

# MODELO ESPACIAL BAYESIANO PARA LA DINÁMICA DE TRANSMISIÓN DE DENGUE EN PUERTO RICO PARA DATOS DEL 2014

## BAYESIAN SPACE MODEL FOR DENGUE TRANSMISSION DYNAMICA IN PUERTO RICO FOR 2014 DATA

Greivin Hernández-González <sup>1</sup>

Recibido: 22.05.19	Aprobado: 29.11.19
--------------------	--------------------

DOI: 10.15517/isucr.v20i42.41848

---

### Resumen

El propósito del presente estudio, ha sido investigar la dinámica de transmisión de dengue en los 76 municipios que componen la isla principal de Puerto Rico, para las semanas 1 a 4 y 32 a 36 del año 2014. Se utilizó un modelo espacial bayesiano para estudiar la relación posible entre incidencia, variables socioeconómicas, climáticas y ambientales. Una vez propuestos los modelos, se comparan los ajustes con índices como el WAIC (Watanabe, 2010), para determinar cual representa mejor a los datos. Además se determina con el índice I-Moran si existe correlación espacial en los residuos, ya que la existencia de la misma es un indicador de que el ajuste no es bueno en alguna medida. Los datos de temperatura y precipitaciones se debieron interpolar previamente, ya que las estaciones que los recolectan no están en los centros poblacionales, para ver los cálculos el lector puede referirse a (Hernández-González, 2017).

### Palabras clave

Puerto Rico; Dengue; Matemáticas; Epidemiología.

### Abstract

The purpose of this study has been to investigate the dynamics of dengue transmission in the 76 municipalities that make up the main island of Puerto Rico, for weeks 1 to 4 and 32 to 36 of 2014. A Bayesian spatial model was used to study the possible relationship between incidence, socioeconomic, climatic and environmental variables. Once the models are proposed, the settings are compared with indices such as WAIC (Watanabe, 2010), to determine which one

---

<sup>1</sup> Costarricense. Matemático. Profesor Universidad de Costa Rica, Sede Guanacaste. Email: [greivin.hernandez@ucr.ac.cr](mailto:greivin.hernandez@ucr.ac.cr),

represents the data best. It is also determined by the I-Moran index if there is spatial correlation in the residuals, since the existence of the index is an indicator that the adjustment is not good to some extent. Temperature and precipitation data should have been previously interpolated, since the stations that collect them are not in the population centers, to see the calculations the reader may refer to (Hernández-González, 2017).

### **Key words**

Puerto Rico; Dengue; Mathematics; Epidemiology.

### Introducción

El Dengue es la infección viral transmitida por artrópodos más importante de las últimas décadas para los seres humanos, tanto así que se estima que cerca de 2500 millones de personas están en riesgo de desarrollar dicha enfermedad (Chien, Lung-Chang. y cols., 2014), el riesgo aumenta en las regiones tropicales o subtropicales, ya que el mosquito *Aedes aegypti*, su principal vector de transmisión, es una especie endémica para la mayoría de estas regiones. A pesar de que existen modelos teóricos cuya finalidad es ajustar un modelo de propagación, los cuales dan gran importancia a variables como la temperatura y la precipitación, y buscan determinar patrones de transferencia; todavía hace falta mayor evidencia empírica para lograr mejores conclusiones.

Para el desarrollo de sistemas de alerta tempranos, que ayuden a enfrentar la transmisión de la enfermedad, es esencial entender las relaciones empíricas entre algunos factores meteorológicos, geográficos, socioeconómicos y la fiebre del dengue (Chien, Lung-Chang. Y cols., 2014). En este artículo se usa un modelo espacial Condicional Autoregresivo (CAR, por sus siglas en inglés) para estudiar el riesgo relativo en cada municipio de la Isla Principal de Puerto Rico. En el modelo propuesto analiza la relación de variables como: temperatura, precipitación y altitud, con el riesgo relativo de transmisión del dengue, aplicado a datos recolectados en el año 2014.

### Datos utilizados en la investigación

El modelo construido incorpora el riesgo relativo de contraer la 165demás165e165 del dengue por municipio de Puerto Rico, este riesgo se define como:

$$(1) \quad RR_i = \left( \frac{c_i}{P_i} \right) / \left( \frac{c_T}{P_T} \right)$$

$RR_i$ : representa el riesgo relativo para las personas que habitan en el municipio  $i$ -ésimo, por ejemplo los datos de la semana 1 de 2014 arrojan el comportamiento descrito en la figura (1) que se presenta a continuación:

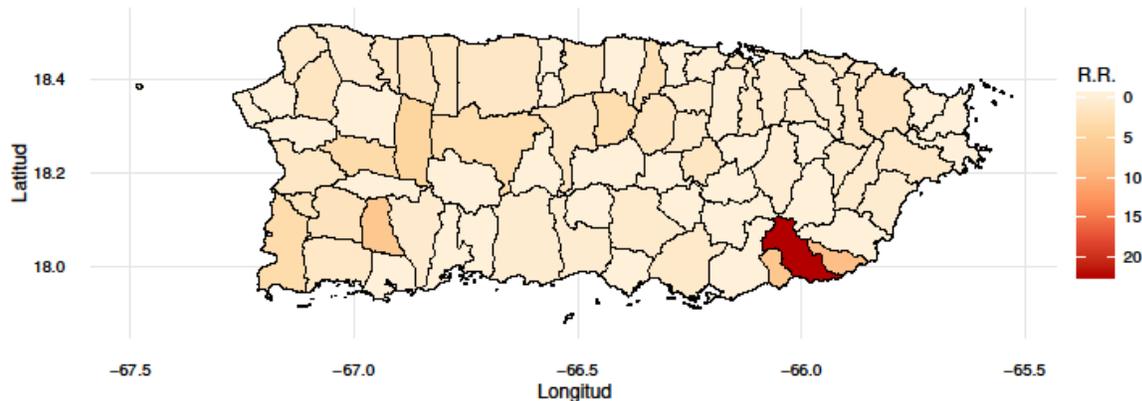


Figura 1: Mapa de colores para el riesgo relativo en semana 1

$c_i$ : es el número de presuntos casos de dengue para el municipio  $i$ . Este dato se obtuvo de los reportes semanales de casos presuntos de Dengue, reportes elaborados en conjunto por el Centro para el Control y Prevención de Enfermedades de Estados Unidos (CDC, por sus siglas en inglés) y el Departamento de Salud de Puerto Rico. Específicamente, fueron obtenidos en la página de internet de dicho departamento de salud, ver (CDC, Subdivisión de Dengue y Departamento de Salud, PR, 2014),

$P_i$ : es la población del municipio  $i$ , según el censo del año 2010, disponibles en la página de la Junta de Planificación del Gobierno de Puerto Rico (Bureau, 2011),

$c_T$ : es el número de casos presuntos totales de Puerto Rico, es decir que es la suma del número presunto de casos de todos los municipios,

$P_T$ : es la población total en Puerto Rico según el censo del año 2010.

