78	0.60	1.01
Proportion of Female cases	1.05	1.01	1.10
Proportion of cases <20 years old	1.06	1.01	1.11
Socio Economic Status (SES)			
Low SES	0.43	0.33	0.60
High SES	0.77	0.50	1.10
Breteau Index (Medium)	1.12	0.89	1.40
Severity (Marks)			
	OR	2.5%	97.5%
Sex (Female)	0.92	0.70	1.20
Age			
15-34 Years	1.05	0.70	1.50
35-54 Years	1.13	0.76	1.70
>55 Years	1.53	0.98	2.40
Contributory Insurance	0.89	0.67	1.20

Proportion of woman and cases under 20 years old indicate 10% increase. SES reference group = Medium SES Level; Breteau Index reference group = Low; Age group reference= <15 years of age; Sex reference= Male; Health Insurance reference is Subsidized scheme

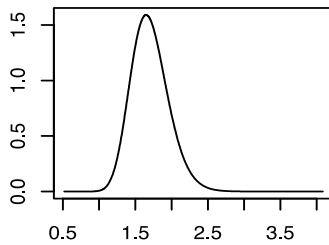
Table S5b. Posterior mean and 95% credible intervals of hyperparameters for the model with two separate spatial structures

Parameter	mean	sd	2.5%	97.5%	mode
Precision for $f_s^j(s_i) + u(s_i)$	1.69	0.26	1.24	2.26	1.63
Phi for $u(s_i)$	0.38	0.14	0.13	0.68	0.35
Precision for $g_s^{(2)}(s_i) + v(s_i)$	2.02	0.59	1.13	3.43	1.75
Phi for $v(s_i)$	0.09	0.09	0.00	0.35	0.01

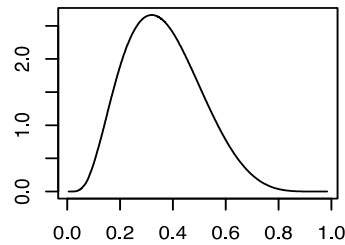
1.3 Posterior density of fixed and random effects of the final joint model using a single spatial component.



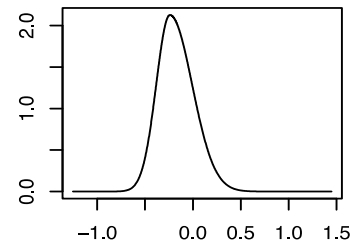
PostDens [Precision for i.spat1]



PostDens [Phi for i.spat1]



PostDens [Beta for i.spat2]



i.spat1= Spatial Effect for overall cases; i.spat2= Spatial effect for severe cases

Supplementary References:

1. Diggle PJ, Moraga P, Rowlingson B, Taylor BM. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. *Statistical Science* 2013;**28**(4):542-563.
2. Illian JB, Martino S, Sorbye S, Gallego-Fernandez JB, Zunzunegui M, Esquivias MP, Travis JMJ. Fitting complex ecological point process models with Integrated Nested Laplace Approximation. *Methods in Ecology and Evolution* 2013;**4**:305-315.
3. Simpson D, Illian JB, Lindgren F, Sorbye S, Rue H. Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *arXiv* 2017:1-26.
4. Illian JB, Sørbye SH, Rue H. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics* 2012:1499-1530.
5. Pinto Junior JA, Gamerman D, Paez MS, Alves Regina HF. Point pattern analysis with spatially varying covariate effects, applied to the study of cerebrovascular deaths. *Statistics in Medicine* 2015;**34**:1214-1226.
6. Waller LA, Gotway CA. Analyzing Public Health Data. . In: Shewhart WA, Wilks SS, eds. *Applied Spatial Statistics for Public Health Data*. Vol. 368 John Wiley & Sons, 2004;7-37.

7. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991;**43**(1):1-20.
8. Blangiardo M, Cameletti M. Spatial modeling. *Spatial and Spatio-temporal Bayesian Models with R-INLA* John Wiley & Sons, Ltd, 2015;173-234.
9. Congdon P. Models for spatial outcomes and geographical association. *Applied Bayesian Modelling* John Wiley & Sons, Ltd, 2014;312-363.