Las variables independientes (también llamadas covariables) utilizadas dentro del modelo son:

1. Altitud por Municipio: esta covariable se obtuvo utilizando el API de Google Maps. El dato obtenido se refiere a la altitud de un punto en el centro poblacional de cada municipio, sin embargo, en el presente modelo se usó como un valor que representa toda el área del municipio.

2. Precipitación, Temperatura máxima y Temperatura mínima: los datos fueron obtenidos desde la base de datos Global Historical Climatology Network, en su página de internet, la cual es manejada por National Center for Environmental Information (NCEI), el cual pertenece al National Oceanic and Atmospheric Administration de los Estados Unidos de América (Menne., Matthew J. y cols., 2012).
3. Porcentaje de pobreza: fue obtenido desde el U.S Census Bureau, siendo datos del censo 2010.
4. Índice de vegetación mejorado (EVI, siglas para Enhanced Vegetation Index): los datos de esta covariable se obtienen utilizando el paquete MODISTools (Tuck, Sean. Y cols., 2015) desarrollado para el 167demás167e R-CRAN, el cual los extrae desde los archivos de NASA LPDAAC (Land Processes Distributed Active Archive Center of NASA). Este índice obtiene la respuesta de las variaciones estructurales del dosel arbóreo (hábitat que comprende la región de las copas y regiones superiores de los árboles de un bosque), incluyendo el índice de área foliar, tipo y arquitectura del dosel, y además de la fisonomía de las plantas. Fue desarrollado para optimizar la señal de la vegetación con sensibilidad mejorada para altas densidades de biomasa, esto se logra al separar la señal proveniente de la vegetación y la influencia atmosférica (Jiang, Zhangyan. Y cols., 2008).

#### Modelación con datos de área

En el presente apartado se estudian 5 modelos distintos, en el sentido que ajustan los datos con enfoques basados en diferentes supuestos, cada modelo realiza un ajuste para los casos de dengue usando regresión de Poisson (ya que se modelan datos de conteo), sin embargo, dos no son espaciales, y los resultados que se dieron en algunas semanas de estudio nos indican que no existe necesidad de modelos espaciales para estos datos. Los modelos son:

Tabla 1: Modelos de ajuste a comparar

<b>Modelo</b>	<b>Tipo</b>	<b>Estimación de Parámetros</b>	<b>Siglas</b>
---------------	-------------	---------------------------------	---------------

Lineal Generalizado (Agresti, Allan, 2013)	No espacial	Máxima Verosimilitud Bayesiano	GLM
Independiente, GLMM (Besag, Julian y cols, 1991)	No espacial	Bayesiano	Ind
Intrínseco (Besag, Julian y cols, 1991)	Espacial	Bayesiano	Int
Besag (Besag, Julian y cols, 1991)	Espacial	Bayesiano	BYM
Leroux (Leroux, Brian G y cols, 2000)	Espacial	Bayesiano	Ler

Datos y análisis exploratorio: La región de estudio es la isla principal de Puerto Rico, que consta de 76 municipios, los cuales tienen un promedio de 48877 habitantes según el censo de 2010. Las variables que se usan en cada modelo son: altitud, nivel o porcentaje de pobreza, número de habitantes (población), superficie por municipio y densidad poblacional, dichas variables no varían entre semanas. Los datos se resumen en el cuadro 2, mostrando los percentiles de sus distribuciones.

Tabla 2: Resumen de la distribución de datos fijos

Variable	Percentiles				
	0 %	25 %	50 %	75 %	100 %
Altitud (m.s.n.m.)	0	17.8996	40.8347	120.6659	602.8044
% Pobreza	0.2730	0.4572	0.5110	0.5685	0.6570
Población	6276	24682	36260	46135	395326
Superficie (mill <sup>2</sup> )	4.8	28.3	40.55	53.45	126
Densidad	171.5	613	877.9	1334.1	8253.2

Por otro lado están: precipitación medida en decimas de milímetros, temperatura mínima medida en grados Celsius, el índice de vegetación mejorado (EVI) y los casos presuntos de dengue, que se usan en lugar de los casos confirmados, ya que estos están fuertemente

correlacionados y además son mas sensitivos porque aproximadamente el 60% de los casos confirmados tienen muestras poco adecuadas para un diagnostico definitivo (Johansson, Michael A. y cols., 2009) y el riesgo relativo, este último calculado como el cociente de los casos presuntos y el riesgo de la población total de cada municipio.

La variable de respuesta en el presente estudio es dada por los casos de dengue, la cual se asume que sigue una distribución de Poisson, es decir,  $c_k \sim Poisson(\mu_k)$ , tal que  $\log(\mu_k) = \log(E_k) + \theta_k$ , en la última ecuación se cumple que  $RR_k = \frac{\mu_k}{E_k}$  y este término  $E_k$  es un offset (número esperado de casos en el municipio k) usado para controlar el tamaño de la población, mientras que el termino  $\theta_k$  contiene el efecto de las covariables, un efecto aleatorio espacialmente estructurado (en el caso de los modelos espaciales) y un efecto aleatorio sin estructura espacial.

En la tabla 3 se presenta un resumen para tres de las semanas de estudio en el presente trabajo, donde se muestra que en los datos de casos hay cierta estabilidad entre los municipios, ya que al menos un 75 % de los mismos presentan un máximo de 2 casos de dengue en las tres semanas, y una media entre 1 y 2 casos. En cuanto a las variables de precipitación, temperatura e índice de vegetación debemos recordar que son el resultado de interpolación espacial mientras que el índice de vegetación se obtuvo por interpolación de la serie temporal, ver (Hernández-González, 2017), por lo que para los modelos ajustados, estas variables podrían aportar un error aún mayor con respecto a si fueran variables medidas directamente (muestras).

Tabla 3: Resumen de la distribución de datos variables semanales

		Percentiles				
	Variable	0 %	25 %	50 %	75 %	100 %
Semana 1	Precipitación	0.4029	2.9044	15.5941	30.2027	39.9786
	Temperatura Mínima	16.3	18.38	19.35	20.58	23.96
	EVI	1773	3695	4276	4730	5860
	Casos	0	0	1	2	13
	Riesgo Relativo	0	0	0.7534	1.5141	22.2353
	Semana 2	Precipitación	0.6208	5.2910	16.93	28.71

	Temperatura Mínima	15.33	17.42	18.72	20.03	22.96
	EVI	1933	3469	4101	4581	5726
	Casos	0	0	1	2	9
	Riesgo Relativo	0	0	0.8441	1.4783	15.9586
Semana 3	Precipitación	0.0711	2.0580	3.507	6.032	49.46
	Temperatura Mínima	15.47	17.20	18.11	19.08	20.70
	EVI	1975	3447	4026	4415	5604
	Casos	0	0	1	2	12
	Riesgo Relativo	0	0	0.7613	1.5146	13.3807

En la figura 1 se puede ver existe un municipio (Patillas) con un riesgo muy superior, también sus vecinos, esto se debe a que dichos municipios según los informes del Departamento de Salud, pertenecieron a un proyecto de vigilancia aumentada (CDC, Sub- división de Dengue y Departamento de Salud, PR, 2014). Para las primeras 3 semanas el municipio de Patillas presenta el mayor número de casos de dengue, San Juan es en esas semanas posee el segundo lugar con mayor número de casos, pero la población de San Juan es poco mas de 20 veces la población de Patillas lo que se traduce en un riesgo relativo muy alto para el este municipio. Dado que el objetivo principal es ajustar modelos que usen todos los datos de la isla principal de Puerto Rico, y teniendo en cuenta la información del Departamento de Salud de Puerto Rico, la cual sugiere la existencia de mediciones muy diferentes para algunos municipios, se decidió usar la información completa de la isla, sin hacer ninguna valoración en el uso de herramientas para detectar valores atípicos (“outliers”) en los datos del estudio. Este comportamiento también se presento en las demás semanas de estudio.

Un tema importante a tratar es el de la autocorrelación espacial de las variables en estudio, en particular para el riesgo relativo, ya que es la variable dependiente del modelo, para llevar a cabo dicha exploración se usan pruebas de hipótesis para los índices de Moran y Geary. Los supuestos subyacentes en las pruebas son sensitivos a la matriz de peso utilizada, por ejemplo, una matriz de pesos estandarizada por las incrementa la influencia en los vínculos para observaciones con pocos vecinos, mientras que una matriz binaria varía la influencia de las

observaciones: las que tienen muchos vecinos están sobre-ponderadas en comparación con aquellas con pocos vecinos (Bivand, Roger S. y cols., 2013); las pruebas tienen como hipótesis nula que no existe autocorrelación espacial y de hipótesis alternativa que existe autocorrelación espacial, ya sea positiva o negativa, ambas pruebas se hicieron usando las funciones moran.mc() y geary.mc() del paquete spdep (Bivand, Roger. y cols., 2015, 2013) de R-CRAN. Los resultados de dichas pruebas se presentan en la tabla 4.

Tabla 4: Índices de Moran asociados a 3 variables de los modelos finales.

	Riesgo Relativo		Precipitación		Temperatura Mínima	
	I-Moran	P-valor	I-Moran	P-valor	I-Moran	P-valor
Semana 1	0.222	0.0013	0.8669	0.0001	0.7032	0.0001
Semana 2	0.074	0.0732	0.8491	0.0001	0.7725	0.0001
Semana 3	0.1963	0.0015	0.5658	0.0001	0.8664	0.0001
Semana 4	-0.051	0.7451	0.5891	0.0001	0.8325	0.0001
Semana 31	0.1664	0.0128	0.0162	0.3242	0.75	0.0001
Semana 32	0.2732	0.0006	0.4282	0.0001	0.7495	0.0001
Semana 33	0.3455	0.0001	0.234	0.0003	0.6958	0.0001
Semana 34	0.2058	0.0034	0.8432	0.0001	0.6009	0.0001

La prueba para altitud da como resultado un valor de 0,3213 para el estadístico con un p-valor de 0,0003, lo que indica que hay evidencia para rechazar la hipótesis nula de no autocorrelación espacial en dicha variable, por otro lado para el registro de pobreza se tiene un índice de Moran de 0,5684, teniendo en este caso un p-valor de 0,0001 por lo que se rechaza la hipótesis nula también es este caso. En la tabla 4 se puede ver que con la excepción de las semanas 2 y 4 para el riesgo relativo, semana 5 para precipitación, las pruebas indican que las variables están autocorrelacionadas espacialmente. Es importante aclarar que dichas pruebas se realizaron con una matriz de pesos estandarizada por filas, sin embargo, también se hizo el ejercicio con la matriz binaria y los resultados obtenidos son similares.

Modelos no espaciales: Los modelos ajustados sin tomar en cuenta una estructura espacial son dos: modelo lineal generalizado (GLM), el cual se ajusto usando máxima verosimilitud y por medio de cadenas de Markov de Monte Carlo, y un modelo lineal generalizado mixto (GLMM), también llamado Independiente en este trabajo (Lee, 2013), ajustado con MCMC. La forma que toma cada uno viene dada por:

$$casos_k \sim Poisson(\mu_k),$$

Donde en el GLM se tiene que:

$$(2) \quad \log(\mu_k) = \log(E_k) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6;$$

Mientras que el modelo Independiente se rige por la ecuación:

$$(3) \quad \log(\mu_k) = \log(E_k) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \theta_k$$

donde  $E_k$  es el offset que controla el tamaño de la población,  $\beta_0$  es el intercepto,  $\beta_1$  es el coeficiente para la variable de altitud,  $\beta_2$  es el coeficiente para el porcentaje de pobreza,  $\beta_3$  es la proporción del índice de vegetación en el modelo,  $\beta_4$  es un coeficiente relacionado a la densidad poblacional,  $\beta_5$  es el coeficiente para la precipitación promedio en el municipio,  $\beta_6$  es el coeficiente para la temperatura mínima y  $\theta_k$  es una variable aleatoria con distribución normal de media cero y varianza  $\sigma^2$ , a dicha varianza se le asigna una previa  $U(0,1000)$ , lo anterior con el objetivo de que sea una previa poco informativa en el modelo.

En el primer escenario se revisa el ajuste realizado al modelo (2) usando máxima verosimilitud y MCMC, donde es importante destacar que la diferencia entre los estimadores es muy poca, sin embargo, con máxima verosimilitud no se puede hablar de un intervalo de predicción para cada uno de los estimadores. Para realizar el ajuste con máxima verosimilitud se uso la función `glm()` del paquete "stats" (R Core Team, 2016), mientras que el ajuste bayesiano se realizo con la función `stan_glm()`, la cual pertenece al paquete "rstanarm" (Gabry y Goodrich, 2016); las previas utilizadas por dicha función para los parámetros de ajuste asociados a las covariables son distribuciones normales  $N(0; 100)$ .

La primera conclusión que se puede dar, es que las covariables de EVI y Densidad no aportan mayor significancia al modelo, ya que para las 8 semanas de estudio el valor de la mediana (y también la media) de los parámetros correspondientes eran cero o muy cercanos a cero, además que el intervalo de predicción para dichos valores es muy pequeño y siempre contiene al 0. Los parámetros para las covariables EVI y densidad se mantienen muy cerca de cero o son cero para las demás semanas de estudio, aún así se hace el ajuste de los otros modelos teniendo en cuenta esas variables, pero el resultado obtenido en cada uno los ajustes fue el mismo, por lo que se volvió a ajustar los datos de todas las semanas con todos los modelos propuestos en la tabla 1, pero sin dichas covariables, es decir, los nuevos modelos a ajustar son:

$$(4) \quad \text{GLM: } \log(\mu_k) = \log(E_k) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_5 X_5 + \beta_6 X_6;$$

En tanto que el modelo Independiente es:

$$(5) \quad \log(\mu_k) = \log(E_k) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_5 X_5 + \beta_6 X_6 + \theta_k;$$

Donde cada elemento en las ecuaciones anteriores ya se han explicado. En el presente estudio los intervalos de predicción al 95 % del modelo lineal generalizado, sugieren que las covariables de mayor peso son altitud y pobreza ya que para todas las semanas ninguno contiene al cero, mientras que para precipitación y temperatura mínima el cero estaba presente en cada intervalo de predicción. Por lo anterior, en caso de existir una estructura espacial que no está siendo tomada en cuenta en dicho modelo, altitud y pobreza tienen un rol importante en explicar el modelo espacial en el riesgo relativo de contraer dengue en alguno de los municipios de Puerto Rico (Lee, 2013). En la tabla 5 se presentan tales intervalos.

Tabla 5: Intervalos de predicción para parámetros de modelo GLM reducido

	Semana1		Semana2		Semana3		Semana4	
	2.50 %	97.50 %	2.50 %	97.50 %	2.50 %	97.50 %	2.50 %	97.50 %
Intercepto	-4.2673	3.9786	-4.3081	4.0453	-4.2947	3.973	-4.304	4.0079

Altitud	-0.0082	-0.0006	-0.0082	-0.0006	-0.0081	-0.0007	-0.0082	-0.0006
Pobreza	-8.0542	-3.9947	-8.0528	-3.9981	-8.0602	-4.0028	-8.0889	-3.9948
Prec	-0.0125	0.0128	-0.0124	0.0127	-0.0126	0.0125	-0.0124	0.0126
Tmin	-0.0388	0.2891	-0.0408	0.2889	-0.0386	0.2877	-0.0404	0.2902
	<b>Semana31</b>		<b>Semana32</b>		<b>Semana33</b>		<b>Semana34</b>	
Intercepto	-4.2945	4.0233	-4.2987	3.9673	-4.2939	4.0512	-4.2798	3.9536
Altitud	-0.0083	-0.0006	-0.0082	-0.0007	-0.0082	-0.0007	-0.0082	-0.0007
Pobreza	-8.0738	-4.0062	-8.0409	-4.0092	-8.0536	-4.0066	-8.0787	-4.0341
Prec	-0.0126	0.0128	-0.0125	0.0126	-0.0124	0.0126	-0.0123	0.0125
Tmin	-0.042	0.2892	-0.0396	0.292	-0.0413	0.2907	-0.0381	0.2898

Con el nuevo ajuste al parecer para estos datos el GLM no es una buena elección según el criterio de información de la devianza, ya que es menor en el modelo independiente para todas las semanas, por otro lado, el WAIC sugiere que las semanas 2 y 3 se ajustan mejor con un GLM en lugar de un GLMM, y las restantes semanas con el Independiente (en adelante GLMM).

A partir de aquí, se hace la comparación y selección de modelos usando únicamente el índice de Watanabe-Akaike, porque el DIC (Spiegelhalter, David J., y cols, 2002) no tiene un enfoque bayesiano en su totalidad ya que se basa en un estimador puntual, mientras que el WAIC tiene un enfoque completamente bayesiano en el sentido que usa la distribución posterior completa, siendo además una mejora para el DIC en modelos bayesianos y es asintóticamente equivalente a la validación cruzada propuesta por Vehtari en (Vehtari, Aki. y cols., 2016), conocida como loo-cv (Leave-One-Out Cross Validation). Otras de las desventajas del DIC es que puede producir estimadores negativos del número efectivo de parámetros en un modelo y no está definido para modelos singulares. Finalmente, a diferencia del DIC, el WAIC es invariante a la parametrización y también es útil en la comparación para modelos con distribuciones singulares (Vehtari, Aki. y cols., 2016).

Tabla 6: Parámetros de ajuste para GLM

Semana	1	2	3	4	31	32	33	34
Intercepto	-0.1296	-0.0997	-0.1126	-0.1116	-0.1235	-0.1210	-0.1014	-0.1223
Altitud	-0.0039	-0.0039	-0.0039	-0.0039	-0.0039	-0.0039	-0.0039	-0.0039
Pobreza	-6.0051	-6.0086	-6.0108	-6.0110	-6.0042	-6.0031	-6.0158	-6.0197
Prec	-0.0002	-0.0002	-0.0002	-0.0002	0.0000	-0.0002	-0.0001	-0.0001
Tmin	0.1251	0.1234	0.1242	0.1240	0.1244	0.1244	0.1235	0.1243

En el GLM, los estimadores se mantienen estables en las semanas de estudio, se puede ver en la tabla 6 que el valor que toma para altitud es prácticamente constante (con 4 dígitos decimales), y los valores asociados a las demás covariables no tienen cambios grandes de semana a semana. Otro resultado obtenido desde el modelo lineal generalizado es que riesgo relativo es inversamente relacionado con altitud, ya que los parámetros de ajuste son negativos para las 8 semanas, eso quiere decir que a mayor altitud existe menor riesgo de contraer dengue, resultado que también obtienen otros autores, por ejemplo (Mena, Nelson. y cols., 2011). Según los resultados del GLM, el índice de pobreza está relacionado de manera inversa al riesgo de contraer el virus, sin embargo, en el modelo Independiente se concluye que a pesar de la fuerte asociación que hay entre riesgo relativo y pobreza, no existe un efecto consistente entre las semanas, porque en algunas semanas el efecto es negativo y en otras es positivo, dicho resultado es semejante al que obtuvieron en un estudio del 2009 realizado por Johansson, Dominici y Glass en (Johansson, Michael A. y cols., 2009).

Resultados de modelos espaciales: Se ajustan 3 modelos espaciales: intrínseco, Besag y Leroux. El primero asume que existe completa autocorrelación espacial en la variable dependiente, el siguiente es una suma entre el modelo intrínseco y el independiente (5) y el último se puede ver como una ponderación entre los modelos intrínseco e independiente. La variable dependiente es el número de casos de dengue para cada municipio, la que suponemos que sigue una distribución de Poisson, es decir:

$$c_k \sim \text{Poisson}(\mu_k), \text{ donde } \log(\mu_k) = \log(E_k) + X_k^T \beta + \theta_k.$$

Para los modelos intrínseco y Leroux se cumple ( Lee, 2013 ):

- Intrínseco:

$$(6) \quad \theta_k | \theta_{-k}, W, \tau^2 \sim N \left( \frac{\sum_{i=1}^n w_{ki} \theta_i}{\sum_{i=1}^n w_{ki}}, \frac{\tau^2}{\sum_{i=1}^n w_{ki}} \right)$$

$$(7) \quad \tau^2 \sim IG(1; 1/1000)$$

- Leroux:

$$(8) \quad \theta_k | \theta_{-k}, W, \tau^2 \sim N \left( \frac{\rho \sum_{i=1}^n w_{ki} \theta_i}{\rho \sum_{i=1}^n w_{ki} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{i=1}^n w_{ki} + 1 - \rho + \sum_{i=1}^n w_{ki}} \right)$$

$$(9) \quad \tau^2 \sim IG(1, 1/1000)$$

$$(10) \quad \rho \sim U(0, 1)$$

Los tres modelos se ajustan con funciones del paquete CARBayes (Lee, 2013), presente en el programa estadístico R-CRAN (R Core Team, 2016). Se realizan 50000 simulaciones con Monte Carlo vía cadenas de Markov, de las que se queman 10000 con el fin de haber alcanzado estabilidad o equilibrio en la cadena, donde los valores de los parámetros de la distribución gamma inversa son los preestablecidos en el paquete, y la matriz W utilizada es la matriz binaria definida como  $w_{ij} = 1$  si los municipios ij comparten alguna fracción de sus fronteras y  $w_{ij} = 0$  en otro caso. Además, para explorar convergencia en las cadenas de Markov se usa el diagnóstico de Geweke, con el que se concluye que en todas las semanas cada parámetro ha convergido a su distribución estacionaria.

En la figura 2 se muestra el comportamiento de las cadenas para los parámetros de precipitación y temperatura mínima en la semana 1 del modelo BYM, a pesar de tener algunos cambios en el patrón, cerca de las iteraciones 15000 y 30000 por ejemplo, se puede decir que se alcanza estabilidad en la cadena, conclusión que se obtiene al ver en conjunto la figura antes mencionada y el diagnóstico de Geweke.

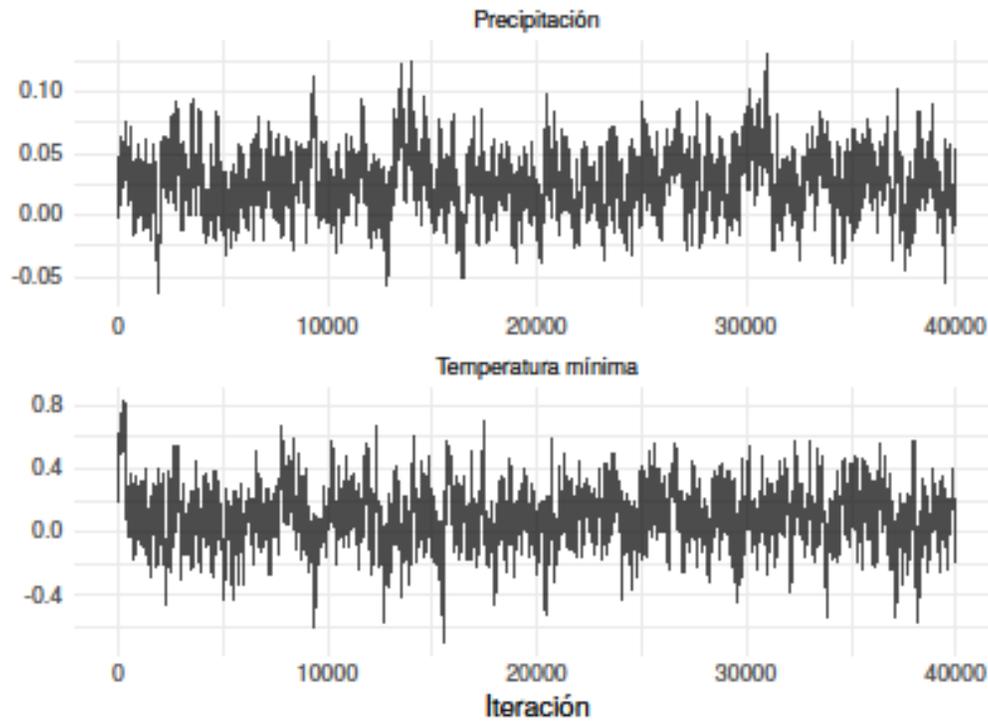


Figura 2: Traceplot para parámetros de Prec y Tmin en semana 1, modelo BYM.

El modelo usado para hacer inferencia es el Besag-York-Mollie, debido a que el criterio de información de Watanabe-Akaike(WAIC) tiene un valor inferior en 6 de las 8 semanas de estudio, y en las semanas donde otro modelo ajusta mejor según el criterio del WAIC, se tiene una diferencia muy pequeña. En la tabla 7 se muestra la medida del criterio para los 5 modelos bayesianos.

Tabla 7: Criterio de Información de Watanabe-Akaike.

Modelo	GLM Bayes	Independiente	Intrínseco	BYM	Leroux
Semana 1	249.4	229.9	231.29	231.22	229.13
Semana 2	249.6	260.11	246.8	244.65	256.92
Semana 3	249	258.45	243.65	243.46	252.98
Semana 4	249.7	244.44	251.65	233.88	259.96
Semana 31	249.6	222.26	212.12	211.98	212.86
Semana 32	249.2	189.14	191.99	190.85	189.81
Semana 33	249.1	213.75	199.82	196.67	200.29
Semana 34	249.1	195.51	190.02	189.83	190.03

3.3.1. *Resultados del modelo Besag-York-Mollie.* En lo siguiente se estudia el resultado obtenido por el ajuste del modelo BYM, dado que en la sección anterior se concluye que hace un mejor ajuste de predicción según el WAIC.

Gráficos de dispersión con líneas de regresión en la figura 3 representan las relaciones (lineales) crudas de las variables independientes con la variable dependiente para la semana 1. Dicho gráfico muestra que la tasa de incidencia de dengue (RR) está asociado positivamente con el porcentaje de pobreza en esa semana, lo mismo pasa en las semanas 2,3 y 4, mientras que para las semanas 31 a 34 cambia a negativa, con temperatura mínima se da una relación positiva en las 8 semanas, con altitud la relación es negativa en casi todas las semanas con excepción de la cuarta, donde parece no existir relación(lineal) alguna, precipitación no presenta una relación definida en la primer semana, negativa en la 32 y positiva en el resto.

El modelo BYM tiene dos efectos aleatorios, uno estructurado espacialmente, donde se incorpora la información de la matriz de vecindarios  $W$ , además de que se estima un parámetro  $\tau^2$  para la varianza y otro efecto sin estructura:  $\theta_k$ , independientes con distribución normal de media nula y varianza  $\sigma^2$ , estimada durante el ajuste con MCMC. Este modelo es el más usado en la práctica (Lee, 2013), sin embargo, se requiere estimar dos efectos aleatorios para cada dato puntual, mientras que solo la suma de estos es identificable en los datos, en la práctica uno de los efectos aleatorios es dominante sobre el otro, pero no se puede saber de antemano (Besag, Julian. y cols, 1991), es por eso que el paquete CARBayes retorna la suma en las muestras del MCMC ( $\varphi_k$ ).

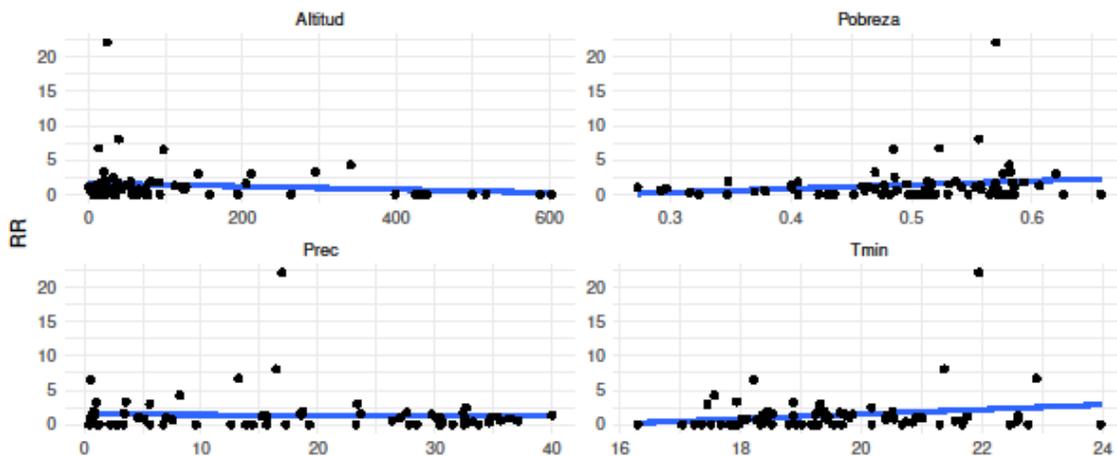


Figura 3: Gráficos de dispersión de variables vs riesgo relativo, semana 1

En la tabla 8 es claro que el parámetro de varianza para el efecto no estructurado es pequeño, eso indica que el efecto es muy cercano a cero ya que la media del mismo es cero. Por otro lado, la varianza para el efecto espacial es mayor que 1 (uno) para las semanas 1, 31, 32, 33 y 34, 0.023 en la semana 4, 0.86 y 0.68 para las semanas 2 y 3 respectivamente. La fila para  $\psi$  representa la mediana de dicho parámetro por semana, pero se debe aclarar que para cada semana existen 76 valores distintos para tal suma, ya que al municipio k-ésimo se le calcula  $\psi_k = \phi_k + \theta_k$ .

Tabla 8: Mediana de parámetros de efectos aleatorios para el modelo BYM

	Semana							
	1	2	3	4	31	32	33	34
$\tau^2$	2,0102	0,8674	0,6850	0,0231	1,9953	2,4995	2,8391	1,1925
$\sigma^2$	0,0019	0,0100	0,0018	0,3301	0,0020	0,0024	0,0021	0,0020
$\psi$	-0,0449	0,0020	0,0088	-0,0036	0,0037	0,0061	-0,0132	0,0040

Los estimadores posteriores para casos de dengue indican que las regiones de mayor incidencia en la semana 1 están localizadas al noreste y sureste de Puerto Rico, donde la región costera es la que más incidencia presenta según los resultados, mientras que los municipios con una incidencia menor se localizan en la región central del país. Por otro lado, en los residuos de la primera semana, se puede ver que no hay una estructura definida, además en un municipio se obtiene un valor residual muy grande comparado con los demás, eso puede dar a entender que existe un valor atípico en los datos para tal región, lo que puede conducir a la eliminación del mismo en el modelo, sin embargo, no se elimina pues al hacerlo se pierde mucha información, ya que no existe independencia entre los datos. En las semanas posteriores se mantiene el comportamiento de los residuos, es decir, son valores sin aparente estructura y el residuo para el municipio en cuestión sigue siendo superior con respecto a los demás, en la figura 4 se resumen los datos de casos, las predicciones del modelo y sus respectivos residuos.

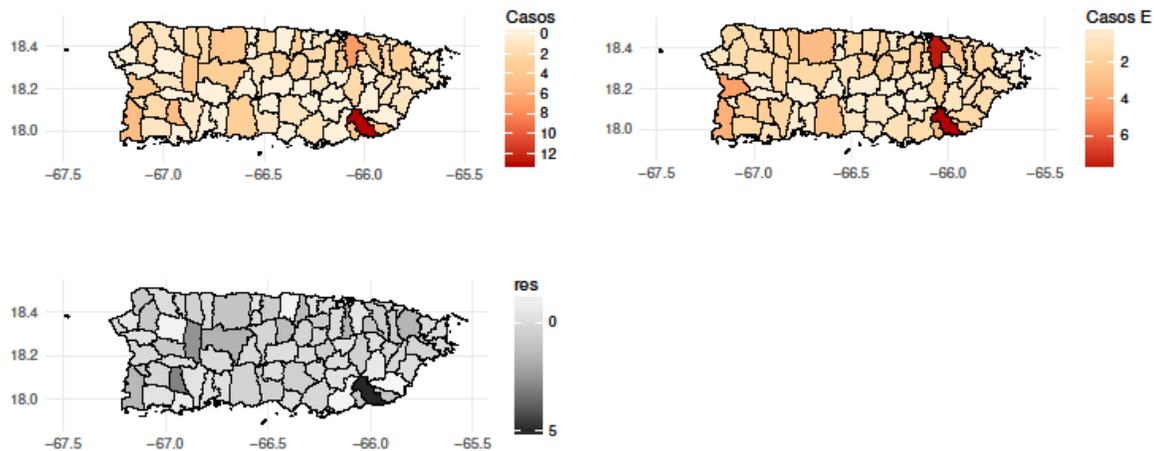


Figura 4: Casos, casos estimados y residuos de predicción en semana 1

La variación estructural estimada después de tomar en cuenta las variables climáticas, geográfica y social, indican que los municipios con alta incidencia de dengue se localizan al sureste en las primeras 4 semanas, también al norte en las semanas 2 a 4, y al suroeste en la semana 1, para todas las semanas los municipios con mayor riesgo se encuentran cerca de la costa. La gura 5 muestra el efecto aleatorio espacialmente estructurado de las semanas 1 a 4 luego de tomar en cuenta las covariables, es decir, una representación suavizada espacialmente del riesgo residual (Hu, Wenbiao. y cols., 2012).

En la tabla 9 se muestran los parámetros de regresión para las variables climáticas, social y geográfica, aunque no se puede definir el tipo de relación entre riesgo relativo y pobreza, si es positiva o negativa, queda claro que es una variable explicativa importante para predecir el riesgo de contraer dengue en Puerto Rico.

Tabla 9: Parámetros de regresión, modelo BYM

Parámetro	Semana							
	1	2	3	4	31	32	33	34
Intercepto	-5.6427	-4.5751	-8.7347	-5.1580	0.5305	4.9486	-3.4229	-3.8443
Altitud	-0.0022	-0.0022	-0.0015	0.0005	-0.0018	-0.0014	-0.0029	-0.0012
Pobreza	6.8104	4.2006	2.0425	1.4259	-5.7478	-6.9323	-4.9921	-2.7388
Prec	0.0261	0.2728	0.0702	0.0015	0.0019	0.0052	0.0292	-0.0003

Tmin	0.0985	0.1028	0.4208	0.2290	0.0661	-0.1232	0.1901	0.2231
------	--------	--------	--------	--------	--------	---------	--------	--------

Con excepción del resultado obtenido en la semana 32 y 34 para los parámetros de temperatura y precipitación respectivamente, el modelo de Besag sugiere que ante un aumento en lluvias o en la temperatura del país, se puede dar un aumento en la incidencia de dengue. También se puede concluir de los parámetros de altitud, que un aumento en altitud puede reflejarse en una disminución en el riesgo de enfermarse por dengue, con la excepción del estimador obtenido en la semana 4.

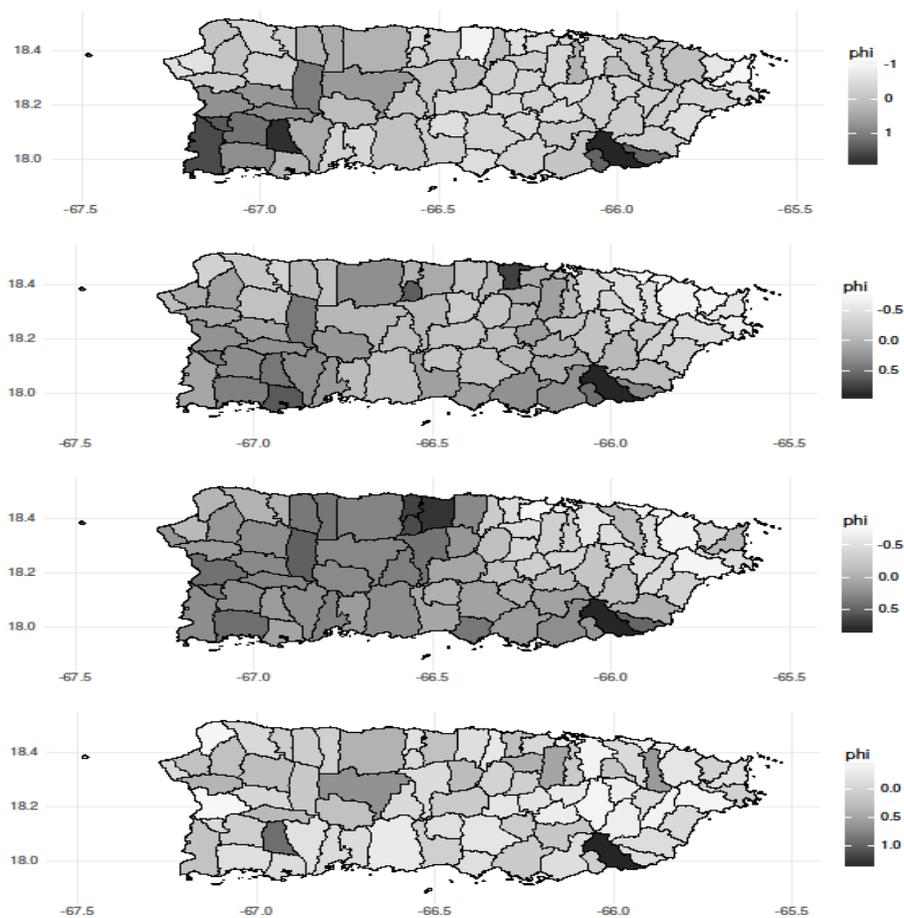


Figura 5: Efecto aleatorio espacial de riesgo relativo, semana 1 a semana 4

En la tabla 10 se presentan los intervalos de predicción para los parámetros de las covariables a un 95 % de credibilidad. Es claro en dichos intervalos que existe al menos una probabilidad de 0.95 de que el parámetro asociado a pobreza sea positivo para las semanas 1 y

2, negativo para las semanas 31 y 32, misma probabilidad para la tercera semana de obtener un parámetro positivo asociado a temperatura.

Tabla 10: Intervalos de predicción en modelo BYM reducido al 95%

	Semana 1		Semana 2		Semana 3		Semana 4	
	2.50 %	97.50 %	2.50 %	97.50 %	2.50 %	97.50 %	2.50 %	97.50 %
Intercepto	-13.3816	1.5322	-10.0228	0.6435	-16.5086	-1.3275	-11.2701	1.0308
Altitud	-0.0059	0.0009	-0.0054	0.0005	-0.0047	0.0013	-0.0025	0.0033
Pobreza	1.9107	11.5726	0.4364	7.5284	-2.0004	5.3757	-2.3018	4.7892
Prec	-0.0194	0.0793	-0.2811	0.8638	-0.4998	0.5451	-0.0348	0.0372
Tmin	-0.2631	0.4505	-0.1421	0.358	0.0561	0.8323	-0.0595	0.5257
	Semana 31		Semana 32		Semana 33		Semana 34	
Intercepto	-8.3289	8.838	-5.3589	17.7279	-14.9854	7.5936	-15.0557	7.062
Altitud	-0.006	0.0021	-0.0057	0.0025	-0.0093	0.002	-0.0055	0.0023
Pobreza	-10.7858	-0.8468	-15.4904	-0.7633	-12.3844	1.6299	-7.6124	1.9686
Prec	-0.0161	0.0189	-0.0211	0.0287	-0.0371	0.0825	-0.014	0.0139
Tmin	-0.277	0.4386	-0.6545	0.3344	-0.2909	0.7074	-0.2511	0.7248

Se realizan pruebas para revisar la presencia de dependencia espacial en los residuos del modelo BYM, tanto con el índice de Geary como con el de Moran, se concluye que no hay autocorrelación espacial en dichos residuos. Nuevamente se hace uso de las funciones de `spdep`(Bivand, Roger. y cols., 2015, 2013) en R-CRAN, `geary.mc()` y `moran.mc()`, donde se utilizan 5000 permutaciones con el objetivo de obtener mejores resultados, ya que en (Gelfand, Allan E. y cols., 2004) aunque sugieren el uso de dichos índices como herramientas exploratorias de asociación espacial, también sugieren usarlos con un enfoque Monte Carlo realizando permutaciones de los datos en las regiones de estudio, y eso es precisamente lo que realizan las funciones que se usan en este apartado.

Tabla 11: Pruebas de significancia espacial en residuos de modelo BYM.

Índice/Semana	1	2	3	4	31	32	33	34

I-Moran	-0.15	-0.15	-0.12	-0.12	-0.09	-0.07	-0.12	0.01
P-valor	0.99	0.99	0.96	0.95	0.87	0.79	0.94	0.33
C-Geary	1.3	1.28	1.27	1.26	1.08	1.08	1.08	1.01
P-valor	0.99	0.99	0.99	1.99	0.83	0.84	0.84	0.57

En la tabla 11 se puede observar que para ninguna de las semanas y los índices hay evidencia suficiente para rechazar la hipótesis nula de no autocorrelación espacial en los residuos. La matriz utilizada en tales pruebas se obtiene al estandarizar por filas la matriz binaria de asociación.

### Conclusiones

El presente estudio tiene la ventaja de usar modelos espaciales sofisticados, que a pesar de no ser propuestos recientemente, es hasta hace poco que son utilizados debido al auge de los programas computacionales. Por otro lado, no dejan de ser modelos sencillos y funcionales para evaluar posibles predictores para la incidencia de dengue.

Los resultados obtenidos en estudios de esta índole, pueden tener implicaciones importantes en las decisiones de salud pública, al identificar factores de riesgo y regiones de alto riesgo, decisiones que se pueden tomar con el objetivo de prevenir y controlar posibles epidemias.

Entre algunas de las limitaciones es que pueden existir sesgos en la medición o falta de información más detallada, un ejemplo puede darse con las notificaciones por casos de dengue, ya que podrían ser menores a los casos reales, debido a que puede haber personas infectadas que no buscan atención médica. Puede haber poca información biológica o de comportamiento a nivel individual e incluso comunitario, por ejemplo la densidad poblacional del mosquito transmisor, inmunidad poblacional al virus, comportamiento humano, lo que puede confundir la asociación entre las variables utilizadas y la transmisión de la enfermedad.

En los ajustes realizados, se concluye que las variables de índice de vegetación mejorada y densidad poblacional no aportan mayor significancia a los modelos, pero este resultado solo habla del comportamiento de las 8 semanas de estudio, por lo que no se puede concluir que siempre será así. Por otro lado, las variables de altitud y pobreza resultaron estar muy asociadas

con el riesgo de contraer dengue según los modelos ajustados, por lo que pueden ser buenas variables predictivas para el comportamiento de transmisión de dicha enfermedad, pero, se necesitan tomar en cuenta más semanas para obtener mejores conclusiones.

Finalmente, una limitación más de los modelos acá presentados, se da en la parte temporal, ya que ninguno de ellos toma en cuenta la posible correlación temporal que pueden tener los casos de dengue con las demás variables e incluso entre sí mismos.

## Referencias

- [1] Agresti, Allan. (2013). *Categorical Data Analysis* (3.<sup>a</sup> Ed.). New Jersey: John Wiley & Sons.
- [2] Besag, Julian, y cols. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 120.
- [3] Bivand, Roger, y cols. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis*, 45, 150-179.
- [4] Bivand, Roger, y cols. (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, 63, 1-36.
- [5] Bivand, Roger S., y cols. (2013). *Applied spatial data analysis with R* (2.<sup>a</sup> Ed.). New York: Springer.
- [6] Bureau, U. C. (2011). Datos del Censo 2010 de Puerto Rico [Web]. Descargado de <http://www.jp.gobierno.pr/PortalJP/Default.aspx?tabid=120> ([Web; accedido el 02-02-2015])
- [7] CDC, Subdivisión de Dengue, y Departamento de Salud, PR. (2014). Informe Semanal de Vigilancia del Dengue [Web]. Descargado de <http://www.salud.gov.pr/Estadisticas-Registros-y-Publicaciones/Pages/Dengue.aspx> ([Web; accedido el 15-02-2015])
- [8] Chien, Lung-Chang., y cols. (2014). Impact of meteorological factors on the spatiotemporal patterns of dengue fever incidence. *Environment International*, 74, 46-56.
- [9] Gabry, J., y Goodrich, B. (2016). *rstanarm: Bayesian Applied Regression Modeling via Stan* [Manual de software informático]. Descargado de <https://CRAN.R-project.org/package=rstanarm> (R package version 2.13.1)
- [10] Gelfand, Allan E., y cols. (2004). *Hierarchical Modeling and Analysis for Spatial Data* (1.<sup>a</sup> ed.). Boca Raton, Florida: Chapman & Hall/ CRC.

- [11] Hernández-González, G. (2017). Modelo Espacial Bayesiano para la Incidencia de dengue en la Isla Principal de Puerto Rico para el año 2014 (Tesis de Máster no publicada). Sistema de Estudios de Posgrado, Universidad de Costa Rica, San Pedro de Montes de Oca, San José, Costa Rica.
- [12] Hu, Wenbiao, y cols. (2012). Spatial Patterns and Socioecological Drivers of Dengue Fever Transmission in Queensland, Australia. *Environmental Health Perspectives*, 120, 260-266.
- [13] Jiang, Zhangyan, y cols. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112, 3833-3845.
- [14] Johansson, Michael A., y cols. (2009). Local and Global Effects of Climate on Dengue Transmission in Puerto Rico. *PLoS Neglected Tropical Diseases*, 3.
- [15] Lee, D. (2013). CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors. *Journal of Statistical Software*, 55, 1-24.
- [16] Leroux, Brian G., y cols. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. *Institute for Mathematics and Its Applications*, 116, 179-191.
- [17] Mena, Nelson, y cols. (2011). Factors associated with incidence of dengue in Costa Rica. *Revista Panamericana de Salud Publica*, 29, 234-242.
- [18] Menne., Matthew J., y cols. (2012). An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*, 29, 897-910.
- [19] R Core Team. (2016). R: A Language and Environment for Statistical Computing [Manual de software informático]. Vienna, Austria. Descargado de <https://www.R-project.org/>.
- [20] Spiegelhalter, David J, Best, Nicola G., Carlin, Bradley P., y Van-Der-Linde, Angelika. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583-639.
- [21] Tuck, Sean, y cols. (2015). MODISTools: MODIS Sub setting Tools [Manual de software informático]. Descargado de <http://cran.r-project.org/package=MODISTools> (R package version 0.94.6)
- [22] Vehtari, Aki, y cols. (2016). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 1-20.
- [23] Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571-3594.